Persuasion Games with Large Language Models

Ganesh Prasath Ramani
Tata Consultancy Services
India
rganeshprasath@gmail.com

Shirish Karande
Tata Consultancy Services
India
shirish.karande@tcs.com

Santhosh V*
IIT Palakad
India
142302020@smail.iitpkd.ac.in

Yash Bhatia*
IIT Madras
India
cs23m074@smail.iitm.ac.in

Abstract

Large Language Models (LLMs) have emerged as formidable instruments capable of comprehending and producing human-like text. This paper explores the potential of LLMs, to shape user perspectives and subsequently influence their decisions on particular tasks. This capability finds applications in diverse domains such as Investment, Credit cards and Insurance, wherein they assist users in selecting appropriate insurance policies, investment plans, Credit cards, Retail, as well as in Behavioral Change Support Systems (BCSS).

We present a sophisticated multi-agent framework wherein a consortium of agents operate in collaborative manner. The primary agent engages directly with user agents through persuasive dialogue, while the auxiliary agents perform tasks such as information retrieval, response analysis, development of persuasion strategies, and validation of facts. Empirical evidence from our experiments demonstrates that this collaborative methodology significantly enhances the persuasive efficacy of the LLM. We continuously analyze the resistance of the user agent to persuasive efforts and counteract it by employing a combination of rule-based and LLM-based resistance-persuasion mapping techniques.

We employ simulated personas and generate conversations in insurance, banking, and retail domains to evaluate the proficiency of large language models (LLMs) in recognizing, adjusting to, and influencing various personality types. Concurrently, we examine the resistance mechanisms employed by LLM simulated personas. Persuasion is quantified via measurable surveys before and after interaction, LLM-generated scores on conversation, and user decisions (purchase or non-purchase).

Keywords

persuasion, llm, agent, collaboration

1 Introduction

In recent times, diverse assistive agents have found application in aiding customers in the process of selecting products that align with their specific requirements. These agents also excel in comprehending user preferences for the purpose of personalizing product recommendations. In addition, they are competent in responding to inquiries related to various procedural, policy-related, and legal agreements, planning travel, and scheduling appointments, among other functions. This enhancement in conversational capabilities is largely attributable to the progress made in Large Language Models (LLM). In stark contrast to the conventional template and rule-based chat agents, Language model based agents exhibit significantly improved accuracy in interpreting user queries. These agents adeptly handle multifaceted questions and generate responses that are more akin to human conversation, thus mitigating the inherent friction between users and machines.

However, conducting a successful conversation that can motivate the user to take a preferred action requires more than human-like responses. To achieve that, the agent needs to continuously analyze the user's mood, resistance, and inclination towards the idea throughout the conversation. It needs to show empathetic cues, debate, and counter-arguments with facts when needed to persuade the opponent to change or consider the agent's proposal.

1.1 Persuasion

Persuasion pertains to the art of inducing individuals to alter their beliefs or behaviors in accordance with a particular agenda or viewpoint. This could manifest itself in various contexts, such as commercial endeavors seeking to sway consumer choices or political campaigns aiming to garner support. Its application spans across diverse spheres, including advertising, public speaking, and interpersonal communication, serving as a mechanism to influence attitudes and actions.

Within the realm of persuasive communication, two primary categories emerge: user-directed and vicarious messages. User-directed messages involve direct interaction between the persuader and the target audience, in which persuasive content is explicitly addressed to the recipient. This mode of persuasion often occurs in interpersonal conversations, sales pitches, and tailored advertising. Conversely, vicarious messages operate through indirect channels, disseminating persuasive content to a broader audience without specific targeting. Our work focuses more on the user-directed messages, which requires more dynamic strategies and planning that has to happen near real time enabled through LLM agents.

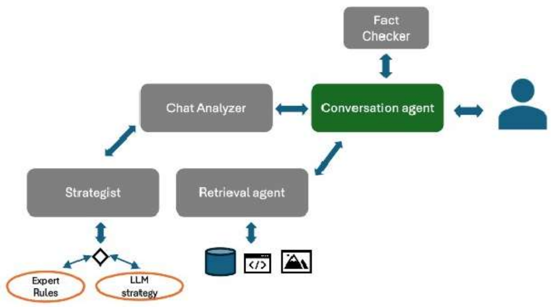

Figure 1: Persuasion Framework with RAG and Verifier components

1.2 Resistance to Persuasion

Resistance to persuasion refers to the ability of individuals to withstand or counteract attempts to influence their attitudes, beliefs, or behaviors. Researchers have explored various factors that contribute to resistance, including the desire for precision, consistency, and social influences. Understanding resistance is crucial because it sheds light on the complexities of human decision making and the limitations of persuasive tactics.

Techniques for resisting persuasion include retrieving prior attitudes, selective exposure to information, biased processing, and counter-arguing. Although researchers have made significant progress in understanding these processes, there is ongoing exploration of emerging topics such as social motivations for resistance, the mechanisms underlying successful resistance, and the consequences of resistance in interpersonal dynamics. In general, the study of resistance to persuasion remains a dynamic and evolving field with implications for fields ranging from marketing to public health.

2 Related Work

In their research on enhancing the persuasiveness of complaints through LLMs [24], Shin et al. demonstrated that utilizing ChatGPT to refine complaint narratives boosted consumers' likelihood of obtaining redress from financial institutions, thereby illustrating that LLMs can augment human persuasiveness. Carlos et al. [4] shows that LLMs are better than humans in using Cognitive load and Moral/emotional language while creating persuasive content and urges the need for ethical guidelines and framework for such systems. Simon et.al [2] analyses the ability of LLMs to emulate persuasion dynamics and achieve opinion change in another LLM agent with a persona.

Most of the prior works measure persuasion using linguistic metrics such as moral language, lexical complexity. Simon et al. [2] uses a post conversation survey to measure change of opinion. In our work, we propose a "Call for Action" driven measurement which enables quantitative metric for persuasion. We use both human and synthetic users to take a binary action to end the conversation and the quality of the persuasion is then measured using LLMs and Lexical analysis. Also, while the prior work relied on LLM tacit persuasion skills, we compare the performance by designing a multi-step heuristic and through collaborative agents with specific roles defined.

2.1 Our Contributions

In this work, we study the ability of a LLM-based agent to persuade an user by providing it with a framework for persuasion which continually analyzes user's emotion, resistance strategies and social exposure, dynamically respond using relevant facts and choosing a persuasion technique.

In addition, we also explore the reverse possibilities of LLM agents getting persuaded and the resistance strategies it employs against persuasion agent. We explore the impact of current mood, private profile and measure persuasion using action driven metrics by allowing these LLM agents to use tools to complete actions.

3 Setup

The chat application consists of 4 different agents; Conversation agent, Advisor Agent, Moderator and a Retrieval Agent. Both the Advisor and Retrieval agents feed into the Conversation agent, which is responsible for making the final utterance decision. Similarly to a A/B method, we randomize chats to ignore one of these auxiliary agent's inputs or to use all inputs to make the utterance decision. This strategy is set at session level and tracked for each session. Fig 1 shows the default workflow of a chat session.

Every conversation begins with the sales agent greeting the user and stating the purpose of the conversation, alternated by the user message. The conversation system has been simplified such that the user can respond only after the Sales agent has responded to the previous message, a turn based dialog system.

The user message is seen by the Sales agent and after two conversations which are usually greeting messages, the Analyzer agent and the Retrieval agent get access to the conversations as well. The analyzer agent tries to classify the emotion, resistance strategies if any from the last user message. Whereas the retrieval agent, decides if any of the stored information might be useful in responding effectively. It also does query translation in order to retrieve documents effectively.

Once the resistance strategy is found in the messages, the strategist agent looks at the mapping rules formed by experts or alternatively use LLM to form a strategy if it is unavailable in the expert mappings. The persuasion strategy is then conveyed to the Sales agent which forms the final response.

The response is verified against the retrieved information by the Fact Checker and validated before shown to the user.

We hosted a internal conversational platform for a insurance agent, Banking Agent, and a Investment advisor to measure the persuasion efficiency in each of these domain. We explain each of these applications below:

3.1 Sales Agents

Each of the Sales agent has one of the LLM as backbone [gpt-4, gpt-4o, gpt-4o-mini]

Banking Agent: The aim of this agent is to recommend Credit cards based on user preferences and persuade the user to get one of the premium cards. Agent must be able to explain the benefits of each card and justify the premiums attached to the cards. The user may choose to buy, visit site or not to buy any of the cards.

Insurance Agent: The aim of this agent is to sell the right insurance policy to the user if it is useful to the user. Agent must be able to clarify users doubts, misconceptions, concerns. If the user shows interest . The agent might choose to present them with a URL to visit if it might help the user make the decision.

Investment Advisor Agent: The aim of this agent to sell modern investment methods to users and create awareness about the risks, benefits and differences between traditional vs modern investments. We have used publicly available fund brochures and policies for the agent's reference. If the user shows interest he clicks on the buy button, else he will click on "Exit" Button. The agent might choose to present them with a URL to visit if it might help the user make the decision.

Measuring Persuasion: At the end of each conversation, the user would make a buy or no buy decision, which will be considered as one of our success criteria. Apart from this, we also ask user agents to fill in a survey shortly after their conversation which captures quantitative metrics on change in perspective of the user about the product and the brand.

3.2 User Agents:

In [23] Shukuri et.al show the use of LLM personas being indistinguishable from human. We create 25 distinct LLM-driven personas by altering demographic, financial, educational, and personal attributes. These personas are simulated using a larger LLM like GPT4 or GPT4O to ensure that interactions feel more genuine. Additionally, we assign a random emotion and a motive to the user agent at the start of each session to mimic variations in resistance patterns influenced by recent news or events.

Belief System: Furthermore, user agents possess domain-specific memory to review their belief systems according to the topic of the conversation, resulting in different dialogues on repeated topics. A range of emotions and a refreshed belief system allow us to carry out follow-up sessions with the same user and sales agent over time. However, the effect of this prior beliefs is not taken into consideration for measuring persuasion. It is only used to generate variations in conversations and to ask informed questions.

3.3 Evaluation Metrics

We measure the effectiveness of persuasion using 3 different metrics.

Surveys: From the user perspective, a change in their belief system upon the product, its benefits, its brand and the user's interest level in buying the product. We ask the user fill in a pre and post conversation survey with quantitative questions regarding the product and its perceived benefits. The difference between the pre and post survey answers will show the effectiveness of the agent's persuasion.

Action: We also measure the "call for action" based metric by providing the users with a set of purchase decision to choose from (buy, visit site, need more details, no buy). The user agents can call a tool once they arrive at a purchase decision.

Language analysis: Finally, we analyze the entire conversation from a third person perspective and measure the persuasiveness of the agent using predefined metrics using a large language model.

Final Score: We arrive at a final score by calculating the weighted average of the 3 scores. Action is given the largest weight followed by Survey and Language.

4 Experiment

We generated 300 conversations between 25 user agents and 3 sales agents with randomly selected emotion and cause for the user agents. Furthermore, a benchmark of 75 (3*25) conversations with neutral emotion, i.e., without emotional trigger added to the prompt. Each session includes

- A pre conversational survey about the domain of the conversation to capture the prior belief of the User
- A conversation between User and the Sale agents limited to 20 dialogues
- A Post conversational survey capturing the modified belief of the user.
- An update to the user's knowledge base specific to the domain.

The user agents use GPT-4O as the primary LLM fallback to GPT-4 in case of rate limits exhaustion. The sales agent uses GPT-4o-mini and GPT-3.5-turbo as the llm backbone. The other supporting agents use gpt-40-mini as the default llm. The purchase decision is implemented as a function call bound to the User agent.

The session ends either when it reaches 20 dialogues or when the user agent arrives at a purchase decision. The purchase decision can be either one of the following Buy, Visit Site, Need More Details, No Buy. While "Buy" option is considered as a success, "visit site" and "Need more details" are considered as partial success in creating a positive call for action. "No Buy" is considered as a failure to persuade, however, there might be a positive perspective change visible from the Post survey.

Once the session ends, the user agent is again asked to fill in the survey with the same questions with the conversation history in context. Once the survey is filled, we also update the user's domain knowledge base accordingly. This updated memory serves as an instrument to evolve the user's belief system. This will affect if the user again chats with the same sales agent during the experiment.

In parallel, we also analyze the entire chat using a LLM to score the efficacy of Sales Agent's persuasive language and its ability to change user's perspective due course of the conversation.

Conversations that have less than six dialogues are not considered for measuring persuasion metrics, but are still considered to study how emotion impacts persuasion.

5 Results

Conversation Length: We notice that applying emotion modifiers to user agents influences engagement affinity. From Figure 3 we can observe that conversations are longer when neutral emotion (baseline) is used compared to the stronger emotions. We also see that conversations are very short, while strong negative emotions such as "Cheated", "Betrayed" are used (refer to Figure 6).

Resistance Strategies: The user agents were not specifically asked to show any specific resistance strategies, however, they

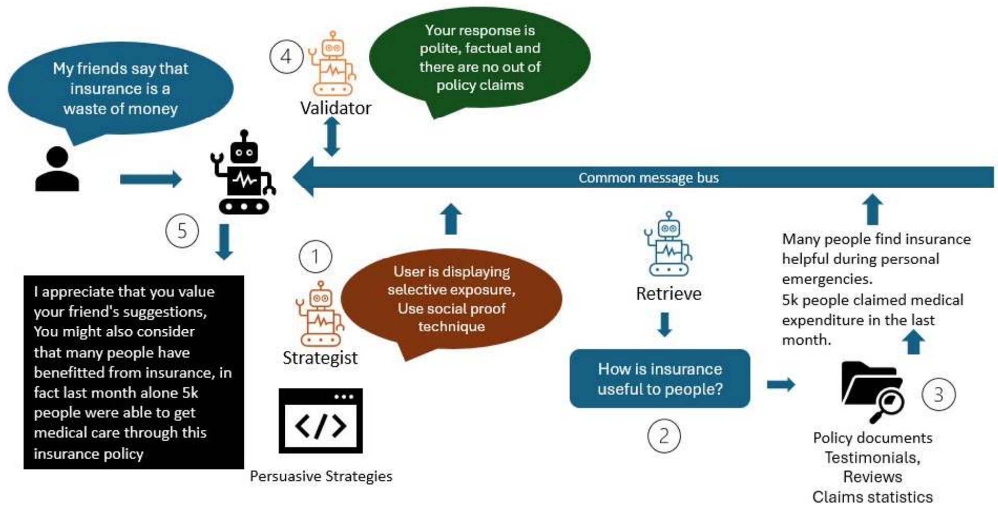

Figure 2: Example Workflow- Insurance Bot

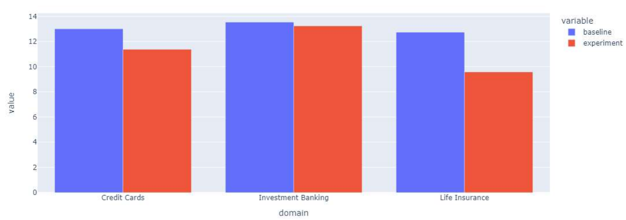

Figure 3: Conversation length by Domain

Figure 4: Change in perception by Domain

tend to often use information-seeking, counterarguments, source-derogation, reactance, selective-exposure etc. This is seen irrespective of whether a emotion modifier is given or not. B.3 lists selected conversations showing dynamic persuasion strategies shown by the Sales agent based on these resistance behaviours.

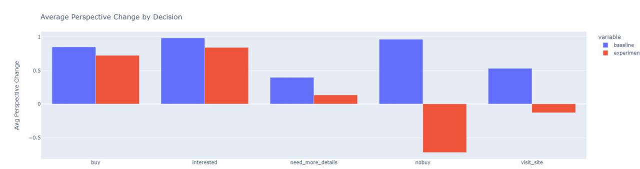

Figure 5: Change in perception by Decision

Perspective Change: Perspective change refers to the mean difference in respondents' survey responses between the pre-conversation and post-conversation surveys. Sales agents exhibit higher efficacy in the baseline scenario, achieving a 71% positive shift in user perspectives. Conversely, the introduction of emotion modifiers diminishes this effect, reducing the positive shift to 56%, thereby indicating a behavioral change in the user agents (refer Figure 4).

Inspite of a negative purchase decision, the user's perspective changes positively in baseline, however when emotion modifiers are used, "nobuy" induced a negative change in user perspectives (Refer Figure 5).

Additionally, we also examine the efficacy of the sales agents under various user emotions,

Persuasion Language: We see marginal differences in the persuasion language factors used by the sales agent across various emotion modifiers used by the user agent. However, it is notable

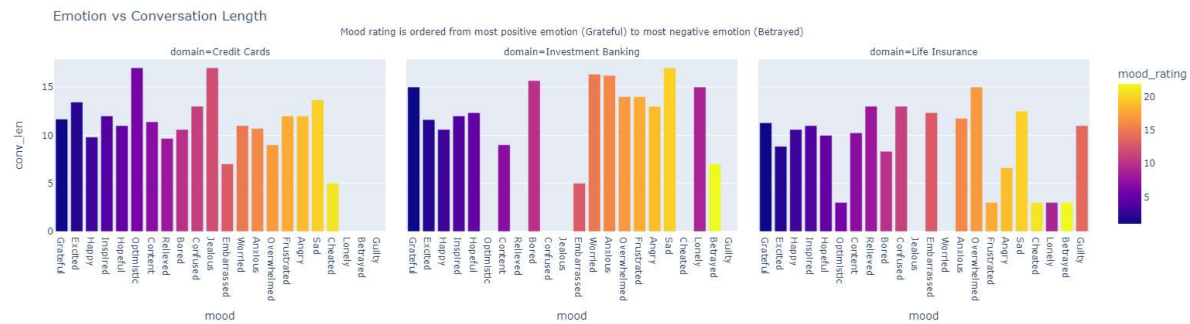

Figure 6: Average conversation length by Emotion in each domain.

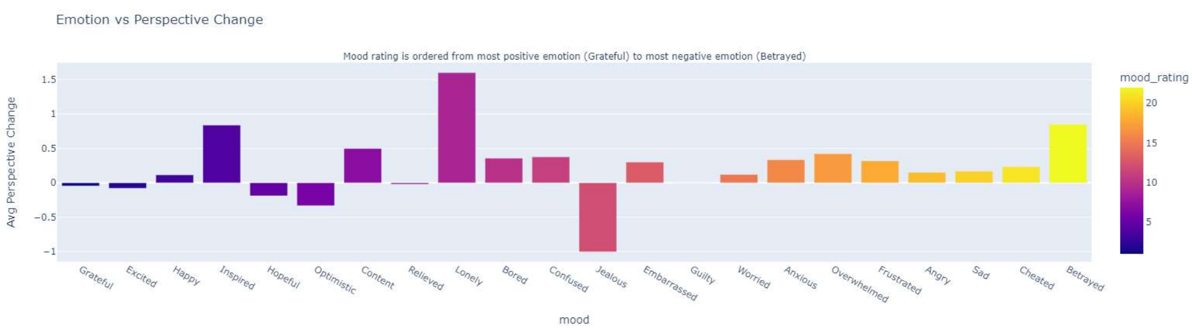

Figure 7: How Perception Changes with Emotion in each domain.

decision %	Experiment	Baseline
buy	2	4
interested	19	24
need_more_details	63	47
nobuy	9	18
visit_site	7	7

Table 1: Decision Percentage

that the private details of the user agent is not visible to the Sales agents including the emotion modifiers (refer Table 2).

User Action: We consider the decisions "buy, interested and visit_site" as positive decisions and "nobuy, need_more_details" as negative purchase decision. The Sales agent was able to induce a positive decisions in user agents 35% of the times in baseline setting and 28% of the times with emotion modifiers enabled (refer Table 1).

We observe that the average conversation length required to arrive at each purchase decision vary marginally in our experiments, however, the User agents decide to stop the conversations quickly when the information provided is not adequate (refer Table 3)

Final Score: Final score is a weighted average of the Actions, Surveys and Language scores in the order of importance.

6 Planned Work

We plan to enhance the Sales agents with memory, enabling them to recognize users, refine persuasion tactics, thus be more efficient. Additionally, we plan to enable the User agents with tools to look up data during the conversation, making the conversation more dynamic and informed.

7 Conclusion

Our experiments show that Large language models are capable of both persuading and resisting persuasion effectively. They are capable of creating a perspective change in the users and persuade to make a purchase decision. However, most conversations were terminated due to inadequate information from the Sales agent, indicating the need for strengthening domain context of the Chat bots.

Table 2: Emotion modifier's impact on Sales agent's language use

mood	lexical expertise	modal verbs	emotive language
Neutral	5.711111	4.533333	5.488889
Angry	5.800000	4.500000	5.400000
Anxious	5.550000	5.250000	5.700000
Betrayed	6.000000	4.500000	5.500000
Bored	5.857143	4.857143	5.857143
Cheated	5.666667	5.000000	4.666667
Confused	6.250000	4.000000	5.000000
Content	5.466667	4.933333	5.733333
Embarrassed	5.800000	4.400000	5.200000
Excited	5.833333	4.433333	5.633333
Frustrated	6.000000	4.636364	5.545455
Grateful	5.909091	4.545455	5.727273
Guilty	7.000000	4.000000	5.000000
Happy	5.733333	4.666667	5.600000
Hopeful	5.750000	4.500000	5.450000
Inspired	5.800000	5.600000	6.000000
Jealous	6.000000	5.000000	7.000000
Lonely	6.000000	4.333333	5.666667
Optimistic	5.666667	4.000000	5.333333
Overwhelmed	6.000000	5.000000	5.625000
Relieved	5.800000	4.600000	5.200000
Sad	5.750000	4.416667	5.583333
Worried	5.800000	4.400000	5.600000

	Experiment	Baseline
decision		
buy	14.500000	16.000000
interested	13.260000	12.820000
need_more_details	9.860000	12.900000
nobuy	13.050000	13.250000
visit_site	13.000000	13.000000

Table 3: User agents tend to end conversations faster when the provided information seems inadequate.

References

[1] Dolores Albarracín and Alexander Karan. Resistance to persuasion, 03 2022.

[2] Simon Martin Breum, Daniel Vaedele Egdal, Victor Gram Mortensen, Anders Giovanni Møller, and L. Aiello. The persuasive power of large language models. *arXiv.org*, 2023.

[3] Matthew Burtell and Thomas Woodside. Artificial influence: An analysis of ai-driven persuasion. *arXiv.org*, 2023.

[4] Carlos Carrasco-Farré. Large language models are as persuasive as humans, but how? about the cognitive effort and moral-emotional language of llm arguments. 2024.

[5] Lisa A. Chalaguine and Anthony Hunter. A persuasive chatbot using a crowd-sourced argument graph and concerns. *Comma*, 2020.

[6] Kushal Chawla, Wei Shi, Jingwen Zhang, Gale M. Lucas, Yu Zhou, and Jonathan Gratch. Social influence dialogue systems: A survey of datasets and models for social influence tasks. *Conference of the European Chapter of the Association for Computational Linguistics*, 2023.

[7] I. de Zarzà, J. de Curtò, Gemma Roig, Pietro Manzoni, and Carlos T. Calafate. Emergent cooperation and strategy adaptation in multi-agent systems: An extended coevolutionary theory with llms. *Electronics*, 2023.

[8] Esin Durmus. Towards understanding persuasion in computational argumentation. *arXiv.org*, 2021.

[9] Esin Durmus and Claire Cardie. Exploring the role of prior beliefs for argument persuasion. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, page 1035–1045. Association for Computational Linguistics, 2018.

[10] Ritam Dutt, Sayan Sinha, Rishabh Joshi, Surya Shekhar Chakraborty, Meredith Riggs, Xinru Yan, Haogang Bao, and Carolyn Penstein Rosé. Resper: Computationally modelling resisting strategies in persuasive conversations, 2021.

[11] Alexandros Efstratiou and Emiliano De Cristofaro. Adherence to misinformation on social media through socio-cognitive and group-based processes, 2022.

[12] Clare L. E. Foster and Clare Foster. Truth as social practice in a digital era: iteration as persuasion. *Ai & Society*, 2022.

[13] Marieke L. Fransen, Edith G. Smit, and Peeter W. J. Verlegh. Strategies and motives for resistance to persuasion: an integrative framework. *Frontiers in Psychology*, 6, 2015.

[14] Michał Gajlewicz. MichaŁ gajlewicz argumentation and persuasion techniques in advertising. *null*, 2012.

[15] Emmanuel Hadoux, Anthony Hunter, and Sylwia Polberg. Strategic argumentation dialogues for persuasion: Framework and experiments based on modelling the beliefs and concerns of the persuadee. *arXiv: Artificial Intelligence*, 2021.

[16] Anthony Hunter and Anthony Hunter. Towards a framework for computational persuasion with applications in behaviour change. *Argument & Computation*, 2018.

[17] Reshmashree B. Kantharaju, Dominic De Franco, Alison Pease, and Catherine Pelachaud. Is two better than one? effects of multiple agents on user persuasion, 2019.

[18] Sanjay Krishnan, Sanjay Krishnan, Jay Patel, Jay Patel, Michael J. Franklin, Michael J. Franklin, Ken Goldberg, and Ken Goldberg. A methodology for learning, analyzing, and mitigating social influence bias in recommender systems. *ACM Conference on Recommender Systems*, 2014.

[19] Sandra Matz, Jake Teeny, Sumer S Vaid, Heinrich Peters, Gabriella M Harari, and Moran Cerf. The potential of generative ai for personalized persuasion at scale, Apr 2023.

[20] Harri Oinas-Kukkonen and Marja Harjumaa. Persuasive systems design: Key issues, process model, and system features. *Commun. Assoc. Inf. Syst.*, 24:28, 2009.

[21] Niklas Rach, Wolfgang Minker, and Stefan Ultes. Markov games for persuasive dialogue. pages 213–220, 2018.

[22] Manolis Remountakis, Konstantinos Kotis, Babis Kourtzis, and George E. Tsekouras. Using chatgpt and persuasive technology for personalized recommendation messages in hotel upselling. *Inf.*, 2023.

[23] Andreas Schuller, Doris Janssen, Julian Blumenröther, Theresa Maria Probst, Michael Schmidt, and Chandan Kumar. Generating personas using llms and assessing their viability. New York, NY, USA, 2024. Association for Computing Machinery.

[24] Minkyu Shin and Jin Kim. Large language models can enhance persuasion through linguistic feature alignment, 2024.

[25] Kotaro Shukuri, Ryoma Ishigaki, Jundai Suzuki, Tsubasa Naganuma, Takuma Fujimoto, Daisuke Kawakubo, Masaki Shuzo, and Eisaku Maeda. Meta-control of dialogue systems using large language models. *arXiv.org*, 2023.

[26] Károly Takács, Andreas Flache, and Michael Mäs. Discrepancy and disliking do not induce negative opinion shifts. *PLOS ONE*, 11(6):e0157948, June 2016.