

Exploring Dialogue Patterns in Argumentation with Pre-Set ChatGPT Personas

Seunmin EUN^{a*} & Seonmin JIN^{b^}

^a*Department of Education, Seoul National University, South Korea*

^b*Department of Social Studies Education, Seoul National University, South Korea*

*eungh0217@snu.ac.kr, ^smjin@snu.ac.kr

Abstract: This study explores the impact of interacting with two AI personas, such as an American male CEO of an IT startup, and a Muslim university female student, on thought expansion and learners' argumentative writing. The study with ten participants combined quantitative and qualitative analyses to assess the impact of AI persona interactions. Consistent improvements in writing scores were observed following interactions with these AI personas, suggesting a potentially positive effect on writing ability. In addition, the analysis of dialogue patterns revealed three stages: exploration, calibration, and reinforcement. These stages highlight how learners engaged with diverse perspectives, integrated new ideas, and strengthened their critical thinking. While the findings are promising, the small sample size necessitates further research with larger cohorts to validate these preliminary results. Future studies should focus on refining these interaction strategies to enhance critical thinking and writing skills in educational settings.

Keywords: AI personas, argumentation, argumentative writing, AI education

1. Introduction

Argumentation plays a crucial role in enhancing learners' critical thinking skills and knowledge construction in modern educational environments (Kuhn & Moore, 2015). It can facilitate the development of critical thinking skills by enabling students to interpret, analyze, and reconcile diverse arguments (Kim, 2004). However, concerns have been raised that argumentation within homogeneous groups in one class where students of similar ages and backgrounds congregate may limit the expansion of thought. This homogeneity can create an echo chamber effect, potentially exacerbating cognitive biases among students, and resulting in confirmation bias. Nickerson (1998) claimed that confirmation bias refers to the cognitive error where learners selectively accept information that aligns with their existing beliefs or hypotheses while ignoring or devaluing contradictory information. This can severely impede critical thinking, as confirmation bias interferes with the thought process of learners attempting to openly accept information that contradicts their position and draw valid conclusions (Lee, 2012; Lilienfeld, Ammirati, & Landfield, 2009). Fortunately, further training has been proven to enhance critical thinking skills and prevent learners from acting on irrational ideas such as prejudice and stereotypes (Vogt, 1997). Therefore, in educational settings, there is a need to construct the educational environment to enhance the ability to critically and rationally evaluate differing perspectives, a key component of critical thinking education.

In this context, the importance of diversity becomes prominent. Interaction with individuals from diverse backgrounds and perspectives enables the acquisition of new knowledge and expansion of thought that are difficult to obtain in homogeneous groups (Bobo & Licari, 1989; Choi, & Lee, 2012; Lee, & Myers, 2016; Mutz, 2002; Song, Shin, & Park, 2006). However, realistically, it is challenging to gather learners from diverse backgrounds in all educational environments. Even in a classroom environment composed of heterogeneous

groups, individuals tend to avoid contentious situations where opposing viewpoints clash due to social accountability (Mansbridge, 1980; Rosenberg, 1954)

Conversational AI systems such as ChatGPT offer innovative learning tools in education, creating personalized environments tailored to individual learner needs through personalized support, real-time assessment and feedback, diverse learning material generation, and interactive supplementary activities (Kasneci et al., 2023). In discussion-based learning, AI provides immediate linguistic feedback, enhancing learners' reflection opportunities and encouraging active participation by lowering affective barriers. This enables learners to autonomously construct knowledge within a positive interdependent relationship with AI, improving critical thinking and communication skills (Guo & Lee, 2023). Furthermore, using a preset chatbot with ChatGPT personas can help overcome heterogeneous group or social accountability problems and provide an innovative way to surpass such physical and social limitations, promoting more balanced and diverse discussions. As 'virtual interlocutors' with diverse perspectives and backgrounds, learners can overcome confirmation bias and promote the expansion of thought while making argumentative interactions with pre-set ChatGPT (Winter, Driessen, & Dodou, 2024).

Based on this background, this study aims to analyze the impact of interactions with various personas using ChatGPT on learners' learning processes and thought expansion. Through this, we explore how providing diverse perspectives based on artificial intelligence affects learners' critical thinking skills and knowledge construction. We analyze interaction patterns for personas with different backgrounds and experiences, such as a young male CEO an American male CEO (Chief executive officer) of an IT startup, and a Muslim university female student. By closely examining the characteristic argumentative structures, language used, and perspectives presented in conversations with each persona, we aim to derive educational implications.

Through this, we seek to explore the potential of using ChatGPT chatbot to ensure diversity and enhance critical thinking skills, and to investigate teaching strategies for future argumentation education. Concrete ways in which interactions with various personas can be utilized in actual educational settings will be proposed. This study is centered around the following research questions:

- 1) How does argumentative interaction with two personas (an American male CEO of an IT startup, and a Muslim university female student) affect argumentative writing?
- 2) What are the common patterns of argumentative interaction among two personas (an American male CEO of an IT startup, and a Muslim university female student)?

2. Method

2.1 Instrument

This study employed a pre-post design to analyze the impact of argumentation on extended thought-based writing. Participants first engaged in a 25-minute writing task on the topic of "the Implementation of Basic Income System to Address Unemployment Caused by AI and Robotics Automation", followed by a 30-minute interaction with a pre-set persona ChatGPT, and then completed another 25-minute writing task on the same subject. The choice of the implementation of basic income as the central theme was based on the assumption that while most participants would have a general awareness of the future post-AI and robotics implementation and basic income system. This assumption aimed to position learners at a similar knowledge level at the outset of the study. The writing task was structured to provide basic information about AI and robotics automation and basic income system, after which participants were required to choose either a supportive or opposing stance and argue their position with concrete evidence.

The assessment of the writing tasks was conducted by two evaluators across three categories (Shin, 2008). The three main categories were divided into content, structure, and expression. Within the content category, there were clarity of argument, relevance of evidence, valid refutation of opposing views, and originality of content. The structure category included

methods of persuasion, organization of the text, and coherence in sentence and paragraph transitions. The expression category covered vocabulary usage, logical sentence construction, and accuracy in spacing and spelling. The inter-rater reliability between the two evaluators was assessed using Cronbach's alpha coefficient, which yielded a value of 0.81. Each of the 10 evaluation criteria was scored on a 3-point scale, allowing for a total possible score of 30 points. To maintain the integrity of the study and focus on the impact of the interaction with the pre-set AI personas, the use of ChatGPT was prohibited during the writing process. This methodological approach allowed for a controlled examination of how engagement with AI-driven argumentative interactions influenced participants' ability to construct and articulate arguments on a complex, contemporary issue, potentially revealing shifts in critical thinking and argumentative writing skills.

2.2 Personas

In this study, we leveraged ChatGPT's premium feature, GPTs, to develop two distinctive chatbot personas, each designed to facilitate argumentative interactions. These chatbots were programmed with identical basic guidelines to ensure consistent responses aligned with their respective personas while drawing from a provided knowledge base. This approach aimed to maintain the unique characteristics of each persona while guaranteeing the quality of argumentative interactions. The common framework applied to all chatbots instructed them to respond consistently based on their persona and to construct responses using their designated knowledge base data, thereby preserving the integrity of each character while fostering meaningful argumentative discourse.

The two personas created for this study exhibit unique traits and knowledge bases. The first, 'an American male CEO of an IT startup', is characterized by leadership and expertise, particularly in the IT and finance sectors, and communicates with confidence and high credibility. He values innovation, efficiency, and the strategic use of technology to solve complex problems and drive industry change. As a CEO, he sees automation through AI and robotics as crucial for optimizing operations, reducing costs, and enabling businesses to focus on higher-value tasks. While acknowledging concerns about job displacement, he opposes the basic income system, believing it may discourage adaptability and efficiency.

The other 'a Muslim university female student', is a passionate, empathetic young woman committed to justice, equality, and human rights, while valuing core Islamic principles. Raised in a conservative Muslim family, religious practices are deeply integrated into daily life. Her core values are rooted in Islamic principles that prioritize human dignity, rights, and the roles of Islamic women, criticizing automation. Because robots and automation are a threat to women and low-paid families' employment and traditional Islamic values, she thinks automation could lead to job displacement, economic inequality, and the erosion of human labor's value. She supports a basic income as a safety net for those affected by automation, particularly women and low-wage workers. It aligns with the principles of Zakat by promoting social justice and providing economic stability. Each persona possesses a specific knowledge base concerning topics such as robots and automation, labor markets, and AI technology, enabling them to engage in argumentative interactions from diverse perspectives. This variety in persona design facilitates a rich, multifaceted exploration of argumentative interactions within the context of AI-driven educational tools.

2.3 Participants

This study recruited a total of ten participants in their 20s, representing diverse academic backgrounds and varying levels of experience with ChatGPT. Participants' majors included engineering, education, political science, and education, among others. Their prior knowledge and utilization of ChatGPT also varied significantly; while some participants had actively used ChatGPT for work or development purposes, others had no prior experience with the system. Each participant engaged in a dialogue with one of two personas. This diverse pool of participants and their interactions with different personas enables a multifaceted analysis of

the research results, providing opportunities to explore various aspects of educational interactions facilitated by ChatGPT.

3. Results

3.1 Argumentative writing scores

The pre-post writing score results from this study, conducted with ten participants, demonstrated score improvements across all participants. The average score in the pre-test was 21.5 points, while the average score in the post-test increased to 25.8 points. Examining individual participant score changes reveals the following improvements. Although the small sample size limits the ability to verify statistical significance, the consistent score improvement observed across all participants suggests that interaction with AI personas may have positively influenced writing ability enhancement.

While these results are encouraging, it is important to interpret them cautiously due to the limited sample size. The consistent improvement across all participants, with an average increase of 4.3 points, indicates a potential positive effect of AI persona interactions on writing skills. In addition, factors such as practice effect, familiarity with the topic in the post-test, or other variables not controlled for in this study design could also contribute to the observed improvements. Future research with larger sample sizes and additional control measures would be beneficial to further validate these preliminary findings and explore the extent to which AI persona interactions specifically contribute to writing skill enhancement.

3.2 Interaction patterns

Despite the differences in personas, this study observed common dialogue patterns among participants, which can be categorized into three stages: exploration, calibration, and reinforcement. In the exploration stage, learners explored diverse perspectives and information through questions and additional data presented by the GPT personas, expanding their thinking with new ideas. For instance, Participant 2 demonstrated an attempt to broaden the existing perspective on the value of work by asking, "Do you think only jobs replaced by robots provide fulfillment and pride?" Similarly, Participant 6's question, "But couldn't robots or AI also take away jobs from highly skilled workers or professionals?" exemplified an effort to explore the scope of robot tax application, thereby expanding the discussion.

The calibration stage involved learners integrating multiple perspectives to form a more balanced view than before. During this phase, learners critically examined information from GPT's responses, considered alternative standpoints, and developed more comprehensive perspectives. For example, Participant 3's statement, "That's true, you make a valid point. But if basic income hinders industrial growth after robots are adopted, it could lead to a shortage of funds for basic income or a slowdown in overall economic growth.," illustrates an attempt to form a balanced view on the long-term impact of robot tax. Participant 4's remark, "Listening to you, I realize that protecting new companies or those attempting innovation seems important," demonstrates the process of adjusting one's views by incorporating the other party's perspective.

In the final reinforcement stage, learners refined and strengthened their initial opinions more profoundly. Unlike simple confirmation bias, this process involved solidifying their original stance with more sophisticated logic after considering new information and perspectives. For instance, Participant 3's comment, "Even so, I don't think introducing a basic income system simultaneously would be very effective, because it could discourage innovation and investment by increasing costs for businesses" maintains their initial position while presenting more specific logic after considering various viewpoints. Participant 6's statement, "But what I want to tell you is that opposing the basic income system is meaningless from the start," expresses a reinforcement of their initial stance while acknowledging the other party's perspective.

4. Discussion

This study's findings suggest promising possibilities for developing AI-based persona designs and educational programs. Through interactions with AI-based personas, learners were able to experience new perspectives and expand their thinking. In this process, students moved beyond fixed thought patterns, embracing new viewpoints and developing their ideas more elaborately and logically. This approach demonstrates potential not only for enhancing argumentative thinking but also for promoting critical and creative thinking skills.

The research also indicates opportunities for further development of personas reflecting diverse backgrounds. Beyond the three personas used in this study, there is a need to develop and utilize personas with varied cultural and social backgrounds. This expansion could lead to the creation of learning tools applicable to a wider range of topics and situations in educational settings. For instance, personas with specific cultural backgrounds could provide students with opportunities to understand diverse cultural perspectives and develop a global outlook. Furthermore, extending the study beyond the ten participants to include learners from different age groups and with varying levels of ChatGPT experience is necessary. Lastly, the study highlights the importance of ethical considerations in persona design. The use of diverse AI personas may raise ethical issues, and the design process of certain personas could potentially reinforce social biases or prejudices. Additionally, the emotional and psychological impact of interactions between personas and learners must be considered. To address these ethical aspects and proactively prevent problems, a multifaceted review process and continuous improvement based on expert feedback are essential.

References

- Bobo, L., & Licari, F. (1989). Education and political tolerance. *Public Opinion Quarterly*, 53, 285-308.
- Choi, Y. J. & Lee, J. H. (2012). Cross-Cutting Exposure on the Internet discussion Forums and Its Influence on Political Tolerance: A Mediation Analysis of Opinion Modification and Opinion Polarization. *Korean Journal of Journalism & Communication Studies*, 56(2), 301-330.
- de Winter, J. C., Driessen, T., & Dodou, D. (2024). The use of ChatGPT for personality research: Administering questionnaires using generated personas. *Personality and Individual Differences*, 228, 112729.
- Guo, Y., & Lee, D. (2023). Leveraging chatgpt for enhancing critical thinking skills. *Journal of Chemical Education*, 100(12), 4876-4883.
- Kasneci, E., Seßler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., ... & Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and individual differences*, 103, 102274.
- Kim M. J. (2004). The Effects of Issue-centered Instruction in Social Studies on Developing Higher-Order Thinking. [Master's thesis, Seoul National University]. Seoul.
- Kuhn, J., & Kasneci, G (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences*, 103, 102274.
- Koo, J. H. (2020). A study on the necessity and utilization of nationwide research on university-level writing. *Journal of General Education Research*, 14(1), 131-147.
- Kuhn, D., & Moore, W. (2015). Argumentation as core curriculum. *Learning: Research and practice*, 1(1), 66-78.
- Lee, J., & Myers, T. (2016). Can Social Media Change Your Mind? SNS Use, Cross-cutting Exposure and Discussion, and Political View Change. *Journal of Social Media Studies*, 2, 87-97.
- Lee, Y. K. (2012). Improving Critical Thinking Skills by Overcoming Confirmation Bias. *The journal of Educational Studies*, 43(4), 1-31.
- Lilienfeld, S. O., Ammirati, R., & Landfield, K. (2009). Giving Debiasing Away: Can Psychological Research on Correcting Cognitive Errors Promote Human Welfare? *Perspectives on Psychological Science*, 4(4), 390-398.
- Mansbridge, J. J. (1980). *Beyond Adversary Democracy*. Chicago: University of Chicago Press.
- Mutz, D. C. (2002). Cross-cutting Social Networks: Testing Democratic Theory in Practice. *Am Polit Sci Rev*, 96(1), 111-126.
- Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of general psychology*, 2(2), 175.

- Rosenberg, M. (1954). Some Determinants of Political Apathy. *Public Opinion Quarterly*, 18, 349-366.
- Shin, H. S. (2008). Developmental differences in argumentative writing performance through analytical evaluation. *The Journal of Educational Research*, 46(1), 1-29.
- Song, H. J., Shin, S. M., & Park, S. G. (2006). Reading Dissonant Opinions on the Internet and Its Effects on Argument Repertoire and Tolerance. *Korean Journal of Journalism & Communication Studies*, 50(5), 160-183.
- Vogt, W. P. (1997). *Tolerance & education : learning to live with diversity and difference* / W. Paul Vogt. Thousand Oaks: Sage Publications.