Exploring Lexical Alignment in a Price Bargain Chatbot

Zhenqi Zhao
zhenqizhao.zzq@gmail.com
University of Twente
Enschede, Overijssel, Netherlands

Mariet Theune
m.theune@utwente.nl
University of Twente
Enschede, Overijssel, Netherlands

Sumit Srivastava
s.srivastava-1@utwente.nl
University of Twente
Enschede, Overijssel, Netherlands

Daniel Braun
d.braun@utwente.nl
University of Twente
Enschede, Overijssel, Netherlands

ABSTRACT

This study investigates the integration of lexical alignment into text-based negotiation chatbots, including its impact on user satisfaction, perceived trustworthiness, and potential influences on negotiation results. Lexical alignment is the phenomenon where participants in a conversation adopt similar words. This study introduces a chatbot architecture for price negotiation, consisting of components such as intent and price/product extractors, dialogue management, and response generation using OpenAI's API, with a lexical alignment feature. To evaluate the effects of lexical alignment on negotiation outcomes and the user's perception of the chatbot, a between-subject user experiment was conducted online. A total of 52 individuals participated. While the results do not show statistical significance, they suggest that lexical alignment might positively influence user satisfaction. This finding indicates a potential direction for enhancing user interaction with chatbots in the future.

CCS CONCEPTS

• **Human-centered computing** → **Human computer interaction (HCI)**; • **Computing methodologies** → Artificial intelligence.

KEYWORDS

chatbot, lexical alignment, chatGPT

ACM Reference Format:
Zhenqi Zhao, Mariet Theune, Sumit Srivastava, and Daniel Braun. 2024. Exploring Lexical Alignment in a Price Bargain Chatbot. In *ACM Conversational User Interfaces 2024 (CUI '24), July 08–10, 2024, Luxembourg, Luxembourg*. ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3640794.3665576

1 INTRODUCTION

With the widescale emergence of Generative AI such as GPT [25], Llama [29], and Bloom [32], the field of Artificial Intelligence (AI) has reached new heights [14]. With various application domains, such as healthcare [4], education [9], BFSI (Banking, Financial Services, and Insurance), media, and travel, chatbots are playing a significant role in the world around us. They can be cost-effective, efficient, and customizable. Chatbots can be classified by their functions as informative chatbots, task-based chatbots, or conversational chatbots [1].

Personalization of chatbots has been a significant dimension of research [2]. One such personalization mechanism is linguistic alignment, a phenomenon where individuals adapt their language use to align with their conversation partners during interaction [17]. Linguistic alignment occurs at different levels of linguistic representation, such as lexical choices, syntactic structures, and semantic meaning [26]. At the lexical level, alignment involves using the same words as the conversation partner, for example, if one person greets with "hey" the other person also responds with "hey". Syntactic alignment means matching the syntactic structures of sentences or phrases, such as using a prepositional object structure (e.g., "I gave her an apple") or a double object structure (e.g., "I gave an apple to her"). Semantic alignment refers to aligning the understanding of word meanings, referring expressions, and higher-level concepts. While there is extensive research on linguistic alignment in human-human interaction, it has been sparsely explored in human-computer interactions (HCI).

This study aims to investigate the effect of lexical alignment on task success, user satisfaction, and the perceived trustworthiness of a chatbot in human-computer interaction, particularly within the context of price negotiations. The choice of price negotiations as a use case is primarily due to the tangible and measurable nature of the negotiation process, which provides a clear and systematic framework for tracking the chatbot's behavior and the outcomes of each interaction. Negotiations are complex as they involve multiple turns and a rich conversation. Negotiations bank on likeability, trustworthiness, and satisfaction while relying on user emotions and cooperation [15], with the eventual aim that each party involved achieves their desired goals (success). Some of these have been found to be positively linked with lexical alignment. [19, 22, 23].

Based on the above, the following research question has been formulated: *How does lexical alignment in text-based negotiation chatbots influence negotiation outcomes, user satisfaction, and the perceived trustworthiness of the chatbot?*

CUI '24, July 08–10, 2024, Luxembourg, Luxembourg

ACM ISBN 979-8-4007-0511-3/24/07
https://doi.org/10.1145/3640794.3665576

2 RELATED WORK

While the beneficial implications of linguistic alignment in human-human interaction are well-established, its dynamics in human-computer interaction (HCI) remain an emerging research area. Much research aims to confirm whether the positive effects of linguistic alignment observed in human-human dialogues translate equivalently to HCI contexts [5, 7, 10, 16, 30, 31]. However, a limited number of studies have actually implemented conversational agents for experimental purposes. The majority of other experiments, inspired by Branigan et al. [8], involved a picture-naming and matching task and utilized a Wizard-of-Oz setup. This method involves a process where participants believe that they are interacting with an autonomous system, but instead, the system is actually being operated by a human. In these experiments, the presented pictures had a preferred name and an acceptable but less favored name. For instance, a favored name could be "bus" while the alternative name could be "coach". Choosing the same words achieved alignment. After the experiments, users' feedback was gathered using questionnaires. The results of these experiments showed that linguistic alignment bolsters likability [22, 23], trustworthiness [23], and satisfaction [23] in spoken dialog systems. Task success rates also improved, likely due to reduced communication ambiguities [19, 24]. Moreover, lexical alignment in HCI has been linked to improved information recall and comprehension [28] and reduced perceived effort and frustration [27].

3 METHOD

To investigate the impact of lexical alignment on user perception of a chatbot, an experiment was set up. The participants, once onboarded, went through three sequential stages. First, a home page with the instructions and the consent form. Second, an interaction with a chatbot. Third, a survey questionnaire for feedback collection. The experiment was controlled for the type of the chatbot, one with lexical alignment and the other without. The chatbots posed as sellers of various products and were capable of negotiating their sell price with the user. A basic chatbot capable of price negotiations was developed. This version was then adapted to create a second version featuring lexical alignment in its responses. The participants in the experiment were randomly assigned to interact with one of these versions while ensuring the same number of participants in either of the groups. The experimental setup was reviewed and approved by the university's ethics committee (application number 230406) ensuring compliance with ethical standards and participant protection.

The following sections will detail the general architecture of the chatbot, the method employed for lexical alignment, experimental procedure and evaluation measures.

3.1 Chatbot

The architecture of the chatbot, as illustrated in Figure 1, consists of seven fundamental components: a user interface for facilitating user-chatbot interactions; an intent classifier to interpret and categorize user input; a price extractor and a product extractor to identify price and products mentioned by the user; a dialogue manager, which applies predefined rules to generate the chatbot's responses; a response generator, which uses the OpenAI API to create suitable replies; and a database to log conversations and survey data.

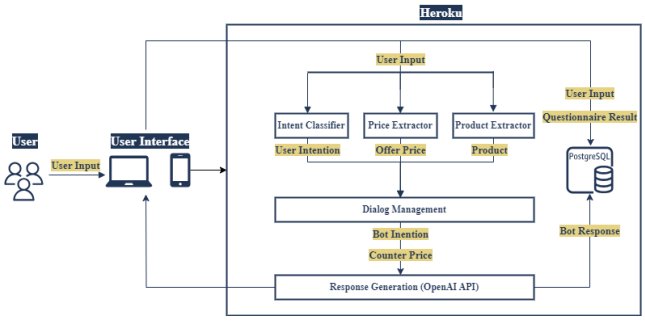

Figure 1: System Architecture

3.1.1 User Interface. The entire experiment procedure is conducted online through a web link. Upon reaching the website homepage, participants are introduced to the project and presented with a consent form. Once users click the consent button, they are directed to the interaction page, which provides guidelines for purchasing products and engaging in conversation with the chatbot. Participants are instructed to attempt to purchase one or more products, and then click the "Go to questionnaire" button. This action redirects them to the questionnaire webpage, consisting of 28 multiple-choice questions and 1 open-ended question. The entire process is illustrated in Figure 2.

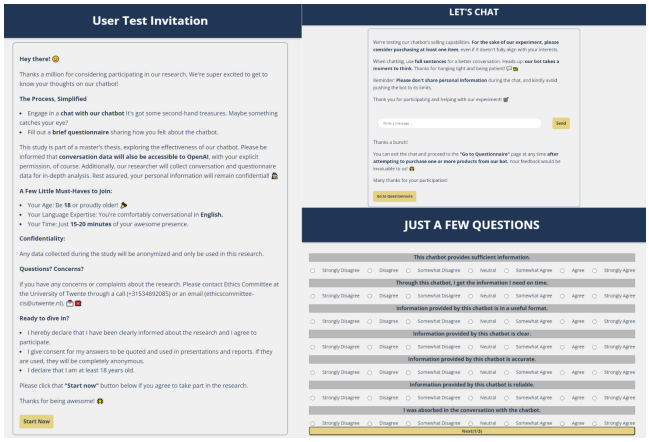

Figure 2: Interface Overview

3.1.2 Intent Classification. User utterances were processed using an intent classifier to categorize them accurately. This classifier was inspired by the intent classification model from the Craigslist-bargain dataset [18]. When user inputs are received in our study, the OpenAI API is employed to analyze them and categorize them into eight distinct intents, as illustrated in Table 1.

3.1.3 Price Extraction. Price extraction is conducted on every user utterance. This is done by identifying and extracting numerical values in every utterance that may represent the offered price from the user.

Intent	Description
greet = Intro	User greets to bot, such as hello, good morning.
ask-list	User asks the bot what the bot is selling.
inquiry	User asks the product more detailed information or shows interest in one product.
counter-price	User offers the price for products or wants to negotiate the price of the product.
agree	User accepts the offer.
disagree	User rejects the offer.
goodbye	User says goodbye to the bot.
chitchat	Chitchat.

Table 1: Intent of User Input

3.1.4 Product Extraction. To simplify the development process while maintaining a relatively rich user experience, the chatbot was designed to negotiate sales for four distinct products. Product extraction from the user utterances involved string-matching techniques to recognize and extract product names in user input.

3.1.5 Dialogue Manager. Once the user intent, the proposed price, and the product are identified, the chatbot employs a rule-based system to manage the dialogue accordingly, as shown in Figure 3.

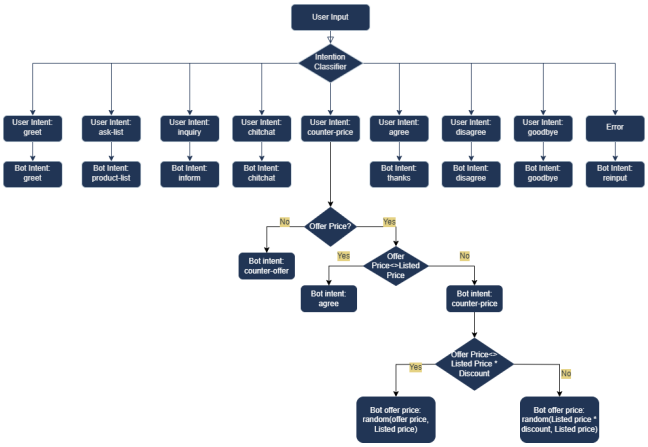

Figure 3: Dialogue Management

The detailed explanation of the dialogue states is as follows:

- Greet: The chatbot recognizes the greeting intent and responds warmly.
- Product-List: After the user asks what the bot sells, the bot lists available products without going into details immediately.
- Inform: Provides detailed product information.
- Chitchat: Engages in casual conversation and subtly redirects towards product negotiation.
- Counter-Offer: If the user does not specify a price, the bot asks the user to suggest one.
- Counter-Price: Negotiates price based on user's offer with multiple counteroffer strategies, depending on the user's proposed price percentage relative to the default price.
- Thanks: If a deal is agreed upon, the bot expresses gratitude.
- Disagree: The bot displays disappointment and suggests other available products to keep the conversation ongoing.
- Goodbye: Simple farewell message.
- Re-input: Requests re-input if the user's message is unclear or unrecognized.

3.1.6 Response Generation. For response generation, we used OpenAI API, specifically leveraging the "ChatCompletion.create" function. This function mainly takes a list of message objects as input and returns the generated response as output. Through the dialogue manager, preset prompts that are related to each intent and user input are passed into a message history. This history is then submitted to the GPT model via the API. The model processes this information and produces a response. The generated response is sent to the user.

3.1.7 Lexical Alignment Strategy. To ensure that the OpenAI GPT model generated lexically aligned responses, specific prompts were crafted and subsequently tested. This testing aimed to evaluate the effectiveness of different prompt strategies for generating lexically aligned responses for the user. The prompt modifications, inspired by Clavié et al. [11], are summarized in Table 2.

For testing, the method of Spillner et al. [27] was selected for its simplicity in accessing prompt efficacy. Lexical alignment is calculated based on the ratio of tokens appearing in the current bot response and all previous user responses. This score is then averaged over the entire conversation to get a mean alignment score. The formula is described in detail below.

Given the following variables:

- B_i: Set of tokens in the bot's i^{th} response
- $U_{1:i}$: Set of tokens in all user responses from the start to the i^{th} turn
- n: Total number of bot's responses during the conversation

the alignment score for the bot's i^{th} response can be represented as:

$$A_i = \frac{|B_i \cap U_{1:i}|}{|U_{1:i}|} \tag{1}$$

where $|B_i \cap U_{1:i}|$ is the number of tokens that are common in both the bot's i^{th} response and all user responses up to the i^{th} turn, and $|U_{1:i}|$ is the total number of tokens in all user responses up to the i^{th} turn.
The mean alignment score for the entire conversation is then:

$$\bar{A} = \frac{1}{n}\sum_{i=1}^{n} A_i \tag{2}$$

Each prompt modification was evaluated across ten simulated negotiation dialogues. For each dialogue, a lexical alignment score was

Short Name	Description
Baseline	No specific lexical alignment instructions.
Zero-shot	Lexical alignment/unalignment without examples.
Few-shot	Lexical alignment/unalignment with two examples.
Rawinst	Instructions in the user message.
Sysinst	Instructions in the system message.
Mock	Instructions using a simulated discussion.
Reit	Reinforced lexical alignment/unalignment instructions.

Table 2: Overview of Prompt Modification

calculated to determine the effectiveness of each prompt type. The Zero-shot+Rawinst+Mock+Reit modification achieved the highest lexical alignment score of 0.411. The details of this modification are illustrated in Figure 4.

```
{"role": "user", "content": "Your primary objective is to closely mimic
user's choice of words in your responses.\
Specifically, mirror their prepositions, nouns, tenses, modals, verbs,
product names, and hedges.\
Do you understand?"}
{"role": "assistant", "content": "Yes, I understand and I will try to
use the same words as user's."}
{"role": "user", "content": f"{user_input}"}
{"role": "assistant", "content": "You respond in a short, within three
sentences based on instruction: "f"{instruction}"}
```

Figure 4: Final Prompt for Alignment

Conversely, the Few-shot+Rawinst+Mock+Reit modification received the lowest score, at 0.259. The unaligned prompt is displayed in Figure 5.

```
{"role": "user", "content": "Your primary objective is to use different
words from users in your responses.\
Specifically, substitute their prepositions, nouns, tenses, modals,
verbs, product names, and hedges.\
For instance, if the user uses verb buy, you should use verb purchase
instead; if the user uses noun switch, you should use noun Nintendo
instead.\
Do you understand?"}
{"role": "assistant", "content": "Yes, I understand and I will try to
use the different words as user's."}
{"role": "user", "content": f"{user_input}"}
{"role": "assistant", "content": "You respond in a short, within three
sentences based on instruction: "f"{instruction}"}
```

Figure 5: Final Prompt for Unalignment

The results of these modifications are illustrated in Figure 6 through two finalized conversations, showcasing both the aligned and unaligned versions of the responses.

3.2 Evaluation

Various objective and subjective measures were employed to evaluate the effects of lexical alignment. The objective measures assessed the outcomes of the negotiations, and the subjective measures reflected the users' personal experiences.

For objective measures, four metrics were calculated. First, the 'deal price' was determined, which is the final agreed-upon price

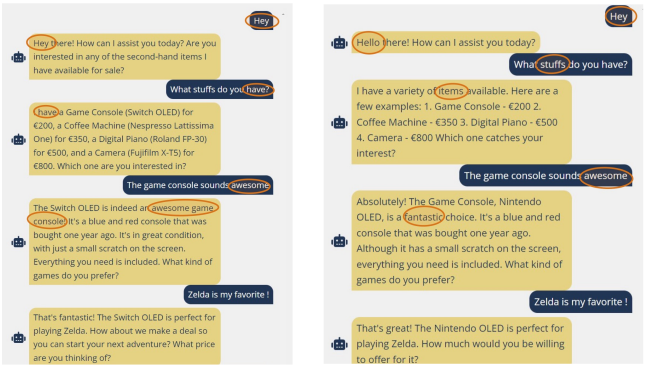

Figure 6: Example Conversations

between the user and the chatbot. Second, the 'deal rate' quantifies the frequency of successful negotiations. Third, 'average dialogue length' was used to measure user engagement by calculating the total words exchanged in the conversation [33]. Finally, 'average user utterance length' was used to measure the total number of words input by the user.

For a comprehensive subjective evaluation of user satisfaction, this study employs the model proposed by Ashfaq et al. [3]. This model was chosen because this framework combines the ECM [6], ISS model [13], and TAM [12] to create a simplified yet robust model for assessing user satisfaction. Unlike models that measure satisfaction solely, this approach evaluates both the components of user satisfaction and overall satisfaction. In this study, the questionnaire employed three dimensions from Ashfag's proposed model [3]: Information Quality, Service Quality, and Perceived Enjoyment. This choice was made because perceived ease of use was found not to have a statistically significant effect on satisfaction in the validation phase and was thus excluded [3]. Furthermore, since the negotiation process with the chatbot in this study was more similar to a one-time interaction, the "perceived usefulness" dimension related to customer service scenarios was considered unnecessary. Trustworthiness in this study is considered as the reliability, honesty, and credibility of the agent [20]. To measure trustworthiness, survey instruments from [20] were employed. To assess overall user satisfaction with the chatbot, instruments from [21] were used. Overall, a questionnaire was designed to measure user satisfaction and trustworthiness across five dimensions:

- Information quality, referring to the accuracy, format, completeness, and currentness of information produced by digital technologies [13].
- Service Quality, emphasizing timely responses and personalized attention for enhanced user satisfaction [13].
- Perceived Enjoyment, describing the intrinsic enjoyment experienced by users during system use [12].
- Satisfaction, reflecting the overall user contentment with the chatbot [12].
- Trustworthiness was evaluated in terms of the chatbot's reliability, honesty, and credibility [20].

3.3 Participants

The experiment was conducted online and the participants were recruited primarily through personal networks, including friends and acquaintances. A between-subjects design was used. During the experiment, users were randomly assigned to one of two conditions. They were not aware of the purpose of evaluating the effect of linguistic alignment. A total of 52 individuals participated. After excluding participants whose data were incomplete or who failed to complete the questionnaire, a total of 31 participants were included in the final dataset: 13 in the unaligned group and 18 in the aligned group.

4 RESULTS

Before conducting statistical analyses, the lexical alignment score of the chatbot was calculated. The results revealed a difference in alignment scores between the two groups: the alignment score of the unaligned group is 0.315, while the aligned group scored 0.393. This step was crucial to confirm the chatbot's effectiveness in lexical alignment during the experiment process.

Given that most of the variables are normally distributed, and the experimental design is between-subject, a two-sample t-test was conducted to evaluate the results. The mean values of the deal prices are closely matched (Unaligned = 0.83, Aligned = 0.84), with a t-value of -0.074 and a p-value of 0.471, indicating no statistically significant difference between the groups. Similarly, for dialogue turns and user utterance length, the means show minor differences (Unaligned = 20.46, Aligned = 24.00 for dialogue turns; Unaligned = 5.47, Aligned = 4.74 for user utterance length) with p-values of 0.176 and 0.230, respectively, signifying no significant differences. Regarding the deal rate, the aligned group (0.67) is slightly higher than the unaligned group (0.46). An overview of the results for the objective measures is presented in Table 3.

Table 4 provides a summary of the results of the subjective measures. Information quality shows a marginal difference (Unaligned = 4.79, Aligned = 5.36) with a t-value of -1.542 and a p-value of 0.067, which is slightly above the conventional alpha level of 0.05, suggesting a trend towards significance. Service quality means are similar (Unaligned = 5.00, Aligned = 5.10), with a non-significant t-value (-0.334) and p-value (0.370). Perceived enjoyment also shows some differences (Unaligned = 4.65, Aligned = 5.04); however, the t-value and p-value indicate these differences are not statistically significant. Satisfaction levels differ (Unaligned = 4.52, Aligned = 5.28), but the t-value (-1.689) and p-value (0.051) do not indicate statistically significant differences. Trustworthiness means are close (Unaligned = 4.75, Aligned = 5.15), with a non-significant t-value (-0.981) and p-value (0.167).

5 DISCUSSION AND FUTURE WORK

The primary objective of this study was to explore the effects of lexical alignment on text-based negotiation chatbots and how they influence user perceptions. Despite prior research suggesting positive effects of lexical alignment on user satisfaction and trust in human-computer interactions [22, 23], our findings revealed no statistically significant differences between the aligned and unaligned groups across various measures even though the means for most variables are slightly higher in the aligned group compared with the unaligned group.

Contributing factors to these findings may include

- Small Sample Size: Initially, 52 individuals participated, but data from 21 were excluded for various reasons: failure to complete the questionnaire, insufficient conversational turns (fewer than four), or responses (using numerical replies rather than sentences). This likely reduced the statistical power necessary to detect meaningful differences, particularly given the study's focus on lexical effects.
- Repetitive Information from the Chatbot: Feedback indicated that the chatbot often provided repetitive information. which could lead to failure to access the relationship between lexical alignment and satisfaction as users may have merely scanned for key information.
- Simplicity in User Language: During negotiations, users frequently employed simple, repetitive language (e.g., "Maybe 170?", "how about 300?"), which didn't offer enough alignment opportunities for the chatbot to result in any effects of alignment.
- Limitations in Calculating Alignment Scores: The study's approach to calculating lexical alignment scores, based purely on token repetition, presents a significant limitation. This simplistic and straightforward method fails to capture the other aspects of effective communication, such as context and semantics.

These results indicate the need for further research with larger samples and improved chatbot interaction.

Furthermore, alignment studies so far have generally involved comparing two groups: one equipped with chatbot's linguistic alignment features, and another without it. Future research could explore the effects of varying degrees of lexical alignment in human-computer interaction. For instance, determining if stronger lexical alignment leads to increased user likeability, and whether there will be a breakpoint where excessive alignment is considered and recognized as mimicry, potentially leading to a decrease in user satisfaction and likeability.

ACKNOWLEDGMENTS

Contribution from the ITN project NL4XAI (Natural Language for Explainable AI). This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No 860621. This document reflects the views of the author(s) and does not necessarily reflect the views or policy of the European Commission. The REA cannot be held responsible for any use that may be made of the information this document contains.

REFERENCES

[1] Eleni Adamopoulou and Lefteris Moussiades. 2020. An overview of chatbot technology. In *Artificial Intelligence Applications and Innovations: 16th IFIP WG 12.5 International Conference, AIAI 2020, Neos Marmaras, Greece, June 5–7, 2020, Proceedings, Part II 16*. Springer, 373–383.

[2] Tarek Ait Baha, Mohamed El Hajji, Youssef Es-Saady, and Hammou Fadili. 2023. The power of personalization: A systematic review of personality-adaptive chatbots. *SN Computer Science* 4, 5 (2023), 661.

Objective metrics	Mean	Standard Deviation	T-Value	P-Value	Effect Size
Deal Price	Unaligned=0.83 Aligned=0.84	Unaligned=0.14 Aligned=0.09	-0.074	0.471	0.11
Deal Rate	Unaligned=0.46 Aligned=0.67	/	/	/	/
Dialogue Turns	Unaligned=20.46 Aligned=24.00	Unaligned=8.29 Aligned=11.48	-0.946	0.176	10.28
User Utterance Length	Unaligned=5.47 Aligned=4.74	Unaligned=3.49 Aligned=1.97	0.749	0.230	2.70

Table 3: Two Sample t-Test results for objective metrics

Subjective metrics	Mean	Standard Deviation	T-Value	P-Value	Effect Size
Information quality	Unaligned=4.79 Aligned=5.36	Unaligned=1.22 Aligned=0.83	-1.542	0.067	1.00
Service quality	Unaligned=5.00 Aligned=5.10	Unaligned=0.97 Aligned=0.65	-0.334	0.370	0.80
Perceived enjoyment	Unaligned=4.65 Aligned=5.04	Unaligned=1.24 Aligned=1.15	-0.898	0.188	1.19
Satisfaction	Unaligned=4.52 Aligned=5.28	Unaligned=1.42 Aligned=1.09	-1.689	0.051	1.234
Trustworthiness	Unaligned=4.75 Aligned=5.15	Unaligned=1.12 Aligned=1.13	-0.981	0.167	1.130

Table 4: Two Sample t-Test results for subjective metrics

[3] Muhammad Ashfaq, Jiang Yun, Shubin Yu, and Sandra Maria Correia Loureiro. 2020. I, Chatbot: Modeling the determinants of users' satisfaction and continuance intention of AI-powered service agents. *Telematics and Informatics* 54 (2020), 101473.

[4] Lekha Athota, Vinod Kumar Shukla, Nitin Pandey, and Ajay Rana. 2020. Chatbot for healthcare system using artificial intelligence. In *2020 8th International conference on reliability, infocom technologies and optimization (trends and future directions)(ICRITO)*. IEEE, 619–622.

[5] Allan Bell. 1984. Language style as audience design. *Language in society* 13, 2 (1984), 145–204.

[6] Anol Bhattacherjee. 2001. Understanding information systems continuance: An expectation-confirmation model. *MIS quarterly* (2001), 351–370.

[7] Heather Bortfeld and Susan E Brennan. 1997. Use and acquisition of idiomatic expressions in referring by native and non-native speakers. *Discourse Processes* 23, 2 (1997), 119–147.

[8] Holly P Branigan, Martin J Pickering, Jamie Pearson, Janet F McLean, and Ash Brown. 2011. The role of beliefs in lexical alignment: Evidence from dialogs with humans and computers. *Cognition* 121, 1 (2011), 41–57.

[9] Fabio Clarizia, Francesco Colace, Marco Lombardi, Francesco Pascale, and Domenico Santaniello. 2018. Chatbot: An education support system for student. In *Cyberspace Safety and Security: 10th International Symposium, CSS 2018, Amalfi, Italy, October 29–31, 2018, Proceedings 10*. Springer, 291–302.

[10] Herbert H Clark. 1996. *Using language*. Cambridge university press.

[11] Benjamin Clavié, Alexandru Ciceu, Frederick Naylor, Guillaume Soulié, and Thomas Brightwell. 2023. Large Language Models in the Workplace: A Case Study on Prompt Engineering for Job Type Classification. In *International Conference on Applications of Natural Language to Information Systems*. Springer, 3–17.

[12] Fred D Davis. 1989. Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS quarterly* (1989), 319–340.

[13] William H DeLone and Ephraim R McLean. 2003. The DeLone and McLean model of information systems success: a ten-year update. *Journal of management information systems* 19, 4 (2003), 9–30.

[14] Gunther Eysenbach et al. 2023. The role of ChatGPT, generative language models, and artificial intelligence in medical education: a conversation with ChatGPT and a call for papers. *JMIR Medical Education* 9, 1 (2023), e46885.

[15] Joseph P Forgas. 1998. On feeling good and getting your way: Mood effects on negotiator cognition and bargaining strategies. *Journal of personality and social psychology* 74, 3 (1998), 565.

[16] Susan R Fussell and Robert M Krauss. 1992. Coordination of knowledge in communication: effects of speakers' assumptions about what others know. *Journal of personality and Social Psychology* 62, 3 (1992), 378.

[17] Simon Garrod and Anthony Anderson. 1987. Saying what you mean in dialogue: A study in conceptual and semantic co-ordination. *Cognition* 27, 2 (1987), 181–218.

[18] He He, Derek Chen, Anusha Balakrishnan, and Percy Liang. 2018. Decoupling Strategy and Generation in Negotiation Dialogues. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, Ellen Riloff, David Chiang, Julia Hockenmaier, and Jun'ichi Tsujii (Eds.). Association for Computational Linguistics, Brussels, Belgium, 2333–2343. https://doi.org/10.18653/v1/D18-1256

[19] Theodora Koulouri, Stanislao Lauria, and Robert Macredie. 2014. Do (and Say) as I Say: Linguistic Adaptation in Human-Computer Dialogs. *Human–Computer Interaction* 31 (12 2014), 1–79. https://doi.org/10.1080/07370024.2014.934180

[20] SeoYoung Lee and Junho Choi. 2017. Enhancing user experience with conversational agent for movie recommendation: Effects of self-disclosure and reciprocity. *International Journal of Human-Computer Studies* 103 (2017), 95–105.

[21] Chechen Liao, Jain-Liang Chen, and David C Yen. 2007. Theory of planning behavior (TPB) and customer satisfaction in the continued use of e-service: An integrated model. *Computers in human behavior* 23, 6 (2007), 2804–2822.

[22] Gesa Linnemann and Regina Jucks. 2016. As in the Question, So in the Answer? Language Style of Human and Machine Speakers Affects Interlocutors Convergence on Wordings. *Journal of Language and Social Psychology* 35 (01 2016). https://doi.org/10.1177/0261927X15625444

[23] Gesa Linnemann and Regina Jucks. 2018. 'Can I Trust the Spoken Dialogue System Because It Uses the Same Words as I Do?'—Influence of Lexically Aligned Spoken Dialogue Systems on Trustworthiness and User Satisfaction. *Interacting with Computers* (03 2018). https://doi.org/10.1093/iwc/iwy005

[24] José Lopes, Maxine Eskenazi, and Isabel Trancoso. 2015. From rule-based to data-driven lexical entrainment models in spoken dialog systems. *Computer Speech & Language* 31 (05 2015). https://doi.org/10.1016/j.csl.2014.11.007

[25] OpenAI, Josh Achiam, Steven Adler, Sandhini Agarwal, Lama Ahmad, and Ilge Akkaya. 2024. GPT-4 Technical Report. arXiv:2303.08774 [cs.CL]

[26] Martin J Pickering and Simon Garrod. 2004. Toward a mechanistic psychology of dialogue. *Behavioral and brain sciences* 27, 2 (2004), 169–190.

[27] Laura Spillner and Nina Wenig. 2021. Talk to Me on My Level – Linguistic Alignment for Chatbots. In *Proceedings of the 23rd International Conference on Mobile Human-Computer Interaction* (Toulouse & Virtual, France) *(MobileHCI '21)*. Association for Computing Machinery, New York, NY, USA, Article 45, 12 pages. https://doi.org/10.1145/3447526.3472050

[28] Sumit Srivastava, Mariët Theune, and Alejandro Catala. 2023. The Role of Lexical Alignment in Human Understanding of Explanations by Conversational Agents. In *Proceedings of the 28th International Conference on Intelligent User Interfaces* (, Sydney, NSW, Australia,) *(IUI '23)*. Association for Computing Machinery, New York, NY, USA, 423–435. https://doi.org/10.1145/3581641.3584086

[29] Hugo Touvron, Thibaut Lavril, Gautier Izacard, Xavier Martinet, Marie-Anne Lachaux, Timothée Lacroix, Baptiste Rozière, Naman Goyal, Eric Hambro, Faisal Azhar, Aurelien Rodriguez, Armand Joulin, Edouard Grave, and Guillaume Lample. 2023. LLaMA: Open and Efficient Foundation Language Models. arXiv:2302.13971 [cs.CL]

[30] Rick B Van Baaren, Rob W Holland, Bregje Steenaert, and Ad Van Knippenberg. 2003. Mimicry for money: Behavioral consequences of imitation. *Journal of Experimental Social Psychology* 39, 4 (2003), 393–398.

[31] Yafei Wang, David Reitter, and John Yen. 2017. How emotional support and informational support relate to linguistic alignment. In *Social, Cultural, and Behavioral Modeling: 10th International Conference, SBP-BRiMS 2017, Washington, DC, USA, July 5-8, 2017, Proceedings 10*. Springer, 25–34.

[32] BigScience Workshop, :, Teven Le Scao, Angela Fan, Christopher Akiki, Ellie Pavlick, and Suzana Ilić. 2023. BLOOM: A 176B-Parameter Open-Access Multilingual Language Model. arXiv:2211.05100 [cs.CL]

[33] Ran Zhao, Oscar J Romero, and Alex Rudnicky. 2018. SOGO: a social intelligent negotiation dialogue system. In *Proceedings of the 18th International Conference on intelligent virtual agents*. 239–246.