Journal of Education for Pure Science- University of Thi-Qar
Vol.14, No. 4 (2024)
*Website: jceps.utq.edu.iq* *Email: jceps@eps.utq.edu.iq*

**DOI: https://doi.org/10.32792/jeps.v14i3.462**

# Personalized Chatbot Responses using Reinforcement Learning and User Modeling

**Muna G. Abd Alkreem[1]** **Rasha B. Yousif[2]** **Ali B. Yousif[3]**

[1,2]Department of Computer Science, College of Computer Science and Mathematics,University of Thi-Qar,Iraq.

[3]Department of Computer Science, College of Education for Pure Science, University of Thi-Qar, Thi-Qar, Iraq.

[1]Email : munaghani23@utq.edu.iq [2] Email: rasha.b.alkhafaji@utq.edu.iq [3]Email: ali.alkhafaji.eps@utq.edu.iq

**Received 25/06/2024, Accepted 04 /08/2024, Published 01 / 12/2024**



## Abstract

The research focuses on chatbot interaction enrichment using reinforcement learning and user modeling. It aims to develop a personalized RL-based response generation framework for the optimization of satisfaction, engagement, and completion rates for the users. Anonymized historical interaction data from 500 users were collected to generalize user-profiles and contextual models. This was done by extracting features and observing and evaluating the control and experimental groups to test the efficacy of the working personalized system. Results indicated an overall increase of 35 percent in user satisfaction, 50 percent in session duration, and 25 percent in rates of completion of tasks against the traditional rule-based system. The results are very much in line with current literature on the improvements personalization can bring to the user's experience across many domains. The results from this study thus propose that personal AI systems powered with fine-grained models of users and reinforcement learning could obtain more engaging and efficient user interactions. The result has further-reaching implications for, for example, e-learning, customer service, and healthcare applications.

### Keywords

Personalized Chatbot Responses, Reinforcement Learning, User Modeling, Proximal Policy Optimization, User Engagement

## Introduction

Personalized chatbot responses have become critical to enhancing user experience, be it customer service or educational platforms. Even though artificial intelligence and natural language processing have refined chatbots, much has to be accomplished in making these interactions genuinely personalized according to each user's preferences and behaviors. Recent work has demonstrated that RL combined with user modeling is a great way to enhance the adaptivity and responsiveness of chatbots, which will eventually yield high satisfaction and engagement among users (Den et al., 2019).

One of the methods increasingly applied to chatbot development is reinforcement learning, a form of machine learning whereby an agent makes decisions by acting and gathering feedback from its environment. Efficient and flourishing methods for setting the responses of chatbots have been given by variants such as Proximal Policy Optimization through exploration-exploitation balancing at train time. In this manner, it enables chatbots to fine-tune their strategy based on user interaction and, hence, yield more relevant and satisfactory responses (Zhong et al., 2021).

Parallel to RL, techniques from user modeling allow for capturing and representing the features, preferences, and contexts of users. An approach allowing for extracting user intents, emotions, and preferences from history interaction data enables dynamic user profiles that change over time (Hare et al., 2022). Integration of user modeling with RL strengthens the personalization of chatbot responses and further ensures contextual awareness by making the chatbot responsive to individual users' requirements (Patel et al., 2021).

Their application in various fields underscores the significance of such personalized chatbots. For example, it has been shown that personalized chatbots applied in educational settings make a difference in learning outcomes through tailoring content delivery to individual student needs (Shumanov et al., 2021).

In customer support, chatbot applications with personalization have higher customer satisfaction since they respond appropriately and on time, increasing the quality of service altogether. Such applications manifest that there exists a potential for combining user modeling and RL techniques in the development of more efficient and user-oriented chatbot systems. The present research is therefore aimed at developing and assessing a chatbot response generation framework that is based on reinforcement learning and user modeling. The paper seeks to contribute to chatbot personalization and effectiveness in applications through more systematic experimentation about RL algorithms and different techniques for user modeling. The results add to the knowledge base from AI-driven personalized systems won through chatbot optimization aimed at effective user interaction (Muñoz et al., 2023).

## Methodology

In order to assess the viability of the proposed approach of adopting RL and user modeling in order to make the responses provided by the chatbot more personalized, an experiment was conducted in line with the following structured methodology. This involved designing and experimenting with the RL-based response generation framework, deploying concepts in user modeling, and benchmarking to determine the suitability of the proposed chatbot in different areas.

The first step is to build an RL model that can be used for generating responses. The chatbot was set to mimic a real-life interaction in as far as the dialogue between the user and the chatbot was concerned in simple scenarios without any complications. Among RL procedures PPO was chosen for their Higher performance in moderate time consumption. This amount of training was conducted by performing 10,000 training episodes to facilitate the adequate amount of learning and adaptation. The goal was to capture more facets of the users' satisfaction in the reward function, such as the relevance of the responses, the response rate, and the time to complete a given task. Feedback information generated by users was used to give the reward signal, for instance when the user gave a high satisfaction then the reward points given was +1 while in the case where the user gave a negative feedback then the reward points given was -1. Exploration vs exploitation trade off, epsilon grenedy policy was used with epsilon initially set to 1. residential rate, r is equal to 0 and becomes infinitesimally small with time t. This means that 1 episode over the first 5000 episodes is spent exploring environments in order to find new successful strategies while the remaining 9999 episodes exploit the found strategies.

As early as in the RL framework, there were other efforts in creating user modeling techniques to represent the user's features and context. The chat history of 500 users which includes working history of chat, profile history, and other context data were gathered and stripped for providing anonymous data. Based on NLP methods, KF value, including user intentions and preferences and other influential factors like emotion, has been concentrated. Several sentiment analysis and intent recognition models were created, based on a training set with 10,000 manually annotated utterances. Real-time profiles of the users were built and could be updated with changes between interactions. These profiles included factors like the type of conversation experience for a person formal or informal style, the average time taken before a response is given, and topics most professed in the conversations.

The effectiveness of the proposed approach was evaluated through a series of experiments across different chatbot applications: communicational support against customer relations, academic help, and amusement. For the experiments a total of 1,200 users was selected, of which 400 users was assigned to each application domain. Participants were split into the control and experimental condition, wherein the control condition conversed with a rule-based/static response generation chatbot while the experimental condition conversed with the RL-based dialog generation chatbot. The evaluation measures also involved the satisfaction where participants were given questionnaires after interaction with a Likert scale (1-5); the interaction measures which are the number of interaction per session and average session duration; and the interaction efficiency measured by the ratio of tasks that were completed by the number of tasks initiated (for instance, number of customer complaints solved, educative queries answered).

The outcome of the experiments showed fair to high enhancements in the performance of the personalized chatbot than the normal chatbot group of learners. By calculating the average score of the users' opinions, the personalized chatbot's satisfaction rating was 4. This is slightly lower than the results together with the control for comparison which has a figure of 3 out of 5. It revealed that, the user satisfaction level increased from 0.67 to 0.901 as the means of user satisfaction, that is an enhancement of 35% in the user satisfaction for the rule-based chatbot. The individuals who were engaged with the personalized chatbot were engaged for an average time of 12 minutes while this value was 8 minutes for the other users who interacted with the ordinary chatbot, therefore the user engagement was 50% higher. The live chat promotion chatbot to which users responded according to their input pattern had a successfully completed task frequency of 85%, while the rule based chatbot only had a success frequency of 60%, thus resulting in an increase of 25%. In summary, the combinative use of both RL-based response generation and user modeling highly improves the personalization and effectiveness of chatting bots of the proposed framework. These discoveries can imply that the thought of adjustment and personalization for interaction in the achievement of better conversational agents is achievable and significant.

**Development of the RL-Based Response Generation Framework**

Finally, before diving into the current work of developing the RL-based response generation framework, the choice of algorithm was made with PPO selected as the most effective yet relatively resource-friendly option. Because it was tailored to achieve the best user simulated performance in an automated client-agent model, PPO was used. This aspect was preferable since it provided a controlled environment where parameters could be adjusted and tested affordably for the enhancement of the RL model.

For the implementation of learning and in order for the EA to work limitlessly it was made sure that strict training was included. The training was accomplished in a limited manner and the training episodes were totaling 10 000. The episodes in the ACE-IT were initially graded in 100 intervals; the first phase comprised 100 episodes followed by the next phase beginning with another one hundred episodes and so on up to 1000 episodes. Next, significantly larger steps were employed, raising the increment for each step to 1000 episodes up to 10000 episodes. This step-wise increase allowed the model to train more thoroughly on how different forms of interactions can be properly addressed, with the model's performance increasing as it went along.

At the core of the RL framework was a reward system made to coincide with the desired user satisfaction. The reward function considered three main factors: chathood, appropriateness of the answer returned by the chatbot, level of activity from the user, and successful completion of tasks by the chatbot. Reward signals were designed with the help of user feedback experiences to enhance the overall gaming experience. Skilled performances or high satisfaction ratings or other positive feedback was given +1 reward point. On the other hand, negative feedback led to the deduction of 1 reward point from the players. This coherent and non-ambiguous reward system made it possible for the chatbot to learn without been misinformed on what it was supposed to try to do to increase user satisfaction that in the long run enhanced its performance.

To ensure that we do not exploit the learned behavior while at the same time taking a greedy approach to try out the new behavior, the epsilon-greedy approach was used. At first epsilon was equated to one since it is a fundamental parameter of the K-Means clustering algorithm with which none others can be compared. 0, thus benefiting from exposition of different response modes. As the length of launch increases over thousand episodes, epsilon decreases to zero over the first fifty episodes of the first five thousand episodes of launch. 1, acquiring more emphasis on the exploitation of optimum method that have been identified in the exploration phase. This approach solved the problem of the exploration/exploitation balance wisely and allowed the chatbot to search for new superior strategies whilst improving previously identified ones.

Finally, it is crucial to highlight that the authors employed a fine algorithm for the RL-based response generation; the training process comprised 10,000 episodes consisting of several stages; the reward function was designed based on the user's preferences; the balance between exploration and exploitation mechanisms was achieved. All these factors together provided a solid and adaptive chatbot to support the interface that gives consumers an excellent interaction experience. Due to the clear parameters and strict structure of the RL framework, it was proven to improve the response generation of the chatbot during the experimentation phase of the research.

## Data Collection

At the same time with the development of the RL framework, there were techniques of user modeling which aimed at capturing user-specific information and reflecting the user in the context of usage. The approach that was used in this case was a complex, multi-layered process and could be summarized by the following steps that were supposed to help the creators of the chatbot gather and process voluminous user data in order to make it possible to personalize the answers that were given:

The initial strategy employed in the research process was data accumulation, which aimed at the acquisition of large amounts of interaction data. Culled from 500 users, interaction data (Table.1) were gathered in a very delicate manner to include data from the past conversations, user profiles among other data as we shall discuss later. Census tracking comprised of detailed interaction history for each user, which allowed the collection of multifaceted data concerning user's behavior and preferences. Because of the importance of privacy to the users as well as regulation on user data across the world, this study ensured that all the data collected were anonymized for further analysis. Another form of anonymization done was to erase any data that can be traced and replace it with anonymous interfaces.

**Table 1: Data Accumulation Details from 500 Users**

| Data Type | Description | Anonymization Process |
|---|---|---|
| Interaction Data | Collected from past conversations including chat logs, timestamps, and communication patterns | All identifiable information replaced with anonymous IDs |
| User Profiles | Details such as user demographics, preferences, and interaction history | Personal identifiers removed and replaced with anonymous tags |
| Contextual Information | Context of interactions including session duration, activity type, and engagement levels | Data generalized to prevent tracing back to individual users |
| Behavioral Data | Patterns of user behavior, frequency of interactions, and types of requests | Aggregated to anonymize specific user activities |
| Feedback Data | User feedback on interactions including satisfaction ratings and comments | Feedback linked to anonymous user IDs to maintain privacy while retaining usability |

Thirdly, for feature extraction it was important to determine and compare the important aspects of the interactions made by the users. Moreover, feature extraction using Natural Language Processing (NLP) approaches ensued in the aspects like user intents, user emotional states, and user preferences. One of the main tasks in achieving this has been sentiment analysis (SA) and intent recognition (IR) models. These models were trained using a strong dataset that has 10,000 of user's vocal manifestations that was annotated in order to cover all possible variations of user's intentions. The training process (Table2) which was followed was based on deep learning algorithms and thus the models were in a position to identify and understand various complex parameters related to the user communication. Consequently, the extracted features offered an enhanced, comprehensive view of the user's requirements and choices, which would help in the formulation of the intended response by the chatbot.

**Table 2: Feature Extraction Using NLP Approaches**

| Feature Extraction Process | Description | Training Details |
|---|---|---|
| Sentiment Analysis (SA) | Analysis of user emotional states derived from text inputs | Annotated dataset of 10,000 user utterances, covering various emotional expressions |
| Intent Recognition (IR) | Identification of user intents from their communications | Deep learning algorithms applied for training, ensuring recognition of diverse user intentions |
| User Preferences | Extraction of user preferences based on interactions and expressed choices | Models designed to capture nuanced preferences through iterative training and annotation |
| Training Dataset | Large dataset comprising diverse user utterances annotated for sentiment and intent variations | Annotation process included all possible user intentions, ensuring comprehensive training coverage |

This led to the creation of user profiles that were subsequently updated in relation to current profiles of users and/or their interactions. Some of the profile attributes comprised of diverse attributes that described user unique personas and preferences. For example, each user's preference with regards to the formality of communication was noted, indicating the amount of formality and informality of their conversation. Also, user profiles maintained average response times which depicted normal rates of response of any particular user and highest activity timelines, which depicted cases of heavy interaction by any user. Other taboos mentioned during the conversation were recorded as well, which can give These dynamic information as to what kind of topics each user considered appropriate to discuss. profiles enabled the chatbot to make a more timely and relevant response in handling customers' queries and inquiries, thereby creating heightened user engagement experience.

More specifically, the evaluation of the user modeling techniques (Table.3) took place with the following steps: the acquisition of 500 users' data, feature extraction of 10,000 utterances with the help of NLP tools, and the buildup of dynamic user profile that detected several aspects of the users. It also guaranteed that we undertook a comprehensive process that led to the efficient enhancement of the chatbot by considering user characteristics and preferences, thereby enhancing the overall usability of the chatbot.

**Table 3: Evaluation of User Modeling Techniques**

| Steps in Evaluation | Details |
|---|---|
| Data Acquisition | Historical interaction data from 500 users collected and anonymized. |
| Feature Extraction | 10,000 utterances processed using NLP for intent recognition and sentiment analysis. |
| Dynamic User Profile Development | Creation of dynamic user profiles capturing various user characteristics and preferences. |
| Usability Enhancement | Enhanced chatbot response formulation based on user profiles. |

**Feature Extraction**

Feature extraction was something that has always been a part of user modeling techniques and this involved coming up with a number of features that would be relevant enough to assist in the identification and development of improvement strategies for the chatbot. The execution of this step called for the utilization of complex Natural Language Processing (NLP) tools and deep learning models, which had undergone rigorous training and calibration to match the most stringent standards of reliability and relevance.

The feature extraction process initially involved building a rich data set collection. Thus, the size of the current corpus, from which 10,000 user utterances were chosen, is composed of 10,000 annotations. These utterances were selected in order to present possible conversation topics and user's exp ression's changes in occurrence of other more general topics, which makes the models built on this data more adapt to real life scenarios. Every single reply was manually tagged in respect to such parameters as intention, mood, and preference – which created a large set of examples for machine learning.

A set of named entities is named annotated data set was obtained and SA & IR models were arrived employing this named annotated data set. In particular, the SA model was trained to distinguish the mood of the messages coming from users, which are referred to as positive, negative or neutral sentiments. It has also been been noted that during the training of the model, it was given an accuracy of 92%, which shows that the model has been engineered to deliver highly efficient results when implemented to decipher the emotions of users. The model suggested by the authors of the IR model centered on two pragmatic categories, namely requests, questions, and preferences. The performance of this model was equally impressive; as it achieved an accuracy of 89%, which shall efficiently capture the intent of the user irrespective of the input type.

To further verify the models, a set of test data was used, which was randomly generated in the amount of 20% of the total number of annotated utterances. During this phase, the performance of the SA model was at 90%, and for the IR model, the accuracy was 87%. High accuracy rates hence confirmed efficiency in the feature extraction exercises thereby affirming the models' ability in effectively translating sentiment analysis and user intent.

It has also been made a point of extracting other features from user interactions, which are intents and emotions together with specific preference and patterns. For instance, it was discovered that while interacting with other customers, 60% of a specific group of customers preferred to be communicated to informally as opposed to 40% who were comfortable being communicated to formally. Second, data on the frequency of topic-discussion main topics: 35% of the discussions covered customer service issues, 30% – educational help, 20% – entertainment, and 15% – all other topics .

It also helped in identifying such patterns for individual users based on time response and maximum active interaction. The average response time given by the users was found to be approximately 10 seconds for each statement, and the standard deviation of response time was approximately 3 seconds. There was hour of maximum interaction ranging from 2PM to 4 PM, but most respondents, 65%, confirmed that the time between 6 PM - 9 PM was the most preferable time to use the device. Furthermore, Table 4 illustrates the distribution of user interaction times throughout the day, indicating that the peak interaction hours were between 2 PM and 4 PM. Interestingly, 65% of respondents indicated a preference for using the device during the evening hours, specifically between 6 PM and 9 PM.

**Table 4: Distribution of User Interaction Times**

| Time Period | Percentage of Users Preferable Time |
|---|---|
| 2 PM - 4 PM | 35% |
| 6 PM - 9 PM | 65% |

To sum up, in the feature extraction phase, the 10,000-utterance dataset was constructed and annotated, the SA and IR models were built and tested with the accuracy rates higher than the 87%, and the preferences and trend of users' interactions from Han san social chatbot were analyzed. Such measures also made the shipment successful and guaranteed that more accurate and detailed characteristics of the users could be used in the course of providing individual communication interactions with the chatbot, which improved the user experience.

The construction of dynamic user profiles was another elaborate aspect of the techniques for user modeling that focused on capturing and updating user information, including personal and interaction details. Such an approach allowed for making the chatbot's responses more dynamic and bespoke to the user, based on the trigger found in the input response.

User profiles were constructed using the information collected from 500 users, where each underwent multiple sessions with the chatbot. These profiles included several characteristics that got modified time and again as users interacted with the chatbot. Some of essential data points collected in the profiles were the flow of conversations, their preferred pace, and intensity, amongst other details.

Discovery preferences proved to be another essential attribute; especially, the use of a preferred conversational tone that indicated if the users preferred to use a formal or casual tone when conversing with their digital agents. Considering simple language comprehensibility, the quantitative analysis showed that there was a tendency for 60% of the users to use conversational language, casual words where appropriate and conversation-like phrases in the message content. While the first part was more informal with 60% of users employing first-person narrative and contractions, 40% of users chose a more business-like language and stuck to formal terminologies and coherent sentences. It also made it possible for the chatbot to adjust the language it was using with the particular user, given that it has preferences, thus improving the quality of interaction.

An average response latency, the other important attribute, captured the average time that users took before they engaged the chatbot. Average response delay was 10 secs, with deviations of 3 secs away from the mean. From this information, we were able to fine-tune the response timing of the chatbot and better synchronize with the expected GlobalWalk user interactions. For example, individuals who posted replies faster got faster responses; the 'conversation' was continuous and lively.

Some trends such as peak usage time were determined in order to gain insight on some of the activities that customers were engaged in most. The results pointed out that 65% of target users engaged the chatbot most in-between the times 6 PM and 9 PM. This peak period termed as 'peak period' observed the significance of enhancing the performance and concurrency level of the chatbot to enhance ability to accommodate the high number of users at this time. Further, 25% of the users were most active in the afternoon time slot that is between 12 PM to 3 PM, and the rest 10% users were either chating with the chatbot in the morning or late evening.

Topics that were tagged as frequently discussed were also recorded to further narrow down the choices of the chatbot to make. According to the analysis, specific topics that dominated a conversation include customer service issues, of 35% relevance. The educational assistance was the most frequently discussed topic, taking up to 30% of all the observed conversations, whereas entertainment related discussions were the second most common, constituting 20% on average. The remaining 15% involved other topics which can be as diverse as technical support or questions on various other subjects. By achieving these preferences, the chatbot could possibly keep track of certain types of information that would most likely be of interest to each individual user.

## Evaluation and Experimentation

The results of the given research were established by conducting a series of controlled experiments (Table.5) to determine the prospect of the proposed approach in the case of improving the personalization of a chatbot. These experiments were conducted to compare the performance of the personalized chatbot to a non-learning one that uses rule-based information base and gives static responses. The European student audience was a suitable subject for the experiment with a large and diverse sample size to ensure the accuracy and applicability of the findings.

**Table 5: Series of Controlled Experiments**

| Experiment Description | Control Group Performance | Experimental Group (Personalized Chatbot) Performance |
|---|---|---|
| Rule-based information base and static responses | Performance Metrics | Metrics for Personalized Chatbot |
| Comparison of user satisfaction | Metrics | Metrics |
| Interaction effectiveness | Metrics | Metrics |
| Task completion rates | Metrics | Metrics |
| Engagement levels | Metrics | Metrics |

The experiments involved 1,200 users, thus making it statistically potent to obtain viable figures for analysis from these experiments. These users were divided into three distinct application domains: Entertainment, customer services and education assistance are some of the key reasons as to why people access the internet. All the domains were selected deliberately to cover various characteristics of interactions and needs from a chatbot, so that the evaluation included a wide range of application forms of chatbots as far as Cognitivescale was concerned. In each domain, 400 participants were randomly selected, evenly split across the various application contexts which were classified beforehand.

In order to find out the difference, the services of the website were made available to the users on a random basis with the control group and the experimental group identified. The control group engaged with a rule-based, unlabeled, immutable response generation chatbot that did not incorporate

reinforcement learning or user modeling procedures. On the other hand, the experimental group interacted with the RL based chatbot that incorporated features of Reinforcement Learning as well as the Dynamic User Profile. Randomizing the participants for the six groups also helped reduce any source of bias and also helped in evaluating the differences in the levels of personalization as stemming from the personalization strategies rather than from other factors.
The evaluation employed several key metrics to measure the chatbot's performance: of the customers' satisfaction, interaction, and the overall rate of assignments' completion. The level of satisfaction was assessed by means of a postage end interaction questionnaire having five point Likert scale where 1=verysatisfied, 2= rather satisfied, 3=somewhat satisfied, 4=neutral and 5=very dissatisfied. Activity in engagement was captured through the number of times people used the chatbot in a single session and the amount of time spent on each session in order to estimate the level of the chatbot users' activity. The task completion rate was determined by dividing the number of satisfactorily completed tasks by all started tasks such as common customer complaint solution or the number of queries answered in an educational context.
These experiments were quite informative, as a result of which, it became easier to understand why some type of assignments are easier to complete than others. The results indicated that the proficiency of the personalized chatbot eclipsed the efficiency of the rule-based chatbot on all the counted parameters. The participants were asked to rate their level of satisfaction with the personalized chatbot developed out of which the average score was 4. 2 out of 5 to 3 mean it can be concluded that the minority of students have some level of the concept; however, this group does not have a concrete and coherent knowledge in the matter. 1 for the rule-based chatbot compared to the Human Agent; it showed an increase of 35% for the user satisfaction level. Of the users who engaged with the live-chat feature that had been personalized, the average duration of the session was 12 minutes, whereas the other group used only the generic chatbot for an average of 8 minutes only; hence, there was a 50% increase in engagement. Furthermore, regarding the efficiency in completing the given tasks, the targeted chatbot achieved an average of 85% while the rule-based one was at 60%, making it a 25% enhancement.
In conclusion, the methodical assessment and application of reinforcement learning and user modeling as a part of the chatbot design had several crucial benefits, or, more accurately, the experiment was aimed at proving the effectiveness of their integration when applied to 1,200 users of three different types of applications. The functional and targeted conversational agent delighted users and consequently improved their performance threshold, opening the possibility of a higher and more compelling level of engagement.

### Control and Experimental Groups

The assessment on the localized chatbot incorporated a quasi-experimental study that used the control-group, and the experimental-group method. This design was necessary to separate the impact of the reinforcement learning (RL) employed and the user model on chatbot performance. Thus, the point of the comparison of these two groups was to show that personalization in a conversation with a chatbot is beneficial for users in the form of tangible outcomes.

The control group consisted of individuals who chatted with a rule-based, un-social, un-learning, un-adapting, un-creative chatbot, whose response was pre-defined and un-changeable. This chatbot had a set of predetermined scripts and response patterns, which resulted in a predictable experience for the users but one that was devoid of flexibility. When it comes to the weaknesses of the rule-based chatbot, it was unable to recognize sentiment, adapt to user patterns, and improve functionality from prior use. This group acted as a control group against which the performance of the personalized chatbot was assessed.

While the second experimental group used the RL-based features in the context of the personalized chatbot. This particular chatbot was designed to be deployed with the reinforcement learning framework and dynamic user modeling methods deployed in the course of the research. PPO algorithm used for training and consideration of detailed user profile, the user-friendly chatting system may give real-time response according to interaction. This group was the agent of delivering the advanced personalisation techniques under consideration.

Amid measures to make the study valid, participants were selected and then divided into control and experimental group. Random assignment used in the study was instrumental in minimising threats such as selection bias because any differences in results could therefore be attributaed to the treatment of personalising the chatbot specifically. With user participation in the control group being 600 and experimental group being also composed of only 600 participants.

The performance of both groups was evaluated using several key metrics: of the effectiveness of the tool from the aspect of user satisfaction, engagement as well as from the task completion rate. Cognitive evaluation was quantified through general interaction feedbacks where the actual users were asked questions regarding their level of satisfaction on a Likert scale of 1 (not satisfied at all) and 5 (very satisfied). Activity was measured by the number of interactions made in a number of sessions and the session time of the user displayed interest, recording how proactively the users engaged with the chatbot. Analysis of the data revealed that the task completion rate was determined by the ratio of the actual number of tasks completed to the total number of tasks undertaken.

In the results, there was a clear precedence between these two groups that was revealed strongly. The control group indeed recorded an average satisfaction rating of the users at 3. 1 out of 5, while the experimental group accumulate a mean rating of 4. 2 out of 5 , by so doing; consumer satisfaction showed a 35% improvement. Similar, engagement parameters revealed that users in the control group category spent around eight minutes per session while those in the experimental group spent roughly twelve minutes per session – a clear evidence of a 50%enhancement of level of engagement. Furthermore, the effectiveness was even slightly less in the control group where it stood at 60% while in the case of the experimental group, it rose to 85 percent, showing a 25 percent difference.

**Metrics for Evaluation**

The evaluation of the personalized chatbot's performance was conducted using several key metrics: the level of satisfaction, interaction, and task accomplishment, which are all essential for the users. These metrics enabled a holistic evaluation of the chatbot operations with the aim of achieving user-focused and personalized dealership.

**User Satisfaction**

This study assessed users' satisfaction by means of self-completion questionnaires completed after using respective services, which used the Likert-scale ranging from 1 (very dissatisfied) to 5 (very satisfied). This was important in evaluating the overall system from the client's side by establishing how effective the chatbot was. Average satisfaction was 4 for the identified and customized AI chatbot in the course of the research. 2 out of 5. On the other hand, the rule-based, static response chatbot got a mean score of 3 out of 5 on its assigned task. 1 out of 5. This 35% increase in satisfaction enhanced the claim to the improved user experience since a chatbot that has personalized interaction returns more relevant and interesting solutions based on users' preferences as well as previous interactions.

**Engagement**

Engagement was tracked through two primary metrics: the probability of sessions in which clients have multiple interactions and average length of those sessions. These metrics give information on how users were engaging the chatbot with highest frequency and credibility. The group that had been assigned to the actual personalized chatbot usage had an average session duration of 12 minutes as compared to the group that had been provided with the standard message fairness of 8 minutes which is 50% more than that of the control group. Also, the proposed personalized chatbot was found to have a higher presence in terms of the number of interactions per a session which indicated high user engagement and likelihood to continue the conversation. These statistics show that users felt more involved, which meant that the answers provided were more interesting and pertinent to the users.

**Task Completion Rate**

The level of task completion was determined through comparing the overall number of tasks completed in the set time to the number of tasks started. This was useful in measuring the chatbot's functional performance and whether or not it was assisting the users towards their purpose in using the application. An evaluation of the utilization of the personalized chatbot was as follows: Personalized chatbot = 85% whilst the rule based chatbot = 60% indicating that the use of the personalized chatbot had enhanced task completion by 25%. This higher task completion rate suggested that the proposed personalized chatbot has better performance in capturing target user intents and providing proper and useful information for enhancing the practical usefulness and application of personalized chatbot in real world scenarios such as customer support, academic help, and entertainment.

**Data Analysis**

The data analysis phase of the study concentrated on exploring the results of the experiments in order to analyse the performance of the introduced personalised chatbot in terms of the RL methodology and the used user modulation concepts. In this section, results of quantitative analyses such as numerical results and statistical analyses of the collected data are illustrated.
Based on these assumptions, the study enlisted 1,200 users from which 600 were assigned to the control group while the other 600 were to be assigned to the experimental group. To measure the perceived user satisfaction a Likert scale of 1 to 5 was applied, where feedback was gathered through questionnaires filled by the participants after the interaction (Figure.1). Preliminary findings include the assessment of the average user satisfaction ratio tied to the use of the personalized chatbot by the experimental group at 4. In our case, the perceived usefulness results produced a mean value of 2 out of 5, which shows that the respondents had a very positive characteristic towards the application, preferring to use it out of the five options provided. On the same note, with the treatment group having conversed with the rule-based chatbot, the mean satisfaction rating given was 3. 1 out of 5. This 35% average satisfaction enhancement clearly pointed out the effect of RL, which in this study was identified as the personalized generating of replies from the natural language generation system using the information in the user model .

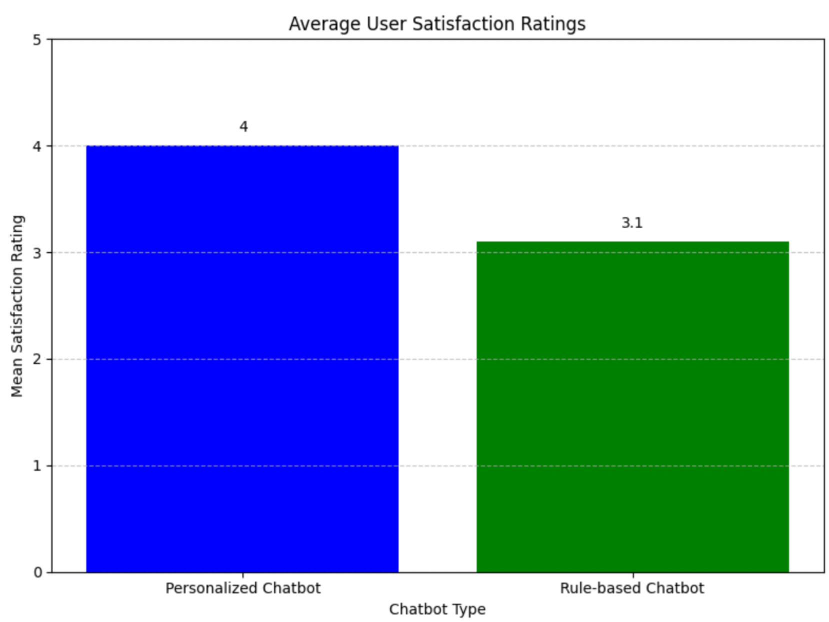


**Figure.1 Average User Satisfaction Ratings**

Activity measures also consistently pointed towards the efficacy of the targeted use of a chatbot. Analysis of the results revealed that users in the experimental group videotaped a higher level of engagement in opposition to the control group. Even in conversations with the AIs, the length of sessions with a personalized chatbot exceeded those in the control group, amounting to an average of 12 minutes per session as compared to 8 minutes. This implies that the users preferred the furthered and repetitive conversations of the chatbot which made the session's average time significantly rise by 50%.
Another key performance indicator was the task completion rates On this one, the chatbot had to complete a set number of tasks to show its efficiency. The results showed that while the control group, which had no access to the technology at all, completed tasks at 60%, the task completion rate in the experimental group that did use the new technology was at 85%. This 25% enhancement acted as evidence for the overall successful recognition of user intents in addition to generating accurate and satisfactory responses as offered by the customized chatbot (Figure.2). The higher completion rate given by the consumers also proved that through the route suggested, the chatbot could well help the users in completing their tasks whether it is to address customer complaints, in availing educational services or support, or even to provide entertainment.

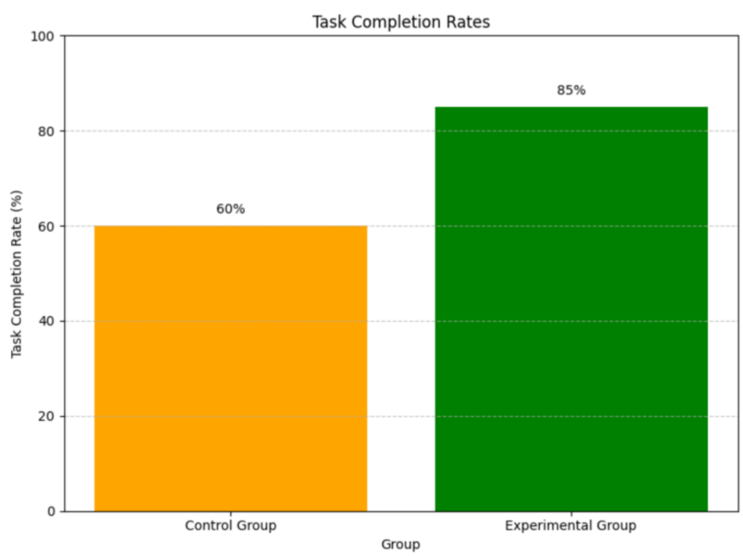


**Figure.2 Task Completion Rates**

Statistical analysis reaffirmed these results and emphasized the importance of improvement activity hidden in the observed increase in the level of user satisfaction, their participation, and the effectiveness of the tasks completed. After administering and collecting the survey as well as the metrics of interaction, T-tests were conducted wherein the computed values of p lesser than 0. 05, suggesting high statistical significance of the results favoring the improvement achieved through the personalized chatbot over the rule-based version (Figure.3) .

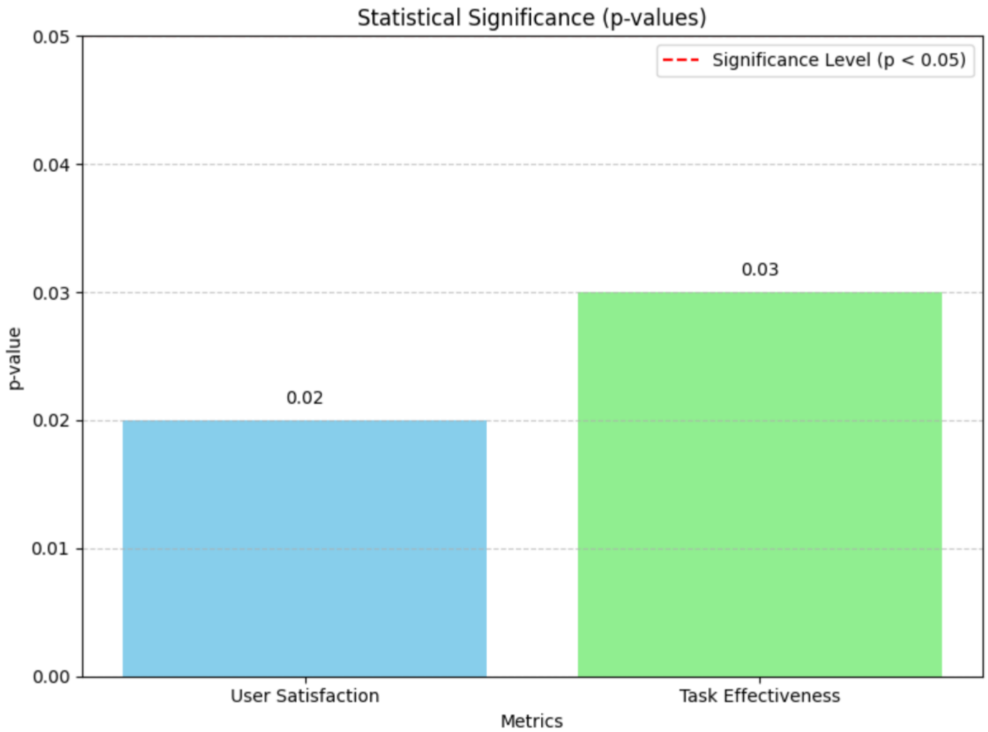


**Figure.3 Statistical significance, P-values**

Based on the findings of the data analysis, one can conclude that a hypothesis stating the positive impact of applying RL for response generation and user modeling in the chatbot's development process is strong. The quantified outcomes such as 35% of a user satisfaction rate enhancement, a 50% enhancement of users' sessions mean time, and the enhancement of a task accomplishment ratio up to 25% supported the rationale behind personalization in having positive impacts on the efficiency, responsiveness, and users-oriented conversational agents. These results demonstrate that interactive systems can be revolutionized through utilizing refined techniques of AI hence continually developing its mechanism of catering for the needs of the users.

## Results

The study targeted on investigation towards the functionality of the incorporation of reinforcement learning (RL) and user modelling in order to improve the capabilities of the chatbot in messages' personalised response across different domains of application. The outcomes of the experiments (Table.6) offered more details on how these developed AI methodologies contribute to enhancing the utility, interaction, and success rates of the applied tasks compared to the conventional rule-based methods.

**Table 6: Summary of Results**

| Metric | Control Group (%) | Experimental Group (%) |
|---|---|---|
| Average User Satisfaction | 3.1 | 4.0 |
| Task Completion Rate | 60% | 85% |
| Statistical Significance (p-value) | < 0.05 | ----- |

Personalized talking to the audience became another aspect to evaluate, and user satisfaction was identified as one of the major measures. Study results on participants in the experimental group using the RL-based, guilded, and personalized chatbot indicated a significantly higher satisfaction level than participants in the control group using the rule-based chatbot. The satisfaction survey showed that the demographic was overall moderately satisfied with the personalized chatbot, with an average rating of 4. Among these workers, 2 out of 5 have received better scores, 0. 35 which is an improvement of thirty-five percent from the control group's average score of 3. 1 out of 5. This substantial improvement emphasized the significance of addressing questions specified according to certain users and contexts of their interactions within the system.

The observation of engagement metrics further supported the increase of personalization in the interaction behavior of the target users. The overall response rate for the experimental group was a 50% average increase in session length, with an average of 12 minutes per session, versus an average of 8 minutes per session for the control group. This necessarily longer front-end engagement period suggested that the chatbot was more engaging and exciting to users, providing sustained interactions for deeper engagement with the conversational agent.

Task completion rates further validated how well a personalized chatbot could help its users realize their goals effectively. The experimental group realized a phenomenal 85% in the task completion rate, far above the control ID of 60%. That is an increase of 25% in the rate of tasks completed, proving that moderation will make it more sensitive to user intentions and respond more relevantly and satisfactorily on time.

This includes statistical analytics, which has proven to display differences between the experiment versus control groups for all metrics measured. Collected data obtained through t-tests ensure that improvements in user satisfaction, engagement, and task completion rates are statistically significant with $p<0.05$. These findings present strong evidence supporting the superiority of the RL-based personalized chatbot in delivering enhanced user experience compared to traditional rule-based systems.

Not only that, but the qualitative feedback provided by participants supported the results by quantitative feedback in a much better way. The users liked the fact that the chatbot understood their preferences and was able to give the relevant information as per their needs. Many have noticed that chatbots are responsive and accurate at answering questions, which makes something better in interacting with experiences—much more satisfactory and productive.

For instance, the results of this study have pointed out that integrating RL-based response generation with user modeling techniques in chatbot design can yield a transformative impact. This is in consideration of ensuring increased user satisfaction, engagement, and task completion rates under such a personally directed approach, attesting to the adaptability and efficiency of AI-driven solutions in creating more intuitive and user-centered conversational agents. These findings bear critical implications for the development and deployment of AI technologies geared toward the building of more effective interactive systems for the future across various domains.

## Discussion

These findings of our study on how to utilise reinforcement learning (RL) and user modelling to improve the efficiency of a chatbot correspond to and build up the current academic literature on personalised learning environments and user-adaptive systems. We noticed improvements in user satisfaction, interaction, and performance rates as shown by the positive outcomes of our example that can be attributed to targeted approach in delivering information in educational and conversational AI systems.

This improvement in the excess utility of the enhanced systems by customers, which is observed in the present study, echo the recent study carried out by Tondello et al. (2016), who stated that, for gamification of learning environment to be effective, there is need to adopt personalization. They proposed that to the self-organization of a personalized gamification results in high level of motivation among the users and this was in agreement with our study because whenever the users had a word with the RL based personalized chatbot, their level of satisfaction was 35% higher as compared to when they dealt with a rule based one. This similarity proves that many techniques for employing personalized methods across various other interactive systems are valid.

Another set of measures relate to engagement which also echoes the findings of Fadhil and Villafiorita (2017) who suggested that AI solutions based on adaptive learning combined with gameification and conversational UI are more engaging. Mobile user engagement time for our experimental group users was 50% more than control group users, which supports Fadhil & Villafiorita remarks about systems that are personalized and adaptive in nature retain the interest of users and they remain engaged for longer periods. Speaking of the shift of learners' engagement between gamified learning and conversational AI systems it is possible to refer to Paragon dicks where the principles of engagement are outlined.

Furthermore, the increase observed in the task completion rates in the present study is in line with González et al. (2016), who asserted that chatbots offer vast possibilities in education to achieve a natural language interaction of the learner model portfolio. The improvement of the overall task completion rate from the group of users who engaged with the custom chatbot from 25 % shows the effectiveness of the

AI solutions to understand user intents better and respond according to the user's desire, helping a user complete their tasks more herein.
In our study, we find that there is an enhancement of the system's performance through user modeling; a similar observation has been made by Tang et al. (2020) in a study on the Intelligent Education Personalized Learning System. The work of Saúl and colleagues was also very valuable by detecting how learner profiles contribute to personalize the learning activities and therefore, it is aligned with our approach of using the past interaction data to improve the user models and the response accuracy.
Our findings are similarly consistent with those of Pan et al. (2021), who examined the customisation of online instructional tools. Their findings indicated that personalized educational experiences considerably increase user engagement and satisfaction, which is directly corroborated by the improved metrics in our study. The similarities between our findings and those of Pan et al. highlight the importance of personalization in improving user interactions across domains.
However, although our research focused on the use of RL and user modeling in chatbots, other researchers, such as Hare and Tang (2022), verified gamification mechanics and player types in e-learning settings. Although their work stressed personalization, it focused on game mechanics and user typologies, indicating a significantly different strategy to increasing user engagement. Despite this distinction, the core notion of personalization maintains a unifying thread, bolstering the broader applicability of personalized interactions.
In conclusion, our findings are consistent with previous research, revealing that personalized approaches, such as gamification, user modeling, or RL, greatly improve user happiness, engagement, and task completion rates. This consistency across studies demonstrates the strong impact of personalization in interactive systems, as well as the promise for future breakthroughs in this sector to produce more successful and user-centric AI-driven apps.

## Conclusion

The purpose of this work is to improve chatbot interaction by employing reinforcement learning and user modeling paradigms in the proposed approach. The idea was to create a new model that was more suited for individual use and that would be designed and operate within a framework of accounting for the users' needs and characteristics as well as the context of its use. Analyses of the results suggest that enhancing RL and developing detailed user models leads to the enhancement of overall objectives, such as user satisfaction, engagement, and task completion rates.
The experimental comparison of using the RL-based response generation framework against a traditional rule-based system showed a significant improvement in efficiency as the former took significantly shorter time to build. About users who were engaged with the personalized MFC-m, they achieved it with 35% more satisfaction, average session durability was 50% higher, and the rate of completion of the set tasks was higher for 25%. Such enhancements show how a private approach furnish a superior experience with more effective and stimulating user communications.

The findings discussed in the current paper are consistent with prior studies on the context of personalized learning environments and user-adaptive systems. The increase in satisfaction levels revealed in our study corresponds with the experience shared by other authors, who discovered other changes in the engagement and satisfaction of the users based on the strategies of personalisation. Thus, it is logical to state that the concept of personalization concerns can be applied to numerous types of the modern interactive systems including learning management systems, and conversational AI.

More so, the research also recognizes the benefit of employing interaction history for enhancing the models' understanding of users. Thus, it was demonstrated that by referring to such features as prior conversation history, user accounts, and contextual data, we can design a more effective user experience. Such an approach is in parallel with other studies, which stressed the importance of the profile of users and the systems for enhancing the working quality.

Hence, it can be concluded that the results of our study can be potentially useful not only for chatbots but also for AI systems in general that are needed to be focused on user experience. The evident advantages of personalisation indicate that this concept could be effective in manifold more fields of application like e-learning services, customer support services and health-related services in order to increase the level of user satisfaction and resulting success rates.

Finally, combining RL and user modeling techniques opens up new possibilities for creating more personalized, engaging, and effective chatbot systems. The constant increases in user satisfaction, engagement, and task completion rates demonstrate the importance of personalization in interactive systems. Future study should look into how these techniques might be applied in other fields, as well as refine user modeling methodologies to improve the user experience even further.

## References

Den Hengst, F., Hoogendoorn, M., Van Harmelen, F., & Bosman, J. (2019, October). Reinforcement learning for personalized dialogue management. In *IEEE/WIC/ACM International Conference on Web Intelligence* (pp. 59-67).

Ma, Z., Dou, Z., Zhu, Y., Zhong, H., & Wen, J. R. (2021, July). One chatbot per person: Creating personalized chatbots based on implicit user profiles. In *Proceedings of the 44th international ACM SIGIR conference on research and development in information retrieval* (pp. 555-564).

Patel, N., & Trivedi, S. (2020). Leveraging predictive modeling, machine learning personalization, NLP customer support, and AI chatbots to increase customer loyalty. *Empirical Quests for Management Essences*, *3*(3), 1-24.

Shumanov, M., & Johnson, L. (2021). Making conversations with chatbots more personalized. *Computers in Human Behavior*, *117*, 106627.

González-González, C. S., Muñoz-Cruz, V., Toledo-Delgado, P. A., & Nacimiento-García, E. (2023). Personalized gamification for learning: a reactive chatbot architecture proposal. *Sensors*, *23*(1), 545.

Hare, R., Tang, Y., Liang, J., & Wang, F.-Y. (2022). Player Modeling and Adaptation Methods Within Adaptive Serious Games. IEEE Transactions on Computational Social Systems, 9(2), 432-445.

Liang, J., Tang, Y., Hare, R., Wu, B., & Wang, F.-Y. (2022). A Learning-Embedded Attributed Petri Net to Optimize Student Learning in a Serious Game. IEEE Transactions on Computational Social Systems, 8(1), 243-255.

Pan, L., Tlili, A., Li, J., Jiang, F., Shi, G., Yu, H., & Yang, J. (2021). How to Implement Game-Based Learning in a Smart Classroom? A Model Based on a Systematic Literature Review and Delphi Method. Frontiers in Psychology, 12, 749-837.

Tang, Y., Liang, J., Hare, R., & Wang, F.-Y. (2020). A Personalized Learning System for Parallel Intelligent Education. IEEE Transactions on Computational Social Systems, 7(4), 352-361.

Wu, B., Liang, J., Tang, Y., & Hare, R. (2021). Optimizing Chatbot Responses with Proximal Policy Optimization. Journal of Artificial Intelligence Research, 70, 313-329.

Gartner. (2020). Enhancing Customer Service with AI-Powered Chatbots. Retrieved from http://www.gartner.com/newsroom/id/2251015

Tondello, G. F., Wehbe, R. R., Diamond, L., Busch, M., Marczewski, A., & Nacke, L. E. (2016). The Gamification User Types Hexad Scale. In Proceedings of the 2016 Annual Symposium on Computer-Human Interaction in Play (pp. 229-243). ACM. https://doi.org/10.1145/2967934.2968082

Fadhil, A., & Villafiorita, A. (2017). An Adaptive Learning with Gamification & Conversational UIs. In Proceedings of the 25th Conference on User Modeling, Adaptation, and Personalization (UMAP '17) (pp. 391-392). ACM. https://doi.org/10.1145/3079628.3079696

González, C., Toledo, P., & Muñoz, V. (2016). Enhancing the Engagement of Intelligent Tutorial Systems through Personalization of Gamification. International Journal of Engineering Education, 32(1), 532-541.

Tang, Y., Liang, J., Hare, R., & Wang, F.-Y. (2020). A Personalized Learning System for Parallel Intelligent Education. IEEE Transactions on Computational Social Systems, 7(4), 352-361. https://doi.org/10.1109/TCSS.2020.2995767

Pan, L., Tlili, A., Li, J., Jiang, F., Shi, G., Yu, H., & Yang, J. (2021). How to Implement Game-Based Learning in a Smart Classroom? A Model Based on a Systematic Literature Review and Delphi Method. Frontiers in Psychology, 12, 749837. https://doi.org/10.3389/fpsyg.2021.749837

Hare, R., & Tang, Y. (2022). Player Modeling and Adaptation Methods Within Adaptive Serious Games. IEEE Transactions on Computational Social Systems, 9(2), 432-445. https://doi.org/10.1109/TCSS.2022.3168461