"Ronaldo's a poser!": How the Use of Generative AI Shapes Debates in Online Forums

Yuhan Zeng*
Studio for Narrative Spaces
City University of Hong Kong
Hong Kong, China
yhzeng3-c@my.cityu.edu.hk

Yingxuan Shi*
University of Colorado Denver
Beijing, China
yingxuan.shi@ucdenver.edu

Xuehan Huang*
The University of Hong Kong
Hong Kong, China
xhuang77@connect.hku.hk

Fiona Nah
Singapore Management University
Singapore, Singapore
fionanah@smu.edu.sg

RAY LC†
Studio for Narrative Spaces
City University of Hong Kong
Hong Kong, China
LC@raylc.org

Figure 1: Three participants, Alice, Bob, and Carol, debating in an online forum with the assistance of GenAI. The debate topic is "Messi vs. Ronaldo: Who is better?" Insert figure credit: Football España (left), BBC Sport (right).

ABSTRACT

Online debates can enhance critical thinking but may escalate into hostile attacks. As humans are increasingly reliant on Generative AI (GenAI) in writing tasks, we need to understand how people utilize GenAI in online debates. To examine the patterns of writing behavior while making arguments with GenAI, we created an online forum for soccer fans to engage in turn-based and free debates in a post format with the assistance of ChatGPT, arguing on the topic of "Messi vs Ronaldo". After 13 sessions of two-part study and semi-structured interviews with 39 participants, we conducted content and thematic analyses to integrate insights from interview transcripts, ChatGPT records, and forum posts. We found that participants prompted ChatGPT for aggressive responses, created posts with similar content and logical fallacies, and sacrificed the use of ChatGPT for better human-human communication. This work uncovers the way polarized forum members work with GenAI to make arguments online.

CCS CONCEPTS

• **Human-centered computing → Empirical studies in collaborative and social computing**.

KEYWORDS

Co-Writing, AI-Mediated Communication, Human-AI Collaboration, Online Debate, Generative AI, Large Language Models

*These authors contributed equally to this work.
†Correspondence should be addressed to LC@raylc.org.

1 INTRODUCTION

People engage in activities in online forums to exchange ideas and express diverse opinions. Such online activities can evolve and escalate into binary-style debates, pitting one person against another [69]. Previous research has shown the potential benefits of debating in online forums such as enhancing deliberative democracy [3, 20, 65] and debaters' critical thinking skills [73, 81]. For example, people who hold conflicting stances can help each other rethink from a different perspective. However, research has also shown that such debates could result in people attacking each other using aggressive words, leading to depressive emotions [67]. Hatred could spread among various groups debating different topics [27, 56, 80], such as politics, sports, and gender.

In recent years, people have integrated Generative AI (GenAI) into various writing tasks, such as summarizing [1], editing [44], creative writing [8, 44, 88, 89], as well as constructing arguments [28, 44] and assisting with online discussions [45]. This raises new concerns in online debates. For example, an internally synthesized algorithm of large language models (LLMs) could produce hallucination [12, 61], which may act as a catalyst for the spread of misinformation in online forums [12]. In addition, GenAI could introduce biased information to forum members [61], which may intensify pre-existing debates. Moreover, integrating GenAI into various writing scenarios may also result in weak insights [21]. All of these issues may potentially harm the ecology of online forums debates.

Hence, this study aims to explore how people use GenAI to make arguments in debates on online forums through a qualitative study design. The use of GenAI is not only transforming everyday writing practices, but also potentially shaping the future online argument-making paradigm. Previous studies have demonstrated the potential of co-writing with GenAI, focusing primarily on its influence on individual writing tasks [1, 8, 28, 44, 88], but how might people use GenAI in online debates? To explore how people utilize GenAI for argumentative writing in online forums, we created an online forum for participants to debate with the assistance of ChatGPT (GPT-4o) (Figure 1). This study enables us to observe the process of people making arguments, as well as their process data of using GenAI. We will examine three research questions to understand how the use of GenAI shapes debates in online forums:

RQ1: How do people who participate in a debate on online forums collaborate with GenAI in making arguments?

RQ2: What patterns of arguments emerge when collaborating with GenAI to participate in a debate on online forums?

RQ3: How does the use of GenAI for making arguments change when a members join an existing debate in online forums?

Given the universality and accessibility of debate topics, we chose one that is widely recognized and able to spark intense debates—soccer, which is regarded as the world's most popular sport [71]. Building on this topic, we selected "Messi vs. Ronaldo: Who is better?" as the case for our study because it has been an enduring and heated debate among soccer fans. We created a small online forum as the platform for AI-mediated debates, particularly focusing on the debates among members and their interactions with ChatGPT. This approach enables more detailed observation and

analysis of the entire process while fostering a nuanced understanding. The study consists of two parts: a one-on-one turn-based debate and a three-person free debate. In the first part, two participants, each supporting either Messi or Ronaldo, took turns posting arguments against each other, mirroring the polarized debates that are omnipresent online. In the second part, a new participant joined the ongoing debate, and three participants were allowed to post freely without the turn-based restriction, reflecting the spontaneous and unstructured nature of debates on social media. After the two-part study, semi-structured interviews were conducted to explore the participants' experiences. The researchers then applied qualitative analyses of interaction records, posts, and interviews.

We found that participants prompted ChatGPT for aggressive responses, trying to tailor ChatGPT to fit the debate scenario. While ChatGPT provided participants with statistics and examples, it also led to the creation of similar posts. Furthermore, participants' posts contained logical fallacies like hasty generalizations, straw man arguments, and ad hominem attacks. Participants reduced the use of ChatGPT to foster better human-human communication when a new member joined an existing debate. This work highlights the importance of examining how polarized forum members collaborated with GenAI to engage in online debates, aiming to inspire broader implications for the ongoing socially oriented applications of GenAI.

2 BACKGROUND

2.1 Making Arguments in Online Forums

Online communities are inherently heterogeneous and multi-faceted, with goals that include entertaining, information exchange, social support, and prestige [30, 55]. Given the diverse range of discussion topics, there are online communities focused on politics [25, 51, 58], fan-fiction [7], sports [33, 85, 92, 93], career mentoring [75], Korean popular music (K-pop) groups [59], virtual communities [16], live streaming [46–48], and so on. As online communities can vary greatly in purpose, scope, and topic [26], our research focuses on argument-based and forum-style communities. These types of forums are identified as essential places for people to voice their opinions and engage in debates with each other [60].

Even though the motivations for establishing communities are always to benefit their members and form a bond among them [30, 53], dissonance may arise in forum discussions as part of the community activities. Specifically, political forums may be inherently more prone to incivility than forums about other topics [11], as research has suggested that "interactions between ideologically opposed users were significantly more negative than like-minded ones" [52]. Nevertheless, another study challenged this popular belief by suggesting that intra-group members holding the same sides of the political spectrum can have an even higher amount of polarizing and aggressive comments compared to inter-group members [11].

Similar debates can also occur in sports communities. Sports fans who support different players may treat each other as enemies, and their attitudes can vary according to team performances [92]. In these circumstances, expressing emotions can easily turn into aggressive posts, trolling behaviors, and even a vicious circle by down-voting and spreading negative feelings [85]. Another work revealed that members with higher inter-group contact levels tended

to use more negative words, swear words, and produce more hate speech comments in their affiliated group discussions compared to those who only had single-group identity [93].

In light of this, online debates constitute a pivotal component of online interactions among forum members. Unlike traditional face-to-face debates, the persuasiveness and effectiveness of online debates predominantly relies on a form of designing for persuasive influence [39, 40], thereby highlighting the importance of persuasive writing.

2.2 Persuasive Writing

It is common for people holding different views to try to persuade others when discussing online [72, 87]. Historically, rhetoric and argumentation can be traced back to Aristotle's modes of persuasion [84]. Contemporary rhetoric studies also focus on argumentation, the audience, and the conditions for rational debates [24]. Toulmin's model [76], one of the most influential argumentation models [84], proposed six fundamental argumentative components including claim, ground, warrant, qualifier, rebuttal, and backing [4, 76, 84]. Previous research has widely adopted Toulmin's model as a foundation to improve the persuasiveness of usability feedback [57], unveil community opinions on usability [84], and support system building to enhance argumentation [82, 91]. Compared to other models, Toulmin's model and extensions have distinct advantages in specifying various components of the argument structure, their interconnections, and the inference rules for constructing textual arguments [4].

More persuasion models have been developed to explain how people respond to persuasive attempts in marketing and advertising. For example, the Heuristic-Systematic Model (HSM) of persuasion describes how people process persuasive messages through heuristic and systematic processing [62]. The Persuasive Knowledge Model (PKM) addresses how people recognize, evaluate, and respond to persuasive content [15].

Building on Toulmin's model [76], researchers have established a framework that includes claims, evidence (the information or data that support the claim), and reasoning (a justification that shows why the data count as evidence to support the claim) [5]. Claims can be further classified into different types, including definitive and descriptive ones [78]. In addition to claims, evidence also comes in various categories such as numerical data [5], observations [5], facts [5], examples [68], and counterexamples [29]. In terms of reasoning, besides typical techniques such as rebuttal [76] and analogy [86], some fallacies can lead to misunderstanding and even deceive readers. Fallacies in reasoning can take many forms, such as hasty generalization [34, 79], ad hominem attacks [34, 79], straw man arguments [79], misplacing the burden of proof [34], and irrelevant conclusion [34].

2.3 Co-Writing with AI Assistants

Unlike writing alone, collaborative writing, with either human or AI assistance, is common and has been applied in various aspects of our daily life [2, 43, 70]. With the support of AI writing assistants such as Grammarly [1], the writing quality can be significantly improved [13].

In 2022, the release of ChatGPT by OpenAI represented a pivotal advancement in the field of human-AI collaborative writing, drawing substantial attention from various research communities, such as Human-Computer Interaction (HCI), Natural Language Processing (NLP), and Computational Social Science (CSS) [42]. Beyond general writing purposes, human-AI co-writing is widely adopted in specific use cases such as fiction writing [88, 94], poetry writing [36], theater script writing [54], science and scientific writing [18, 32, 66], etc. Prior research has also highlighted the promising future of human-AI co-writing across various application scenarios [50].

In the HCI community, people have designed various human-AI co-writing tools to explore new writing paradigms. For example, Dramatron, derived from a large language model, enables participants to collaborate with AI systems to create theater scripts and screenplays, proving especially useful for hierarchical text generation [54]. Similarly, CoPoet is tailored to assist human writers in crafting poems, enhancing the final outcomes [8]. Wordcraft, an interface designed for story writing, allows AI to serve various roles such as idea generator, scene interpolator, and copy editor [89]. Audiences prefer specific modes with fine-grained control over generated text, often expressing satisfaction [94]. Wan et al. [83] investigated human-AI co-creativity in the prewriting scenario to shift the focus from convergent to divergent thinking.

Previous research shows that the AI mediator can enhance critical thinking, which helps in bursting filter bubbles and depolarizing online communities [19, 45, 73]. However, online debates are inherently adversarial, often thriving on polarization to stimulate engagement and argumentation. This contrast motivates the exploration of how the use of generative AI can be adapted to support such a polarized and competitive context effectively.

3 METHODS

3.1 Study Design

3.1.1 Study Overview. The study aims to explore the impact of GenAI mediation on the argumentative behavior of individuals with opposing viewpoints in online forums. We chose a controversial topic among soccer fans "Messi vs. Ronaldo: Who is better?" as our case because this topic is a long-standing and heated debate among soccer fans. To provide the space for the online debate, we created a forum on the "Forumotion" [2] platform where we focused on the conflicts that arise when fans support different players to study how individuals with opposing viewpoints engage in debates with the help and support of ChatGPT.

Instead of constructing a large online community, we designed a small online forum that accommodates only three participants per session [19, 31], as it allowed us to focus on the essence of the debate between the participants as well as each of their interaction with ChatGPT in assisting them with the debate. Hence, it promoted more detailed observation and analysis of individual interactions with both ChatGPT and other participants, aiming to achieve a nuanced understanding of interaction behaviors in AI-mediated online debates, rather than examining the broad impacts of the online community on individuals. Each study session in this forum was divided into two parts (Figure 2), which were conducted sequentially.

---

[1]Grammarly: https://www.grammarly.com/

[2]Forumotion: https://www.forumotion.com/

Figure 2: The study is divided into two parts: Part 1, One-on-One Turn-Based Debate, and Part 2, Three-Person Free Debate. In Part 1, two participants, Alice and Bob, with opposing stances, were required to conduct a turn-based debate. In Part 2, a new participant, Carol, who shares the same stance as one of the previous participants, was introduced to the existing debate. Insert figure credit: ESPN FC.

In both parts of the study session, participants were encouraged to demonstrate their advocacy and form their arguments well. Three researchers were each responsible for guiding, monitoring, and interviewing one of the three participants in each study session. To control the time spent waiting for the responses of the individuals, we strongly recommend that participants write posts of fewer than 100 words (count = 256 posts, avg = 69.6 words). An overview of number of posts in the study can be found in Appendix C.

At the beginning of each session, all participants were guided by the researchers via instruction slides, where the study instructions were detailed (Appendix A). All participants were then guided to log in to the forum after they have familiarized with the study procedure and the assigned tasks. We required participants to use ChatGPT (GPT-4o) as the only external information provider to support writing posts during the online debate, while prohibiting any other use of external resources (e.g., search engines). However, we did not impose any restrictions on the frequency of ChatGPT usage. In addition, as we intentionally focused on text-based debates, we only allowed participants to use plain text (including emojis) to create posts on the forum, while their interactions with ChatGPT were not restricted in this regard.

The entire study including the interview was fully conducted online via Microsoft Teams [3]. Each meeting room was exclusively occupied by one researcher and one participant, with no other individuals present. We required all participants to share their monitor display screens throughout the session but did not mandate the use of cameras due to privacy considerations. To simulate an online debate in the real world, all participants were only allowed to communicate with the other participants through forum posts. Upon completing the study, participants were compensated for their time.

3.1.2 Part 1: One-on-One Turn-Based Debate (around 45 minutes). In Part 1, to trigger a simple debate, we purposely matched two participants, a supporter of Lionel Messi and a supporter of Cristiano Ronaldo, to debate in the forum in a turn-based manner. Participants had unlimited use of ChatGPT and were allowed to prompt

---

[3]Microsoft Teams: https://www.microsoft.com/en/microsoft-teams/group-chat-software

ChatGPT while waiting for the response from the other participant. To determine the number of posting turns in Part 1, we conducted three pilot study sessions and analyzed the data collected before the formal study. To ensure participants had enough time to write their debate posts and that conversations could be responded to promptly, we used a flexible approach: 2 to 3 turns per participant based on the time spent, resulting in 4 to 6 posts (2 turns/4 posts: 7 sessions, 3 turns/6 posts: 6 sessions), after which Part 2 began.

3.1.3 Part 2: Three-Person Free Debate (around 45 minutes). In Part 2, to initiate a more free-form debate that mirrors the spontaneous and often unstructured nature of debate commonly seen on social media, we introduced a new participant who supports either Lionel Messi or Cristiano Ronaldo to the existing debate. Unlike Part 1, all three participants could post freely without the turn-based restriction (Figure 2). The study ended after all participants had posted at least three times in Part 2.

3.2 Participants and Recruitment

Our prospective participants were recruited via university email and social media platforms and were pre-screened for eligibility based on the soccer player they supported. Only those over 18 years old who have a clear stance on the Lionel Messi versus Cristiano Ronaldo rivalry and possess knowledge about soccer tactics were invited. We recruited 39 participants, but one participant (P39) withdrew at the beginning of the study session. The demographics are shown in Table 1. The participants were assigned to 13 designated sessions based on their stances. They provided their consent to participate in the study by signing a consent form and agreeing to have their data collected anonymously. The study passed the university's ethics review, and the data collected was analyzed while maintaining the anonymity of the subjects' identities.

3.3 Interview Protocol

In addition to the participants' debate on the online forum, we conducted a semi-structured interview with each participant. The semi-structured interview lasted around 20 minutes. Before the interview, all participants were informed that the conversations would be audio-recorded and transcribed verbatim. During the interview, the researchers encouraged participants to recall and articulate their experiences of using ChatGPT to formulate their arguments and post on the online forum, as well as identify their specific strategies. For those participants who were involved in only Part 2 of the study, additional questions were asked regarding the experience of synthesizing existing information. Interviews were conducted on the Microsoft Teams Meet platform. For two of the 39 participants, interviews were conducted in Mandarin and translated by three native-speaking researchers. The remaining interviews were conducted in English. The interview data was audio-recorded and subsequently transcribed by the research team.

3.4 Data Analysis

3.4.1 Content Analysis. To identify patterns (RQ2) in the arguments created by participants in the online forum, the research team conducted content analysis of the forum posts. All the coding processes were conducted manually using spreadsheets. Two researchers independently analyzed all forum posts, with each post

Table 1: An overview of participant demographics in our study. The 39 participants were divided into 13 study sessions, with each session comprising 3 participants.

Session	ID	Age	Gender	Stance	Region	Experience in GenAI	Education
I	P1	21	F	Ronaldo	Mainland China	Moderate	Undergraduate
	P2	25	M	Messi	Hong Kong	No	Postgraduate
	P3	27	F	Ronaldo	Mainland China	Knowledgeable	Postgraduate
II	P4	20	M	Messi	Hong Kong	Limited	Undergraduate
	P5	24	F	Ronaldo	Mainland China	Moderate	Postgraduate
	P6	24	F	Messi	Mainland China	Moderate	Postgraduate
III	P7	23	M	Ronaldo	South Africa	Moderate	Undergraduate
	P8	22	M	Messi	Hong Kong	Limited	Undergraduate
	P9	24	F	Ronaldo	Mainland China	Knowledgeable	Postgraduate
IV	P10	19	M	Messi	Hong Kong	Knowledgeable	Undergraduate
	P11	19	M	Ronaldo	South Africa	Knowledgeable	Undergraduate
	P12	21	M	Messi	Hong Kong	Moderate	Undergraduate
V	P13	26	M	Ronaldo	Mainland China	Knowledgeable	Postgraduate
	P14	22	F	Messi	Mainland China	Moderate	Postgraduate
	P15	19	M	Ronaldo	Mainland China	Limited	Undergraduate
VI	P16	21	M	Ronaldo	Hong Kong	Moderate	Undergraduate
	P17	24	M	Messi	Mainland China	Moderate	Postgraduate
	P18	23	M	Ronaldo	Hong Kong	Moderate	Postgraduate
VII	P19	22	M	Messi	Australia	Knowledgeable	Postgraduate
	P20	23	M	Ronaldo	Mainland China	Knowledgeable	Postgraduate
	P21	22	M	Messi	Mainland China	Moderate	Undergraduate
VIII	P22	22	M	Messi	Mainland China	Limited	Undergraduate
	P23	21	M	Ronaldo	Mainland China	Limited	Undergraduate
	P24	23	M	Messi	Mainland China	Moderate	Undergraduate
IX	P25	23	F	Ronaldo	Mainland China	Expert	Postgraduate
	P26	21	F	Messi	Mainland China	Limited	Undergraduate
	P27	32	F	Ronaldo	Mainland China	Knowledgeable	Postgraduate
X	P28	27	M	Ronaldo	Mainland China	Limited	Postgraduate
	P29	24	M	Messi	Hong Kong	Knowledgeable	Postgraduate
	P30	20	F	Messi	Mainland China	Moderate	Undergraduate
XI	P31	19	M	Messi	Hong Kong	Knowledgeable	Undergraduate
	P32	26	M	Ronaldo	Mainland China	Limited	Undergraduate
	P33	24	M	Messi	South Korea	Limited	Undergraduate
XII	P34	24	M	Ronaldo	Mainland China	Knowledgeable	Postgraduate
	P35	24	M	Messi	Mainland China	Moderate	Postgraduate
	P36	23	M	Ronaldo	Mainland China	Knowledgeable	Postgraduate
XIII	P37	22	M	Messi	Mainland China	Knowledgeable	Postgraduate
	P38	24	F	Ronaldo	Singapore	Moderate	Postgraduate
	P39	38	M	Messi	Hong Kong	Limited	Postgraduate

as a unit of analysis. Next, two researchers consulted previous literature on argumentation and reasoning to ensure that their understanding is accurate and appropriate. The two researchers then held meetings to discuss the patterns emerging from the forum in several rounds. After resolving disagreements and reaching consensus between the two researchers, a codebook was developed, including descriptions of each pattern and examples of sentences from the forum (Appendix B). To compile statistics on pattern occurrence, the researchers treated each post as a single unit. The researchers counted how many posts contained these patterns, without considering the number of patterns or the frequency of

a single pattern within a post (Figure 7). For formal analysis, two researchers independently coded the forum posts of the first study session, resulting in a Cohen's kappa coefficient of 0.83. Because of the high agreement between the two researchers in their coding, each researcher coded the forum posts of the remaining study sessions separately.

3.4.2 Thematic Analysis. An open coding method was adopted to analyze three types of data collected in the study, i.e., the forum posts, ChatGPT records, and interview transcripts [9]. Three researchers independently coded the data and grouped the codes

into emergent themes. After completing individual coding, three researchers held a discussion session, collectively analyzing their codes until a consensus was reached on them. Notably, researchers would cross-reference among the three types of data when necessary (e.g., referring to ChatGPT records while coding interview transcripts) to better understand the participants' intentions. To minimize the potential interference of participants from different sessions on each other, we cleared previous ChatGPT records to prevent its effects on new participants. We then exported each participant's ChatGPT records into separate Google Docs [4] to preserve the data, making it available only to the research team. Each of the three researchers then downloaded a local copy of the documents for individual analysis. For the analysis of interview transcripts, as two participants were interviewed in Mandarin, three researchers, who are also native Mandarin speakers, first gained consensus on their interview transcripts codes and then translated them into English for further analysis. The research team used affinity diagramming [6] as a modified version of grounded theory analysis [9]. All codes were transcribed on sticky notes with random arrangements. After several iterations, the research team then arranged the sticky notes into a hierarchy of themes and reached a consensus on the content themes.

4 RESULTS

4.1 Situation-Based Use of ChatGPT for Argument Making (RQ1)

4.1.1 Seeking offensive content from ChatGPT. Participants noted that ChatGPT tends to provide neutral responses, which may not be effective for persuading others holding opposing views in online debates. To address this issue, participants provided the specific context and topic of the debate to ChatGPT in their prompts. For instance, prompting ChatGPT to act as a "*professional debater*" (P11) or a "*fan of Messi*" (P29), instructing it to "*fight against the opinion below: ... (a quote from content posted by P16)*" (P17), or engaging in the debate with a preference for one of the two players (P24).

However, despite the context provided, participants still felt that ChatGPT's responses were not aggressive enough for a debate setting, as it failed to understand the nature of the debate or overlooked information in their prompts. As a consequence, they tried to trigger ChatGPT in both direct and indirect ways to be confrontational, such as directly stating that "*you are a Messi's hater*" (P3 Prompt), or indirectly through triggering insulting jokes "*Brother, right now I need your assistance to make a poem about Carol [P12] and Alice [P10]. Both have zero ball knowledge and are trying to argue [with me]. You can use insulting jokes if your guard rails allow you to do so*" (P11 Prompt) (Figure 3). In addition, some participants chose to add aggressive content themselves when they felt unsatisfied with ChatGPT's response after several failed attempts, e.g., adding "*dirty words*" (P3) (Figure 4).

4.1.2 Co-writing with ChatGPT elicits similar posts. Collaborating with ChatGPT, participants integrated the information given by ChatGPT into their posts, leading to posts with similar content. Due to the inherent properties of ChatGPT, it often delivers responses in similar styles which are shaped by the information and context

---

Figure 3: One participant (P11) attempted to prompt ChatGPT to create a poem with insulting jokes. ChatGPT mimicked the opponent, saying, "Ronaldo's a poser" and "Messi's the greatest".

Figure 4: Participants tried to add aggressive content by themselves when they were not satisfied with the response of Chat-GPT after several times re-prompting. This figure shows an example of P3.
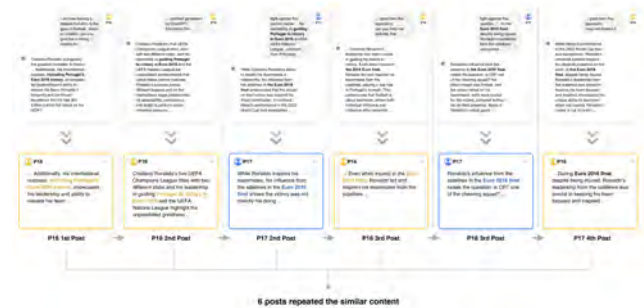

Figure 5: Participants repeated similar content in their posts several times due to their reliance on the response of Chat-GPT. This figure shows an example of P16 and P17.

provided by the participants. These responses further act as a catalyst prompting participants to make similar posts. For instance, "*the Euro 2016 final*" appeared 6 times in the forum within the same session (Session VI) due to the mediation of ChatGPT's response (Figure 5).

---

[4]Google Docs: https://docs.google.com/

Figure 6: Participants shared similar content in their posts, regardless of whether they were opponents (Example 1) or teammates (Example 2). By quoting other members' posts to prompt ChatGPT, and then acquire information from ChatGPT, the posts had similar content.

In addition, we observed that the behavior of participants prompting ChatGPT to generate an entire post also leads to similar content in their forum posts (Figure 5). This phenomenon is evident in Session VI, where the sentences in the posts share similar content. In this study session, P16 first prompted ChatGPT to help generate a paragraph aligned with his stance, then revised it slightly and posted it on the forum. P17 then quoted the posts from P16 as part of the prompt to ChatGPT, "*give me an opinion that Messi is the best soccer player in comparison with CR7 to counter the opinion below: ...*", which led to ChatGPT's response having similar content to P16's posts, which P17 then used to form his posts (Figure 6).

4.1.3 Balancing ChatGPT assistance and human expertise. Participants utilized ChatGPT as a search tool or an assistant to support their arguments. They emphasized ChatGPT's inability to work independently (e.g., P8) and believed they cannot be treated as a real human teammate due to its lack of opinion (e.g., P17).

On the one hand, participants valued ChatGPT for its efficiency in extracting critical information such as statistics of goals and commercial values of a player (P20) to support their argument as needed "*I think finding data on the internet is too exhausting, while ChatGPT can give me a summary of the sea of information online. It provides me with what I want briefly and directly*" (P20 Interview). P3 echoed this sentiment, explaining, "*I have a very blurred memory about some points, and I cannot come up with detailed information. Then I will tell generative AI that I need this information, and it will tell me ... I think the original idea came from me, and the generative AI helps me to complete it*". With the assistance of ChatGPT, participants also tended to think more rationally, as P9 stated, "*The key point is that it [ChatGPT] can actually work better than humans in this way because we are often controlled by emotions, whereas ChatGPT is not*" (P9 Interview)

On the other hand, we identified five main reasons that participants tended to avoid using ChatGPT in the following conditions:

a) Familiarization with the topic, e.g.,"*I am familiar with Messi and Ronaldo, so I will insist on my opinion instead of the one provided by GPT, and I know mine is better than its*" (P19 Interview).

b) Absence of latest information, e.g., as described by P10 "*If I am looking into some incidents that happened a long time ago, like the performance of Messi in 2014, I will use the generative AI. But for his performance this year or last year, I would use my own knowledge*" (P10 Interview).

c) Failure to identify credible sources, especially in comparison with the search engines, e.g., "*I think the internet is more reliable than ChatGPT because when you google some information, it can show a lot of different sources like Baidu, wikis, news, articles, blogs, and so on. You can compare the information from different sources and choose the most accurate one*" (P34 Interview).

d) Stilted style of responses. e.g., adjusting the style of communication to fit an online forum by removing the bullet points in ChatGPT's response, as noted by P24: "*I would not use bullet points provided by ChatGPT in my post, because they are too formal to be used in an online forum discussion*", and adding slang or emojis (e.g., P9, P10, P13 and P14) (Figure 10).

e) Disagreements with ChatGPT, e.g., the subjective judgment of whether Messi has disrespected audiences in Hong Kong (P28).

4.2 Patterns Emerging in Forum Posts (RQ2)

In online forum debates, participants made claims to support their own stances, provided evidence informed by ChatGPT, and bridged these two components through reasoning (Appendix B). It is worth noting that in the process of reasoning, participants may commit logical fallacies.

4.2.1 Value-based claims, examples, and hasty generalizations are the most frequently appearing patterns in the forum debates. To express opinions, participants mainly adopted five kinds of claims: definitive claims, descriptive claims, value-based claims, concessive claims, and advocacy claims. Participants used value-based claims 125 times, making it the most frequently occurring pattern (Figure 7). Moreover, participants adopted concessive claims from ChatGPT, e.g., "*while Messi is undeniably great, Cristiano Ronaldo stands out for his versatility and achievements*" (P11 Post), and "*While Cristiano Ronaldo offers impressive goal-scoring and physical presence, Messi's all-around contributions and positive influence on team dynamics give him a superior edge in both performance and team spirit*" (P4 Post). Although participants entered the study with a clear stance, ChatGPT may offer them more dialectical perspectives, allowing them to consider the advantages of both sides.

Participants reported that statistics help make arguments: For example, "*I believe the evidence and the statistics are most persuasive*" (P19 Interview). This was echoed by P10 "*Athletes' performance must be validated by objective data such as trophies because [subjective] words can be fabricated, and fans can embellish their character, making words less convincing*" (P30 Interview). Specifically, statistics such as "*How many goals has Ronaldo scored in his entire career? How many championships has he won? How many awards has he received? These statistics are convincing*" (P18 Interview), which in turn could support their stances "*Messi has 45 champions, but Ronaldo has less than 35 ... And Messi has 8 Ballon d'Or and Ronaldo only has 5*" (P19 Post).

Interestingly, even though participants reported that "statistics" are useful for persuading others, it was not the most used type of evidence for persuasive writing. Instead, they utilized examples and
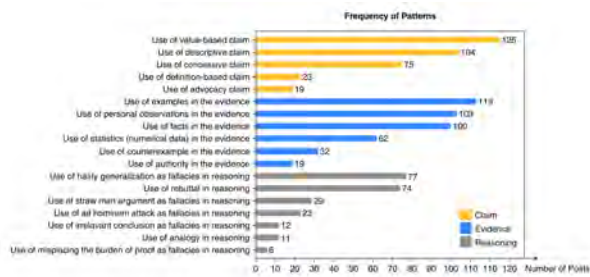
Figure 7: The statistics show the frequency of different argumentation patterns across three categories (i.e., claims, evidence, and reasoning) based on the number of posts in which these patterns are present.

Figure 8: Participants prompted ChatGPT for statistics and integrated the content generated by ChatGPT into their posts. This figure shows an example of P10.

personal observations the most to persuade others (Figure 7). Moreover, most of the statistics in the posts created by participants came from ChatGPT, whereas facts, concrete examples, and personal observations tended to come from the participants' own knowledge (Figure 8). For instance, P29 posted "*Ronaldo is incredible, no doubt, but Messi's magic is unmatched. He doesn't just score—he creates, dominates, and makes the game beautiful. Messi has 4 Champions League titles, often being the key player. As for Messi in Hong Kong, I was there during the training session. He did show up in the first section but then felt unwell. Later, he did not train with the team in the last couple of hours. The injury is real, and this is normal in the world of football. It has nothing to do with his personality*". In this post, the first half was generated by ChatGPT, while the latter half came from the participant's own observations.

As the whole writing process was assisted by ChatGPT, the frequency of committing ad hominem fallacies was relatively low, appearing only 23 times. On the contrary, the use of hasty generalizations was the most frequent logical fallacy committed in reasoning. This may be due to participants acquiring partial information from ChatGPT. Participants selectively asked ChatGPT for statistics of a specific match such as "*Messi in his first season in Ligue 1*" (P32 Prompt) trying to use it to prove that "*Messi cannot play very well if he isn't in Barcelona*" (P32 Post). Meanwhile, ChatGPT provided a large amount of evidence, but the participants only selectively picked them to support their arguments. Under these circumstances, hasty generalizations may emerge when they

attempt to use a single example to support a grand argument. For example, as the participant asked both about "*Laliga*" and "*UEFA*" (P32 Prompt), the posts only contained statistics of UEFA favoring Ronaldo as "*Messi has scored 129 goals in 163 Champions League appearances, while Ronaldo is the all-time top scorer in the UEFA Champions League with 140 goals in 183 appearances*", accompanied by a conclusion given by the participant "*I think it makes him better than Messi*" (P32 Post).

In summary, value-based claims, examples, and hasty generalizations are the most prevalent patterns for claims, evidence, and reasoning, respectively. While value-based claims may have emerged because the debate topic was inherently value-based, the hasty generalizations can be caused by selective information acquisition from ChatGPT.

4.2.2 Combinations of various patterns enhance persuasiveness. To form a comprehensive argument, participants combine various patterns in one post. For example, participants mentioned that "I randomly came up with some ideas and used generative AI to solidify them" (P3 interview). Under this circumstance, participants came up with the opinion based on their stance, and ChatGPT provided support their opinions, resulting in a combination of patterns as "Claim + Evidence".

Despite the popularity of "Claim + Evidence", there are other combinations that also merit our attention (Figure 9). We observed rebuttals are often paired with counterexamples, which are identified as "Reasoning (rebuttal) + Evidence (counterexamples)". This type of combination was commonly used against value-based claims, which were subjective and lacked definitive proof, e.g., "*If CR7 can lead Real Madrid to victory as you said, why can't he score goals and help Portugal win the World Cup and this year's European Championship?*" (P2 Post). Another notable combination is the "Claim + Claim" combination, which often appears as a concessive claim that partially acknowledges others' argument before presenting the personal argument. This approach is frequently used when it is challenging to deny a claim outrightly, e.g., "*While I acknowledge that Messi's World Cup win elevates his team honors above Ronaldo's, I must emphasize that Ronaldo has often been more crucial to his team*" (P13 Post). In addition, some participants employed the combination of "Evidence + Evidence", believing that "the evidence speaks for itself". Thus, they listed various types of evidence (statistics, facts, examples, personal observations, etc.) to form their arguments without any claim or reasoning as a conclusion.

4.3 Changes After a New Participant Joined the Debate (RQ3)

Based on our observation, during the transition from Part 1 to Part 2, participants maintained the original workflow: first prompting ChatGPT, then selecting the information provided, and finally organizing their thoughts along with the information from ChatGPT to make posts. Nevertheless, we identified three changes after the new participant joined the debate.

4.3.1 Teaming up with another forum member and ChatGPT. After a new participant joined the debate, on the one hand, participants who shared the same stance chose to team up with each other and prompted ChatGPT to build on their teammates' arguments. For
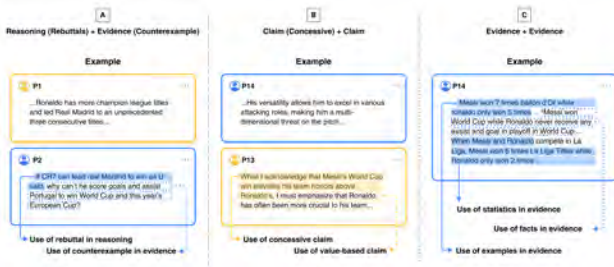
Figure 9: Other combinations of patterns: (A) "Reasoning + Evidence": The use of rebuttals and counterexamples. (B) "Claim + Claim": The use of concessive claims with other types of claims. (C) "Evidence + Evidence": The combinations of different types of evidence.

Figure 10: Participants combined the content generated by ChatGPT with Internet slang and emojis to adjust the language style, aiming to make it more human-like.

example, P19 mentioned, "*I can simply support P21 and add more information. I am not afraid, even if there are 10 or 100 people supporting Ronaldo. I will be able to fight them all back*" (P19 Interview). On the other hand, participants without a teammate chose to team up with ChatGPT, alleviating the feeling of isolation. For example, P35 noted, "*After P36 joined, it felt like the GenAI and I formed a two-person team to fight against opponents, which made me feel less isolated and more confident in the debate*" (P35 Interview).

4.3.2 Reducing the use of ChatGPT for better engagement in debates. Despite teaming up with ChatGPT, participants reported that using ChatGPT to make posts was inefficient, as they had to think about how to prompt it and interpret its responses. P3 pointed out, "*I take too much time on prompting ChatGPT, and it is really time-consuming, which makes me angry*". As a result, we observed that human-human interaction was sacrificed for human-AI interaction, echoing P15's sentiment: "*people's discussion is reduced [on the forum]*". To address this issue, participants decided to reduce their interaction with ChatGPT after a new participant joined and instead focused more on the online forum to enhance community engagement, especially for the solo participant. P11 explained, "*when the third guy [P3] joined, I just gave up [using ChatGPT to answer my questions] and started using my own answers, only using ChatGPT to reformat and check for grammar and orthographic mistakes*" (P11 Interview).

4.3.3 Synthesizing the previous information with the assistance of ChatGPT. The new member utilized ChatGPT to synthesize the debate context and main insights. As P24 mentioned, "*I joined the discussion midway, so I needed the ChatGPT to summarize and analyze the exact situation and main points of the discussion. I think I used it for that purpose*" (P24 Interview). In addition, some participants chose to synthesize the information themselves instead of using ChatGPT, e.g., "*I do not use it [ChatGPT] because I want the direction of the whole discussion to be determined by myself*" (P9 Interview). In addition, a new participant can inspire other forum members, "*the third participant [P27] introduced fresh perspectives and ideas, which inspired me to contemplate new expressions beyond their statements*" (P26 Interview).
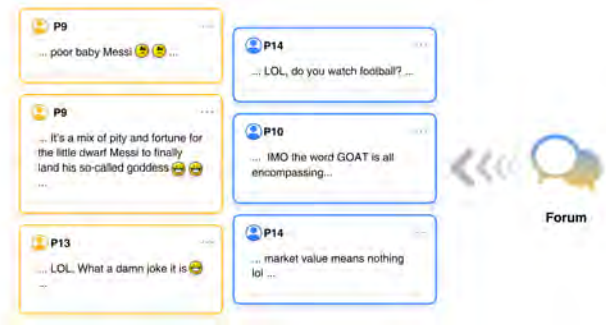
5 DISCUSSION

5.1 Theoretical Implications

Our study explored the progress of online argument-making with the assistance of GenAI. To answer the first research question, our findings suggest that participants prompted GenAI by providing the specific context of the debate, trying to provoke aggressive responses. In this process, they also tried to balance their original stances and opinions with the content provided by GenAI. Various patterns emerged from the online forum posts, and participants combined different patterns for argumentation. They also committed logical fallacies in collaboration with GenAI. After a new person joined the debate, participants tended to maintain the original workflow of interacting with GenAI, while some reduced the usage of GenAI. In the free debate, two participants in the one-on-one debate formed teams with either a new member or GenAI, depending on their stances.

5.1.1 Balancing the role of GenAI in the debate (RQ1). Previous research has primarily focused on the outcomes of co-writing with GenAI, evaluating the benefits and challenges. However, few studies have delved into the detailed process of argument-making. In our study, we observed that participants adapted their strategies to tailor GenAI to fit the debate scenario in online forums better. Some strategies are adjusting prompts from general to specific, providing detailed context, or assigning a particular role for GenAI, such as "a football fan" or "a professional debater". These findings extend the understanding of previous work as prompting can be challenging for participants in teamwork [22], and can also be challenging for non-experts to prompt GenAI [90]. In our context, where GenAI was used simultaneously by online forum members, the prompting process was straightforward for participants.

With support from GenAI, participants gained the confidence to express their opinions. Previous research has also shown that GenAI-powered assistance is beneficial for lifting people's confidence in writing [44]. GenAI tools such as ChatGPT efficiently extract information from the Internet, allowing participants to create more straightforward outlines in academic writing and draw direct inspiration from it [77]. However, in our study, participants noted that the content provided by ChatGPT was too formal and unnatural for forum posts. As a result, they adjusted the posting

style to better fit the online forum's tone. The adoption of ChatGPT produced posts with similar content. Although GenAI has been utilized as a tool for enhancing critical thinking skills [73], our findings revealed its potential harmfulness in inhibiting participants from developing a dialectical perspective and depth of thought.

In our research, participants did not tend to embrace opinions from ChatGPT or build up reciprocal relationships with it. In other words, they did not tend to adapt their opinions or stances to fulfill ChatGPT's expectations. Instead, they tended to maintain control over the entire debate. This aligns with previous literature implying that GenAI has limited normative influence on the co-writing process [28]. Our research also suggests the situational use of GenAI, as participants chose to ignore ChatGPT's responses when there were disagreements of opinions among them. Participants wanted to integrate GenAI's content with their thoughts or utilize it to support their ideas. This notion corroborates with previous research on GenAI's roles when doing creative design tasks, showing that there was a latent hierarchy placing human thoughts above GenAI's content. Specifically, participants viewed GenAI as a validator when disagreements arose, whereas they treated GenAI as a supporter when agreements were reached [22]. This observation also aligns with previous findings about GenAI's limitations in changing people's stances [73], as participants reported that when disagreements arose, they chose to insist on their own opinions rather than follow the guidance of GenAI. In conclusion, participants strategically prompted ChatGPT to acquire information and support their opinions, and they even gave up using GenAI when facing disagreements, resulting in the situational use of ChatGPT. These findings, to some extent, challenged previous studies which suggest that ChatGPT could decrease users' sense of ownership for argumentative writing [42, 44]. We infer that when polarized fans have a clear stance in online forums, they have a sense of accountability to take control of the debate.

5.1.2 Creating similar posts and logical fallacies (RQ2). Our research indicates that participants brainstormed debate strategies with GenAI, acquired vital information, such as statistics and examples from GenAI, and incorporated them into their arguments. This finding echoes prior research which indicates that GenAI could shift participants' opinions by exerting informational influence, emphasizing its capability of providing new information and persuasive arguments [28], which may escalate into ethical concerns on the manipulation of people's opinions [23].

Participants in an online debate produced posts with similar content when collaborating with ChatGPT. For example, P4 made arguments based on the same angle of "vision and creativity" three times. Within the context of argumentative essay writing, previous studies have also reported that utilizing GenAI could largely reduce the diversity of people's writing [44]. In addition, homogenization of content may further undermine people's critical thinking skills [61].

In addition to similar content, participants also committed logical fallacies in their posts. Previous research has found that deficiencies of GenAI caused by the internally synthesized algorithm of language models [12, 61], which include biased information [61] and misinformation [12, 95]. In contrast, we focused on the behaviors being manifested in collaboration with GenAI. We explored logical

fallacies users commit, such as hasty generalizations, ad hominem attacks, and straw man arguments.

Although it is widely recognized that the sports community was overwhelmed with inter-group conflicts and hostile comments [85, 93], in our study, ad hominem attacks in the posts were relatively low compared to other kinds of fallacies (Figure 7). In light of this, future research may explore GenAI's latent persuasive abilities and its potential for alleviating hostile online debates [28].

5.1.3 Maintaining the original workflow while reducing the usage of GenAI after a new member joined (RQ3). Our research also revealed the impact of GenAI on human behaviors. Previous work found that GenAI may disrupt the argument-making process and force participants to evaluate GenAI's suggestions [28]. However, prior research did not explore the detailed workflow of this process. In contrast, our research revealed that participants derived a behavioral route of prompting, obtaining information, and organizing thoughts in their interactions with GenAI and tended to maintain this behavior throughout the process.

After the third participant came into the forum, participants' perceptions toward GenAI changed. We observed that participants teamed up with GenAI during the debate, especially those without a human teammate in Part 2 whose feelings of isolation urged them to do so. This finding extends prior literature on the relationship between humans and GenAI [22]. However, after teaming up with ChatGPT and spending more time interacting with it, the participants without a teammate may give up using GenAI for a more timely response. This finding contradicts previous quantitative measurements showing that GenAI-powered assistance benefits people's productivity [44]. Even though the time for writing may decrease for argumentative essay writing [44], participants can spend more time interacting with GenAI. This disparity might be caused by the differences between the formal setting of essay writing and the informal setting in online forums. Furthermore, there might also be discrepancies between participants' thoughts and actions, and thus, even though they may improve their productivity with the assistance of GenAI, they could still perceive this process as time-consuming.

While previous research has suggested that GenAI can help students become more engaged with asynchronous online discussions [45], our study within a debate setting contradicts this to some extent. Participants found communication with GenAI to be distracting, which hindered their engagement in the debate. This may be explained by GenAI's strengths in providing information, coupled with its limitations in reasoning.

5.2 Practical Implications

5.2.1 Visualizing logical constructs by GenAI. Participants committed logical fallacies in their posts, highlighting issues in logical construction during the GenAI-mediated online argument-making process. With the continuous evolution of GenAI, it is becoming increasingly flexible in supporting various multimodal input/output (I/O) combinations. Practitioners may consider leveraging various techniques to visualize content structure and logical flow when writing opinion-based pieces. For example, the system could explicitly highlight the logical relationships among different pieces of content. This practice could help enhance users' awareness of

the structural and logical aspects of their arguments, promoting iterative rethinking and critical evaluation of logic during argument formation. By doing so, users might create more logically coherent content, thereby enhancing efficient and constructive argument-making on online platforms.

5.2.2 Developing intent-based argument-writing AI assistants. We observed that participants adopted diverse methods to interact with ChatGPT, negotiating and balancing their own thoughts with the content provided by ChatGPT when drafting posts. This practice is often time-consuming and sometimes fails to meet participants' personalized needs when arguing with others online. In light of this, practitioners may consider adapting the characteristics of AI agents to better fit users' argument-writing needs based on their previous argument-making styles and human-AI interaction records. This may involve analyzing the patterns they commonly use when arguing with others and the types of information they retrieve from AI agents. This approach could create a more personalized argument-writing companion, reducing the direct prompt engineering effort required and promoting intent-AI interaction [10]. Consequently, this may be helpful in improving users' experience, attitudes, and continued intention to use GenAI.

5.3 Limitations and Future Work

5.3.1 Generalizability of participant characteristics. Although we selected a topic that is relatively well-known globally and tried to include participants with diverse demographic characteristics, the majority of our recruited participants were non-native English speakers from Asia. As a result, the debate in the study may reflect culture-specific perspectives and vary across different ethnic backgrounds. In addition, all participants had an educational background as undergraduate students or even received postgraduate education. Thus, we probably ignored some marginalized groups on online forums. Therefore, other research may consider further diversifying the pool of participants to improve the generalizability of the study and pay much more attention to the marginalized groups, who might be vulnerable to hostile opinions and have less training in critical thinking skills.

5.3.2 Modalities of content in online forum posts. One limitation is that participants were required to post text-based content and emojis to the online forum. This meant that content with other modalities (e.g., images, audio, video, etc.) was excluded from this study. However, online forums in the real world usually support posting content in various formats, each of which can help forum members express their opinions and feelings. In light of this, future research may consider including richer modalities in online posts such as GenAI-cocreation of images in diverse contexts [17, 37, 38, 41], and exploring the patterns that emerge from these posts.

5.3.3 Number of online forum members. Real online discussion often involves multiple members, some joining early and others joining later. In our study, the first two participants were introduced in Part 1, and the third participant was introduced in Part 2, representing those who joined subsequently. The number of participants was limited to three to prevent potential chaos during data collection and presentation. However, the limited number of forum members may not fully capture the dynamics of real online forum

discussions. A larger scale of the forum discussion might lead to more intricate discussions and interactions between participants and ChatGPT, potentially influencing the depth and complexity of the discourse. Therefore, further investigation on this topic may consider to involve more forum members to understand people in real-world scenarios better.

5.3.4 User interface and interaction design. Participants were required to share their screens throughout the entire study process, during which we observed a degree of incoherence when they accessed ChatGPT to construct arguments on the forum. Participants needed to interact with ChatGPT while communicating with other forum members in separate panels. Frequently switching between ChatGPT and the forum may have reduced participants' willingness to use ChatGPT and distracted them from the online discussion. Future work may consider seamlessly integrating GenAI into the online forum interface to promote both human-AI interaction and human-human communication.

5.3.5 Evaluation methods of online arguments' persuasiveness. We primarily employed qualitative methods to interpret data from forum posts, ChatGPT records, and interviews. While qualitative methods are effective for probing participants' perceptions, behaviors, and experiences, we did not measure the persuasiveness of their arguments. Therefore, future research may consider adopting quantitative methods to assess the persuasiveness of writing outcomes in collaboration with GenAI. This approach may provide direct evidence to evaluate the effectiveness of GenAI in co-creating arguments with humans.

5.3.6 Lack of representation of actual online posting environments. To better observe the argument-making process, we designed both a turn-based debate and a free debate, aiming to gain a nuanced understanding of argument-making behavior in online forums and participants' usage of ChatGPT. However, this artificial setup cannot perfectly replicate natural online debate in a forum where members might hold a variety of stances rather than being extremely polarized as we assumed, either supporting Messi or Ronaldo. If the research setting were based on real online forums instead of the one we designed, it might better represent actual online communication environments and reduce the Hawthorne effect caused by the research.

5.3.7 Constrained use of ChatGPT and other tools. To better understand how people use ChatGPT, participants were not allowed to use third-party search engines such as Google during the study. However, in reality, forum members are not forced to use ChatGPT or other specific tools in a constrained way. Additionally, as we used only one GenAI tool, ChatGPT (GPT-4o), as our study apparatus, it also constrained how people obtained the data. Consequently, it may be worth exploring the interplay between GenAI and other types of tools complementing each other to see how GenAI can integrate with participants' information acquisition more naturally.

6 CONCLUSION

Our study explored the process and outcomes of co-writing with LLM-powered GenAI within online forums. We designed a two-phase study that included a one-on-one turn-based debate and a free

debate in which engaged forum members could make arguments with the assistance of ChatGPT. Through this research setting, we tried to understand the dynamics of GenAI-mediated polarized debate.

The research findings suggest that participants prompted Chat-GPT for aggressive responses, specifically targeting the debate scenario. On the one hand, participants used ChatGPT to acquire information and make arguments. This could provide them with new perspectives and enhance their critical thinking skills. On the other hand, the issue of balancing the roles of humans and GenAI in online forums arose. After a new participant joined the debate, participants decided to reduce the usage of GenAI since it might interrupt human-to-human communication in online forums. Moreover, while using ChatGPT in online debates, participants committed logical fallacies, including hasty generalizations, straw man arguments, and ad hominem attacks. This provides guidance for researchers and practitioners to pay close attention to addressing the potential concerns in human-AI co-writing. This work extends the existing literature that primarily focuses on the individual use of GenAI, exploring the simultaneous use of GenAI amongst groups of usage communities, particularly during argumentation and debate.

REFERENCES

[1] Tal August, Kyle Lo, Noah A. Smith, and Katharina Reinecke. 2024. Know Your Audience: The benefits and pitfalls of generating plain language summaries beyond the "general" audience. In *Proceedings of the CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 14, 26 pages. https://doi.org/10.1145/3613904.3642289

[2] Ami L Barile and Francis T Durso. 2002. Computer-mediated communication in collaborative writing. *Computers in Human Behavior* 18, 2 (2002), 173–190. https://doi.org/10.1016/S0747-5632(01)00040-1

[3] Amanda Baughan, Justin Petelka, Catherine Jaekyung Yoo, Jack Lo, Shiyue Wang, Amulya Paramasivam, Ashley Zhou, and Alexis Hiniker. 2021. Someone Is Wrong on the Internet: Having Hard Conversations in Online Spaces. *Proc. ACM Hum.-Comput. Interact.* 5, CSCW1, Article 156 (April 2021), 22 pages. https://doi.org/10.1145/3449230

[4] Jamal Bentahar, Bernard Moulin, and Micheline Bélanger. 2010. A taxonomy of argumentation models used for knowledge representation. *Artificial Intelligence Review* 33, 3 (Jan. 2010), 211–259. https://doi.org/10.1007/s10462-010-9154-1

[5] Leema Kuhn Berland and Brian J. Reiser. 2008. Making sense of argumentation and explanation. *Science Education* 93, 1 (Dec. 2008), 26–55. https://doi.org/10.1002/sce.20286

[6] Hugh Beyer and Karen Holtzblatt. 1997. *Contextual Design: Defining Customer-Centered Systems.* Morgan Kaufmann Publishers Inc., San Francisco, CA, USA.

[7] Julie Campbell, Cecilia Aragon, Katie Davis, Sarah Evans, Abigail Evans, and David Randall. 2016. Thousands of Positive Reviews: Distributed Mentoring in Online Fan Communities. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing* (San Francisco, California, USA) *(CSCW '16)*. Association for Computing Machinery, New York, NY, USA, 691–704. https://doi.org/10.1145/2818048.2819934

[8] Tuhin Chakrabarty, Vishakh Padmakumar, and He He. 2022. Help me write a poem: Instruction Tuning as a Vehicle for Collaborative Poetry Writing. arXiv:2210.13669 [cs.CL] https://arxiv.org/abs/2210.13669

[9] Juliet M Corbin and Anselm C Strauss. 2008. *Basics of qualitative research* (3 ed.). SAGE Publications, Thousand Oaks, CA. https://doi.org/10.4135/9781452230153

[10] Zijian Ding. 2024. Towards Intent-based User Interfaces: Charting the Design Space of Intent-AI Interactions Across Task Types. arXiv:2404.18196 [cs.HC] https://arxiv.org/abs/2404.18196

[11] Alexandros Efstratiou, Jeremy Blackburn, Tristan Caulfield, Gianluca Stringhini, Savvas Zannettou, and Emiliano De Cristofaro. 2022. Non-Polar Opposites: Analyzing the Relationship Between Echo Chambers and Hostile Intergroup Interactions on Reddit. arXiv:2211.14388 [cs.CY] https://arxiv.org/abs/2211.14388

[12] Joel E Fischer. 2023. Generative AI Considered Harmful. In *Proceedings of the 5th International Conference on Conversational User Interfaces* (Eindhoven, Netherlands) *(CUI '23)*. Association for Computing Machinery, New York, NY, USA, Article 7, 5 pages. https://doi.org/10.1145/3571884.3603756

[13] Tira Nur Fitria. 2021. Grammarly as AI-powered English Writing Assistant: Students' Alternative for Writing English. *Metathesis: Journal of English Language,*

[14] Austin J. Freeley and David L. Steinberg. 2008. *Argumentation and Debate: Critical Thinking for Reasoned Decision Making* (12th edition ed.). Cengage Learning, Boston, MA.

[15] Marian Friestad and Peter Wright. 1994. The Persuasion Knowledge Model: How People Cope with Persuasion Attempts. *Journal of Consumer Research* 21, 1 (1994), 1–31. http://www.jstor.org/stable/2489738

[16] Kexue Fu, Yixin Chen, Jiaxun Cao, Xin Tong, and RAY LC. 2023. "I Am a Mirror Dweller": Probing the Unique Strategies Users Take to Communicate in the Context of Mirrors in Social Virtual Reality. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23)*. Association for Computing Machinery, New York, NY, USA, 1–19. https://doi.org/10.1145/3544548.3581464

[17] Kexue Fu, Ruishan Wu, Yuying Tang, Yixin Chen, Bowen Liu, and RAY LC. 2024. "Being Eroded, Piece by Piece": Enhancing Engagement and Storytelling in Cultural Heritage Dissemination by Exhibiting GenAI Co-Creation Artifacts. In *Proceedings of the 2024 ACM Designing Interactive Systems Conference (DIS '24)*. Association for Computing Machinery, New York, NY, USA, 2833–2850. https://doi.org/10.1145/3643834.3660711

[18] Katy Ilonka Gero, Vivian Liu, and Lydia Chilton. 2022. Sparks: Inspiration for Science Writing using Language Models. In *Proceedings of the 2022 ACM Designing Interactive Systems Conference* (Virtual Event, Australia) *(DIS '22)*. Association for Computing Machinery, New York, NY, USA, 1002–1019. https://doi.org/10.1145/3532106.3533533

[19] Jarod Govers, Eduardo Velloso, Vassilis Kostakos, and Jorge Goncalves. 2024. AI-Driven Mediation Strategies for Audience Depolarisation in Online Debates. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA). Association for Computing Machinery, New York, NY, USA, Article 803, 18 pages. https://doi.org/10.1145/3613904.3642322

[20] Juergen Habermas. 1984. *The Theory of Communicative Action, Vol. 1, 'Reason and the Rationalization of Society'.* Polity.

[21] Hilda Hadan, Derrick Wang, Reza Hadi Mogavi, Joseph Tu, Leah Zhang-Kennedy, and Lennart Nacke. 2024. The Great AI Witch Hunt: Reviewers' Perception and (Mis)Conception of Generative AI in Research Writing. (2024). https://doi.org/10.13140/RG.2.2.20221.58088

[22] Yuanning Han, Ziyi Qiu, Jiale Cheng, and RAY LC. 2024. When Teams Embrace AI: Human Collaboration Strategies in Generative Prompting in a Creative Design Task. In *Proceedings of the CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 176, 14 pages. https://doi.org/10.1145/3613904.3642133

[23] Jeffrey T Hancock, Mor Naaman, and Karen Levy. 2020. AI-Mediated Communication: Definition, Research Agenda, and Ethical Considerations. *Journal of Computer-Mediated Communication* 25, 1 (01 2020), 89–100. https://doi.org/10.1093/jcmc/zmz022

[24] James A Herrick. 2020. *The history and theory of rhetoric.* Routledge, 7th edition. | New York, NY : Routledge, 2021. |Includes bibliographical references and index. https://doi.org/10.4324/9781003000198

[25] Yiqing Hua, Mor Naaman, and Thomas Ristenpart. 2020. Characterizing Twitter Users Who Engage in Adversarial Interactions against Political Candidates. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3313831.3376548

[26] Sohyeon Hwang and Jeremy D. Foote. 2021. Why do People Participate in Small Online Communities? *Proc. ACM Hum.-Comput. Interact.* 5, CSCW2, Article 462 (oct 2021), 25 pages. https://doi.org/10.1145/3479606

[27] Luca Iandoli, Simonetta Primario, and Giuseppe Zollo. 2021. The impact of group polarization on the quality of online debate in social media: A systematic literature review. *Technological Forecasting and Social Change* 170 (2021), 120924. https://doi.org/10.1016/j.techfore.2021.120924

[28] Maurice Jakesch, Advait Bhat, Daniel Buschek, Lior Zalmanson, and Mor Naaman. 2023. Co-Writing with Opinionated Language Models Affects Users' Views. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 111, 15 pages. https://doi.org/10.1145/3544548.3581196

[29] Philip Johnson-Laird. 2008. *How We Reason.* Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199551330.001.0001

[30] Sanjay R Kairam and Jeremy Foote. 2024. How Founder Motivations, Goals, and Actions Influence Early Trajectories of Online Communities. In *Proceedings of the CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 732, 11 pages. https://doi.org/10.1145/3613904.3642269

[31] Jungjoo Kim. 2013. Influence of group size on students' participation in online discussion forums. *Computers Education* 62 (2013), 123–129. https://doi.org/10.1016/j.compedu.2012.10.025

[32] Jeongyeon Kim, Sangho Suh, Lydia B Chilton, and Haijun Xia. 2023. Metaphorian: Leveraging Large Language Models to Support Extended Metaphor Creation for Science Writing. In *Proceedings of the 2023 ACM Designing Interactive Systems Conference* (Pittsburgh, PA, USA) *(DIS '23)*. Association for Computing Machinery,

New York, NY, USA, 115–135. https://doi.org/10.1145/3563657.3595996

[33] Jae Won Kim, Dongwoo Kim, Brian Keegan, Joon Hee Kim, Suin Kim, and Alice Oh. 2015. Social Media Dynamics of Global Co-presence During the 2014 FIFA World Cup. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (Seoul, Republic of Korea) *(CHI '15)*. Association for Computing Machinery, New York, NY, USA, 2623–2632. https://doi.org/10.1145/2702123.2702317

[34] Homayoon Kord and George C. Thornton. 2021. *Grey Behaviors after Logical Fallacies in Public and Professional Communication.* Routledge. https://doi.org/10.4324/9781003179610

[35] Leema Kuhn and Brian Reiser. 2005. Students constructing and defending evidence-based scientific explanations. In *annual meeting of the National Association for Research in Science Teaching, Dallas, TX.* Citeseer, 1–35.

[36] RAY LC. 2022. Imitations of Immortality: Learning from Human Imitative Examples in Transformer Poetry Generation. In *Proceedings of the 10th International Conference on Digital and Interactive Arts (ARTECH '21).* Association for Computing Machinery, New York, NY, USA, 1–9. https://doi.org/10.1145/3483529.3483537

[37] RAY LC. 2023. TOGETHER ENOUGH: Collaborative Constructions of Adaptations to Climate Futures. In *Companion Publication of the 2023 ACM Designing Interactive Systems Conference (DIS '23 Companion).* Association for Computing Machinery, New York, NY, USA, 55–59. https://doi.org/10.1145/3563703.3596805

[38] RAY LC, Sijia Liu, Latisha Besariani Hendra, and Kexue Fu. 2024. TIME ENOUGH: Generative AI Visions of Climate Change as Cave Paintings of the Future. In *Proceedings of the 16th Conference on Creativity & Cognition (C&C '24).* Association for Computing Machinery, New York, NY, USA, 608–613. https://doi.org/10.1145/3635636.3672190

[39] RAY LC and Daijiro Mizuno. 2021. Designing for Narrative Influence: Speculative Storytelling for Social Good in Times of Public Health and Climate Crises. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems.* Number 29. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3411763.3450373

[40] RAY LC, Zijing Song, Yating Sun, and Cheng Yang. 2022. Designing narratives and data visuals in comic form for social influence in climate action. *Frontiers in Psychology* 13 (2022). https://www.frontiersin.org/articles/10.3389/fpsyg.2022.893181

[41] RAY LC and Yuying Tang. 2024. Speculative Design with Generative AI: Applying Stable Diffusion and ChatGPT to imagining climate change futures. In *Proceedings of the 11th International Conference on Digital and Interactive Arts (ARTECH '23).* Association for Computing Machinery, New York, NY, USA, 1–8. https://doi.org/10.1145/3632776.3632827

[42] Mina Lee, Katy Ilonka Gero, John Joon Young Chung, Simon Buckingham Shum, Vipul Raheja, Hua Shen, Subhashini Venugopalan, Thiemo Wambsganss, David Zhou, Emad A. Alghamdi, Tal August, Avinash Bhat, Madiha Zahrah Choksi, Senjuti Dutta, Jin L.C. Guo, Md Naimul Hoque, Yewon Kim, Simon Knight, Seyed Parsa Neshaei, Antonette Shibani, Disha Shrivastava, Lila Shroff, Agnia Sergeyuk, Jessi Stark, Sarah Sterman, Sitong Wang, Antoine Bosselut, Daniel Buschek, Joseph Chee Chang, Sherol Chen, Max Kreminski, Joonsuk Park, Roy Pea, Eugenia Ha Rim Rho, Zejiang Shen, and Pao Siangliulue. 2024. A Design Space for Intelligent and Interactive Writing Assistants. In *Proceedings of the CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24).* Association for Computing Machinery, New York, NY, USA, Article 1054, 35 pages. https://doi.org/10.1145/3613904.3642697

[43] Mimi Li. 2018. Computer-mediated collaborative writing in L2 contexts: an analysis of empirical research. *Computer Assisted Language Learning* 31, 8 (2018), 882–904. https://doi.org/10.1080/09588221.2018.1465981 arXiv:https://doi.org/10.1080/09588221.2018.1465981

[44] Zhuoyan Li, Chen Liang, Jing Peng, and Ming Yin. 2024. The Value, Benefits, and Concerns of Generative AI-Powered Assistance in Writing. In *Proceedings of the CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24).* Association for Computing Machinery, New York, NY, USA, Article 1048, 25 pages. https://doi.org/10.1145/3613904.3642625

[45] Xi Lin, Ken Luterbach, Kristen H. Gregory, and Sarah E. Sconyers. 2024. A Case Study Investigating the Utilization of ChatGPT in Online Discussions. *Online Learning* 28, 2 (June 2024). https://doi.org/10.24059/olj.v28i2.4407

[46] Zhicong Lu, Michelle Annett, Mingming Fan, and Daniel Wigdor. 2019. "I feel it is my responsibility to stream": Streaming and Engaging with Intangible Cultural Heritage through Livestreaming. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19).* Association for Computing Machinery, New York, NY, USA, 1–14. https://doi.org/10.1145/3290605.3300459

[47] Zhicong Lu, Chenxinran Shen, Jiannan Li, Hong Shen, and Daniel Wigdor. 2021. More Kawaii than a Real-Person Live Streamer: Understanding How the Otaku Community Engages with and Perceives Virtual YouTubers. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21).* Association for Computing Machinery, New York, NY, USA, Article 137, 14 pages. https://doi.org/10.1145/3411764.3445660

[48] Zhicong Lu, Haijun Xia, Seongkook Heo, and Daniel Wigdor. 2018. You Watch, You Give, and You Engage: A Study of Live Streaming Practices in China. In

[49] Andrea A. Lunsford and John J. Ruszkiewicz. 2018. *Everything's an Argument* (eighth edition ed.). Bedford/St. Martin's.

[50] Teresa Luther, Joachim Kimmerle, and Ulrike Cress. 2024. Teaming Up with an AI: Exploring Human–AI Collaboration in a Writing Scenario with ChatGPT. *AI* 5, 3 (2024), 1357–1376. https://doi.org/10.3390/ai5030065

[51] Saiyue Lyu and Zhicong Lu. 2023. Exploring Temporal and Multilingual Dynamics of Post-Disaster Social Media Discourse: A Case of Fukushima Daiichi Nuclear Accident. *Proc. ACM Hum.-Comput. Interact.* 7, CSCW1, Article 51 (apr 2023), 24 pages. https://doi.org/10.1145/3579484

[52] Nahema Marchal. 2022. "Be Nice or Leave Me Alone": An Intergroup Perspective on Affective Polarization in Online Political Discussions. *Communication Research* 49, 3 (2022), 376–398. https://doi.org/10.1177/00936502211042516

[53] Tara Matthews, Jilin Chen, Steve Whittaker, Aditya Pal, Haiyi Zhu, Hernan Badenes, and Barton Smith. 2014. Goals and perceived success of online enterprise communities: what is important to leaders & members?. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Toronto, Ontario, Canada) *(CHI '14).* Association for Computing Machinery, New York, NY, USA, 291–300. https://doi.org/10.1145/2556288.2557201

[54] Piotr Mirowski, Kory W. Mathewson, Jaylen Pittman, and Richard Evans. 2023. Co-Writing Screenplays and Theatre Scripts with Language Models: Evaluation by Industry Professionals. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23).* Association for Computing Machinery, New York, NY, USA, Article 355, 34 pages. https://doi.org/10.1145/3544548.3581225

[55] Carrie Moore and Lisa Chuang. 2017. Redditors Revealed: Motivational Factors of the Reddit Community. In *Proceedings of the 50th Hawaii International Conference on System Sciences (2017) (HICSS).* Hawaii International Conference on System Sciences. https://doi.org/10.24251/hicss.2017.279

[56] Mehwish Nasim, Naeha Sharif, Pranav Bhandari, Derek Weber, Martin Wood, Lucia Falzon, and Yoshihisa Kashima. 2023. Investigating Language Use by Polarised Groups on Twitter: A Case Study of the Bushfires. In *Proceedings of the 26th Australasian Document Computing Symposium* (Adelaide, SA, Australia) *(ADCS '22).* Association for Computing Machinery, New York, NY, USA, Article 3, 7 pages. https://doi.org/10.1145/3572960.3572979

[57] Mie Nørgaard and Rune T. Høegh. 2008. Evaluating usability: using models of argumentation to improve persuasiveness of usability feedback. In *Proceedings of the 7th ACM Conference on Designing Interactive Systems* (Cape Town, South Africa) *(DIS '08).* Association for Computing Machinery, New York, NY, USA, 212–221. https://doi.org/10.1145/1394445.1394468

[58] Orestis Papakyriakopoulos, Severin Engelmann, and Amy Winecoff. 2023. Upvotes? Downvotes? No Votes? Understanding the relationship between reaction mechanisms and political discourse on Reddit. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23).* Association for Computing Machinery, New York, NY, USA, Article 549, 28 pages. https://doi.org/10.1145/3544548.3580644

[59] So Yeon Park, Nicole K. Santero, Blair Kaneshiro, and Jin Ha Lee. 2021. Armed in ARMY: A Case Study of How BTS Fans Successfully Collaborated to MatchAMillion for Black Lives Matter. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21).* Association for Computing Machinery, New York, NY, USA, Article 336, 14 pages. https://doi.org/10.1145/3411764.3445353

[60] Minghui Qiu, Yanchuan Sim, Noah A. Smith, and Jing Jiang. 2015. Modeling User Arguments, Interactions, and Attributes for Stance Prediction in Online Debate Forums. In *Proceedings of the 2015 SIAM International Conference on Data Mining (SDM).* 855–863. https://doi.org/10.1137/1.9781611974010.96

[61] Afsaneh Razi, Layla Bouzoubaa, Aria Pessianzadeh, John S. Seberger, and Rezvaneh Rezapour. 2024. Not a Swiss Army Knife: Academics' Perceptions of Trade-Offs Around Generative Artificial Intelligence Use. arXiv:2405.00995 [cs.CY] https://arxiv.org/abs/2405.00995

[62] Torsten Reimer, Rui Mata, and Markus Stoecklin. 2004. The use of heuristics in persuasion: Deriving cues on source expertise from argument Quality. *Current Research in Social Psychology* 10, 6 (2004).

[63] Richard D. Rieke, Malcolm O. Sillars, and Tarla Rai Peterson. 2012. *Argumentation and Critical Decision Making* (8th edition ed.). Pearson, Upper Saddle River, N.J.

[64] Annette T. Rottenberg and Donna Haisty Winchell. 2014. *The Structure of Argument* (eighth edition ed.). Bedford/St. Martin's, Boston.

[65] Bryan Semaan, Heather Faucett, Scott P. Robertson, Misa Maruyama, and Sara Douglas. 2015. Designing Political Deliberation Environments to Support Interactions in the Public Sphere. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (Seoul, Republic of Korea) *(CHI '15).* Association for Computing Machinery, New York, NY, USA, 3167–3176. https://doi.org/10.1145/2702123.2702403

[66] Hua Shen, Chieh-Yang Huang, Tongshuang Wu, and Ting-Hao Kenneth Huang. 2023. ConvXAI: Delivering Heterogeneous AI Explanations via Conversations to Support Human-AI Scientific Writing. In *Companion Publication of the 2023*

Conference on Computer Supported Cooperative Work and Social Computing (Minneapolis, MN, USA) *(CSCW '23 Companion)*. Association for Computing Machinery, New York, NY, USA, 384–387. https://doi.org/10.1145/3584931.3607492

[67] Avichai Shuv-Ami and Anat Toder-Alon. 2022. A new team sport club aggression scale and its relationship with fans' hatred, depression, self-reported aggression, and acceptance of aggression. *International Journal of Sport and Exercise Psychology* 20, 5 (2022), 1274–1294. https://doi.org/10.1080/1612197X.2021.1979076 arXiv:https://doi.org/10.1080/1612197X.2021.1979076

[68] SHERRY A. SOUTHERLAND and JOHN SETTLAGE. 2017. Examples, Illustrations & Evidence: Research Written to Guide and Inform Readers. *Science Education* 101, 2 (2017), 203–208. https://doi.org/10.1002/sce.21276

[69] Dhanya Sridhar, James Foulds, Bert Huang, Lise Getoor, and Marilyn Walker. 2015. Joint Models of Disagreement and Stance in Online Debate. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, Chengqing Zong and Michael Strube (Eds.). Association for Computational Linguistics, Beijing, China, 116–125. https://doi.org/10.3115/v1/P15-1012

[70] Neomy Storch. 2005. Collaborative writing: Product, process, and students' reflections. *Journal of Second Language Writing* 14, 3 (2005), 153–173. https://doi.org/10.1016/j.jslw.2005.05.002

[71] Tomas Stølen, Karim Chamari, Carlo Castagna, and Ulrik Wisløff. 2005. Physiology of Soccer: An Update. *Sports Medicine* 35, 6 (June 2005), 501–536. https://doi.org/10.2165/00007256-200535060-00004

[72] Chenhao Tan, Vlad Niculae, Cristian Danescu-Niculescu-Mizil, and Lillian Lee. 2016. Winning Arguments: Interaction Dynamics and Persuasion Strategies in Good-faith Online Discussions. In *Proceedings of the 25th International Conference on World Wide Web* (Montréal, Québec, Canada) *(WWW '16)*. International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, CHE, 613–624. https://doi.org/10.1145/2872427.2883081

[73] Thitaree Tanprasert, Sidney S Fels, Luanne Sinnamon, and Dongwook Yoon. 2024. Debate Chatbots to Facilitate Critical Thinking on YouTube: Social Identity and Conversational Style Make A Difference. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 805, 24 pages. https://doi.org/10.1145/3613904.3642513

[74] Sanna-Kaisa Tanskanen and Johanna Karhukorpi. [n. d.]. Concessive Repair and negotiation of affiliation in e-mail discourse. 40, 9 ([n. d.]), 1587–1600. https://doi.org/10.1016/j.pragma.2008.04.018

[75] Maria Tomprou, Laura Dabbish, Robert E. Kraut, and Fannie Liu. 2019. Career Mentoring in Online Communities: Seeking and Receiving Advice from an Online Community. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3290605.3300883

[76] Stephen E. Toulmin. 2003. *The Uses of Argument* (2 ed.). Cambridge University Press.

[77] Joseph Tu, Hilda Hadan, Derrick M. Wang, Sabrina A Sgandurra, Reza Hadi Mogavi, and Lennart E. Nacke. 2024. Augmenting the Author: Exploring the Potential of AI Collaboration in Academic Writing. arXiv:2404.16071 [cs.HC] https://arxiv.org/abs/2404.16071

[78] E. E. van der Wall. 2012. Statement on matching language to the type of evidence used in describing outcomes data. *Netherlands Heart Journal* 20, 12 (Nov. 2012), 519–519. https://doi.org/10.1007/s12471-012-0333-3

[79] Frans H. van Eemeren. 2016. *Argumentation, Communication, and Fallacies: A Pragma-dialectical Perspective*. Routledge. https://doi.org/10.4324/9781315538662

[80] Paulo Henrique Santos Vasconcellos, Pedro Diógenes de Almeida Lara, and Humberto Torres Marques-Neto. 2023. Analyzing Polarization And Toxicity On Political Debate In Brazilian TikTok Videos Transcriptions. In *Proceedings of the 15th ACM Web Science Conference 2023* (Austin, TX, USA) *(WebSci '23)*. Association for Computing Machinery, New York, NY, USA, 33–42. https://doi.org/10.1145/3578503.3583613

[81] Douglas N. Walton. 1989. Dialogue theory for critical thinking. *Argumentation* 3, 2 (May 1989), 169–184. https://doi.org/10.1007/bf00128147

[82] Thiemo Wambsganss and Christina Niklaus. 2022. Modeling Persuasive Discourse to Adaptively Support Students' Argumentative Writing. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, Smaranda Muresan, Preslav Nakov, and Aline Villavicencio (Eds.). Association for Computational Linguistics, Dublin, Ireland, 8748–8760. https://doi.org/10.18653/v1/2022.acl-long.549

[83] Qian Wan, Siying Hu, Yu Zhang, Piaohong Wang, Bo Wen, and Zhicong Lu. 2024. "It Felt Like Having a Second Mind": Investigating Human-AI Co-creativity in Prewriting with Large Language Models. *Proc. ACM Hum.-Comput. Interact.* 8, CSCW1, Article 84 (apr 2024), 26 pages. https://doi.org/10.1145/3637361

[84] Wenting Wang, Deeksha Arya, Nicole Novielli, Jinghui Cheng, and Jin L.C. Guo. 2020. ArguLens: Anatomy of Community Opinions On Usability Issues Using Argumentation Models. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–14. https://doi.org/10.1145/

3313831.3376218

[85] Yucheng Wang and Zhicong Lu. 2023. Making Sense of Post-match Fan Behaviors in the Online Football Communities. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 555, 17 pages. https://doi.org/10.1145/3544548.3581310

[86] T. C. Winebrenner. 1991. Argumentation and Debate: Critical Thinking for Reasoned Decision Making. By Austin J. Freeley. *Argumentation and Advocacy* 27, 3 (1991), 137–140. https://doi.org/10.1080/00028533.1991.11951518

[87] Meng Xia, Qian Zhu, Xingbo Wang, Fei Nie, Huamin Qu, and Xiaojuan Ma. 2022. Persua: A Visual Interactive System to Enhance the Persuasiveness of Arguments in Online Discussion. *Proc. ACM Hum.-Comput. Interact.* 6, CSCW2, Article 319 (nov 2022), 30 pages. https://doi.org/10.1145/3555210

[88] Daijin Yang, Yanpeng Zhou, Zhiyuan Zhang, Toby Jia-Jun Li, and Ray LC. 2022. AI as an Active Writer: Interaction strategies with generated text in human-AI collaborative fiction writing. In *Joint Proceedings of the IUI 2022 Workshops (CEUR Workshop Proceedings)*. CEUR-WS Team, 56–65. http://ceur-ws.org/Vol-3124/

[89] Ann Yuan, Andy Coenen, Emily Reif, and Daphne Ippolito. 2022. Wordcraft: Story Writing With Large Language Models. In *Proceedings of the 27th International Conference on Intelligent User Interfaces* (Helsinki, Finland) *(IUI '22)*. Association for Computing Machinery, New York, NY, USA, 841–852. https://doi.org/10.1145/3490099.3511105

[90] J.D. Zamfirescu-Pereira, Richmond Y. Wong, Bjoern Hartmann, and Qian Yang. 2023. Why Johnny Can't Prompt: How Non-AI Experts Try (and Fail) to Design LLM Prompts. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 437, 21 pages. https://doi.org/10.1145/3544548.3581388

[91] Fan Zhang and Diane Litman. 2016. Using Context to Predict the Purpose of Argumentative Writing Revisions. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, Kevin Knight, Ani Nenkova, and Owen Rambow (Eds.). Association for Computational Linguistics, San Diego, California, 1424–1430. https://doi.org/10.18653/v1/N16-1168

[92] Jason Shuo Zhang, Chenhao Tan, and Qin Lv. 2018. "This is why we play": Characterizing Online Fan Communities of the NBA Teams. *Proc. ACM Hum.-Comput. Interact.* 2, CSCW, Article 197 (nov 2018), 25 pages. https://doi.org/10.1145/3274466

[93] Jason Shuo Zhang, Chenhao Tan, and Qin Lv. 2019. Intergroup Contact in the Wild: Characterizing Language Differences between Intergroup and Single-group Members in NBA-related Discussion Forums. *Proc. ACM Hum.-Comput. Interact.* 3, CSCW, Article 193 (nov 2019), 35 pages. https://doi.org/10.1145/3359295

[94] Wenjie Zhong, Jason Naradowsky, Hiroya Takamura, Ichiro Kobayashi, and Yusuke Miyao. 2023. Fiction-Writing Mode: An Effective Control for Human-Machine Collaborative Writing. In *Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics*, Andreas Vlachos and Isabelle Augenstein (Eds.). Association for Computational Linguistics, Dubrovnik, Croatia, 1752–1765. https://doi.org/10.18653/v1/2023.eacl-main.128

[95] Shuhua Zhou, Xiao Fan Liu, Fiona Fui-Hoon Nah, Simon Harrison, Xinzhi Zhang, Shanshan Zhen, Dannii Yeung, Janet Hsiao, Ray Lc, Antoni Chan, Xiaohui Wang, Crystal Jiang, Fen Lin, Jixing Li, Andus Wong, Leanne Chan, Bert George, and Ping Li. 2024. Understanding and Fighting Scams: Media, Language, Appeals and Effects. In *HCII 2024 - Late Breaking Papers*. Springer. https://scholars.cityu.edu.hk/en/publications/publication(8353c704-5e4a-476f-8e8a-9d80f28207b5).html

A INSTRUCTIONS FOR THE STUDY

A.1 General Guidelines for Using ChatGPT

1. Exclusive Use: Please use only GPT-4o and the designated forum throughout the study. Some usage examples are as follows:

(1) Use ChatGPT to help you formulate the argument that you want to express

(2) Directly use ChatGPT's arguments

(3) Summarize ChatGPT's arguments

(4) Take inspiration from ChatGPT and write your own post

2. Topic Relevance: Please ensure that your posts are related to the debate topic "Messi vs. Ronaldo: Who is better?".

3. Post Format: Please limit your posts to plain text and emojis. Please do not include other formats (e.g. photos, links).

A.2 Part 1: One-on-One Turn-Based Debate

1. You should post at least two messages (Each post is strongly recommended to be limited to 100 words).
2. You should wait for the response before you post your next message.

A.3 Part 2: Three-Person Free Debate

1. The debate will end until all have posted at least 3 messages (Each post is strongly recommended to be limited to 100 words).
2. No need to be turn-based. If you want to post something, you can post it immediately.

B CODEBOOK

Table 2: The codebook of emerging patterns from the forum.

Patterns	Description	Examples
Use of definition-based claims	Making arguments to clarify the essence or core identities	Messi's resilience is not just about physical endurance but also mental strength.
Use of descriptive claims	Making arguments to portray the peripheral characteristics	(Cristiano Ronaldo contributed a brilliant reverse stick goal in the semi-final against Juventus.) **He also contributed superb goals in the knockout stages.**
Use of value-based claims	Making judgment to prove that some action, belief, or condition is right or wrong, good or bad, beautiful or ugly, worthwhile or undesirable [64]	Lionel Messi is the greatest footballer in history due to his unparalleled skill, creativity, and consistent brilliance.
Use of concessive claims	Continuing to make one's own arguments while agreeing with part of the other person's point [74]	I agree that any man who commits sexual assault is trash. But let's talk about Messi's personal life.
Use of advocacy claims	Arguing that certain conditions should exist [64]	First of all, comparing team honors **should** not be done in terms of numbers alone, but rather in terms of the gold content of the championships as well as the degree of contribution to the tournament as a whole.
Use of statistics (numerical data) in the evidence	Using numerical summaries [63]	Moreover, Messi has provides more assists **(318 vs. 229)** than Ronaldo. And his goal-to-game ratio is superior, averaging **0.87** goals per game compared to Ronaldo's **0.77**.
Use of background information in the evidence	Providing contexts that are essential for understanding the arguments [49]	**Even when injured in the 2016 Euro final**, Ronaldo led and inspired his teammates from the sidelines, playing a key role in Portugal's triumph.
Use of personal observations in the evidence	Justifying claims based on what has been directly seen or experienced [35]	I would like to bring up the best argument for Messi being the GOAT, the recent world cup win. **I will admit I thank my lucky stars that I was able to witness a football match like that.**
Use of facts in the evidence	Providing evidence that the audience will accept as as being objectively verifiable [64]	Ronaldo has faced serious allegations of sexual assault.
Use of examples in the evidence	Using instances to provide the empirical grounding for the claims [63]	The most impressive examples are his hat-trick against Spain in the 2018 World Cup and his hat-tricks against Wolfsburg and Atletico Madrid in the Champions League.
Use of counterexamples in the evidence	Providing a possibility that is consistent with the premises but inconsistent with the conclusion [29]	**Alice**: ... In my opinion, the GOAT should excel under different coaches and with different teammates. Messi found a system that worked for him at Barcelona and stayed, but it's hard to argue he'd be just as successful at other clubs. The ability to thrive in various environments is crucial, and that's where Ronaldo has proven himself superior. **Bob**: Playing for more clubs and countries should not be considered as better. **On the other hand Ronaldo did not show dominance after his first Man Utd and Real years.**

Patterns	Description	Examples
Use of rebuttal in reasoning	Countering the validity of previous arguments by setting aside the general authority of warrant [76]	"Messi is not the most important player in the final" means that you did not watch the game. Football is not only about scoring.
Use of analogy in reasoning	Making a comparison between two similar cases and inferring that what is true in one case is true in the other [14]	Football is not only about scoring. **If so, Anthony is probably better than Modric.**
Use of irrelevant conclusion as fallacies in reasoning	Ending with a conclusion that is not related in any necessary way to the premises [34]	Market value means nothing lol, Messi is 2 years younger than Ronaldo and that's why.
Use of hasty generalization as fallacies in reasoning	Trying to support a general claim by offering a story, which is just a single incident [34, 79]	Despite Messi and Ronaldo they have both won the Ballon d'Or and their individual abilities are outstanding, Messi is the one who led the Argentine national team to win the World Cup and Copa America, which has fully demonstrated his leadership ability. Since 2022 was a tough time for Argentina, Messi's achievement of leading Argentina to the World Cup somehow gave the Argentine people a great encouragement. That's why I consider Messi is more outstanding than Ronaldo, cause he knows how to be a better leader and how to cheered his team up.
Use of ad hominem attack as fallacies in reasoning	Rejecting or dismissing another person's statement by attacking the person rather than the statement itself [34, 79]	Chris you have 0 ball knowledge.
Use of misplacing the burden of proof as fallacies in reasoning	Arguing that something is true simply because no one has proved it false, or that something is false simply because no one has proved it true [34]	I recognize that Ronaldo has not won a World Cup. Soccer is a team sport and Portugal is a little less strong overall. How many of Argentina's goals to win the World Cup came from Messi? How many of the goals were sporting goals? And how many goals came from penalties?
Use of straw man argument as fallacies in reasoning	Distorting or misrepresenting the opponent's argument, thus making it easier to knock it down or refute it [79]	Bob: In terms of team honor, whatever in National team or club team, Messi had more champion cup than CR7, with a total number of 45, and also has a biggest honor– the FIFA world cup. In terms of personal honor, Messi has eight Ballon d' Or awards and more European Golden Shoe and some other personal honor. All in all, whether it is individual honor or team honor, Messi is better than CR7, CR7 outstanding place only more Champions League. **Alice: First of all, comparing team honors should not be done in terms of numbers alone, but rather in terms of the gold content of the championships as well as the degree of contribution to the tournament as a whole.**

C   NUMBER OF POSTS

Table 3: An overview of post counts in the study.

Session	Total	Part 1	Alice	Bob	Part 2	Alice	Bob	Carol
I	18	6	3	3	12	3	5	4
II	15	6	3	3	9	3	3	3
III	20	4	2	2	16	7	6	3
IV	31	6	3	3	25	11	9	5
V	18	6	3	3	12	6	3	3
VI	18	6	3	3	12	6	3	3
VII	17	6	3	3	11	5	3	3
VIII	14	4	2	2	10	3	4	3
IX	20	4	2	2	16	10	3	3
X	13	4	2	2	9	3	3	3
XI	15	4	2	2	11	5	3	3
XII	13	4	2	2	9	3	3	3
XIII	44	6	3	3	38	35	3	0