Collective intelligence in the digital age: facilitating dialogic education on climate science with GPTs

Xiaoxiao Sun

A Thesis in The

Department of

Geography Planning and Environment

Presented in Partial Fulfillment of the

Requirements for the Degree of Master of Science

(Environment) at Concordia University

Montreal, Quebec, Canada

January 2025

CONCORDIA UNIVERSITY

School of Graduate Studies

This is to certify that the thesis

prepared By: Xiaoxiao Sun

Entitled: Collective intelligence in the digital age: facilitating dialogic education on climate science with GPTs

and submitted in partial fulfillment of the requirements for the degree of

Master of Science (Environment)

complies with the regulations of the University and meets the accepted standards with respect to originality and quality.

Signed by the final examining committee:

_____ Chair

Dr. Ann-Louise Davidson

_____ Examiner

Dr. Ann-Louise Davidson

_____ Examiner

Dr. Fenwick McKelvey

_____ Supervisor

Dr. Damon Matthews

Approved by

_____

Dr. Pascale Biron

_____

Dr. Pascale Sicotte

_____ 2025

Defended on: February 10, 2025

Abstract

Collective intelligence in the digital age: facilitating dialogic education on climate science with GPTs

Xiaoxiao Sun

My research aims to investigate to what extent Artificial Intelligence (AI), like Generative pre-trained transformers (GPTs) can be used to facilitate or enhance dialogic education (DE) on complex topics like climate science, particularly in university settings. Discussions involving climate science are often polarizing because they concern not only the sciences, but a multitude of interconnected socio-economic concerns. Consequently, the likelihood of reaching consensus or compromise is complicated by the diversity of perspectives involved. The ability to have constructive conversations on these topics is therefore essential to this process because environmental action and policy decisions are collaborative problem-solving endeavors among individuals from multiple backgrounds and perspectives. My research involved developing a custom GPT (Agora) based on dialogic principles and testing its performance while collecting data on user interactions. Agora was used as part of a focus group study that included 11 participants who engaged with Agora on topics related to climate science. The study also included discussion periods before and after the Agora interactions. Data from surveys, reflection exercises, and chat logs were analyzed to determine the extent to which Agora facilitated DE, and to better understand user experiences and preferences when conversing with Agora. The results of the study showed that Agora was able to facilitate DE by asking questions that encouraged participants to reflect on their beliefs and assumptions. Participants also expressed increased interest in using GPTs for personal applications in the future, despite their initial reservations about AI. The study also revealed that the quality of interactions with Agora depended largely on the participants' expectations, perceptions, motivations, and approaches. Additionally, reflection exercises indicated that many participants currently struggle with having conversations about climate change, including with those in their personal networks. Some of these struggles included barriers to perspective-taking and perspective-getting, emotional regulation, and conflict resolution. As such, participants expressed a strong interest in learning how to better communicate with others on climate topics, and in receiving personalized guidance on how to navigate interpersonal challenges without alienating others. These insights suggest that there is a need for skills development in climate communication, and that a DE approach in combination with AI could be a promising avenue for addressing this need.

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to my supervisor, Dr. Damon Matthews, for his guidance, support, and mentorship throughout my research process. Under his supervision, I was given the freedom to explore and test my ideas, however unconventional or ambitious. My research would not have been possible without his trust in my ability to see it through to completion.

I also thank my thesis committee members, Dr. Ann-Louise Davidson and Dr. Michael Barcomb, for shaping the context for my research. Their expertise in education was an immense source of inspiration, which further refined the methodologies used to answer my research questions. Working with them has also led to my involvement with other research and faculty communities at Concordia University, which have helped me grow both personally and academically.

I would like to acknowledge the support of the Geography Planning & Environment department at Concordia University, which provided resources and facilities that were essential to the completion of this study. I am particularly grateful to Dr. Donny Seto from the Climate Scenarios, Impacts and Modelling (CSIM) Lab, for investing both time and energy toward troubleshooting and solving the technological challenges I encountered while working in the lab. I also extend my appreciation for my colleagues at the CSIM Lab, whose feedback have helped strengthen the basis of my research.

My participation in The Faculty Interest Group (FIG) for Teaching with AI, through The Center for Teaching and Learning (CTL), presented unique insights into faculty perspectives and experiences on using generative AI in the classroom. As well, Concordia's AI for Workplace Practices community hub equipped me with a platform to not only share my research, but to learn from faculty and staff regarding their thoughts and experiences using AI. Much of my background knowledge regarding experiential learning was drawn from my experiences though the Innovation Lab at Concordia, which also informed my knowledge on communication strategies and collaborative learning.

Both Nilushi Kumarasinghe (CODES[1] internship supervisor) and Odile Joblin (LEADS[2] program coordinator) have been a constant source of support and encouragement throughout my academic journey. Their openness to receiving my ideas and allowing me to apply them in real-world settings has been both personally rewarding and academically enlightening. Likewise, I am most appreciative of Dr. Valeria Kebets from the GeminAI program (Applied AI Institute) for her optimism and enthusiasm in supporting my research; her mentorship has made the final stretch of my Master's all the more fulfilling and enjoyable.

Importantly, I thank the participants of this study, whose time and effort contributed to the richness of the data collected. I am grateful for their candor and openness in sharing their experiences, however personal; most of all, for entrusting me with this knowledge. I am fortunate to have learned so much from their unique perspectives, and intend to incorporate their insights into future research endeavors.

Finally, I would like to acknowledge my family and partner for keeping me grounded and motivated not only throughout this research project, but throughout my life as a whole. I am forever grateful for your love and encouragement, which has been an invaluable source of strength and comfort.

[1]CODES: Coalition for Digital Environmental Sustainability

[2]LEADS: Leadership in Environmental and Digital innovation for Sustainability

Contents

1 Introduction

1.a Problems & Misconceptions in Science Communication

Discussions involving complex STEM[3] topics, like climate change, are often polarizing because they concern not only the sphere of STEM, but a multitude of interconnected socio-economic concerns. Because of this, the likelihood of reaching consensus or compromise becomes complicated due to the diversity of perspectives involved. A common misconception in science communication is that simply communicating factual evidence-based information is sufficient to change perspectives, but evidence shows that this is largely unsubstantiated [1], [2], [3], [4], [5], [6], [7]. As with all human interactions, and specifically dialogues, the outcome is dependent on the way in which they are approached and facilitated. Knowledge communication and education is therefore better served when it is based on evidential research that provides insight into human cognition and behavior with respect to learning and communication [8], [9], [10].

Recent attempts to address this problem have primarily focused on appeals to either individual or social values and attitudes in order to steer behavior. Experimental evidence suggests, however, that long-term changes are more likely to occur through reinforcing *habits* and *structural changes* than through attitudinal changes alone. Focusing *primarily* on values and attitudes is unreliable for long-term behavioral changes, given that values and attitudes are subject to constant fluctuations due to various psycho-social factors [1], [2], [3], [4], [5], [6], [7], [11], [12]. In addition, this approach remains largely behaviorist because the ultimate objective is to influence predetermined beliefs and behaviors rather than expanding one's cognitive ability to process and engage critically and deeply with knowledge [8], [10]. Individuals are influenced through a primarily top-down approach that limits their personal agency and free-thinking, based on preconceptions about what constitutes "acceptable" behavior. This therefore poses an important ethical dilemma: Is the purpose of science communication to influence predetermined values, attitudes, and behaviors, or to encourage deep and critical engagement with knowledge? If the former: who determines these values, attitudes, and behaviors, and why are they preferable to others?

[3]STEM: Science, Technology, Mathematics

To reconcile the dilemma between the purpose of science communication and how to approach it ethically, we need only recall the purpose of **scientific enquiry** as a *process open to multiple perspectives and possiblities* [8]. How do we account for the diversity of perspectives involved in these discussions while collectively deriving innovative solutions to the problems faced in our daily lives? One way is by providing individuals with the *skills* necessary to have constructive conversations about issues like climate change and other complex STEM topics. This represents a more *dialogic* approach that is better able to adapt to different contexts.

Research in learning and cognition shows that individuals who develop critical thinking, creative thinking, collaboration, and communication skills (4C's) are more likely to contribute to innovative approaches to problem-solving [8], [13], [14], [15], [16], [17], [18], [19]. Experimental results also show that critical thinking improves problem-based learning performance because it requires a disposition of *open-mindedness* that is a primary factor in improving problem-solving, self-directed learning, lifelong learning, information gathering, and practical executive skills [18], [20]. This open-mindedness is also necessary for encouraging independent thinking and curiosity which, consequently, enables one to think creatively and challenge the social world [18]. Evidence suggests that limiting the scope of varying perspectives in educational settings also limits the development of these skills because this limits the development of open-mindedness [8]. The 4C's are of particular importance to mitigating and preventing the negative impacts of climate-related phenomena, because environmental action and policy decisions are collaborative problem-solving endeavors among individuals from multiple backgrounds and perspectives.

Case studies show that **Dialogic Education** is a viable method for developing the 4C's across multiple contexts, online and offline [8]. Much of the knowledge acquired today is primarily *monologic* because it is dependent on print-based mediums and is delivered in a one-to-many or many-to-one process; knowledge is created, recorded, and disseminated (e.g., news, textbooks, podcasts, lectures) [8]. It is then received passively through reading or listening, with minimal to no reciprocal interaction; its meaning remains frozen at the time it was created [8]. Dialogic knowledge, however, is interactive (e.g., discussions, self-reflective processes) and inhabits the present and future; meaning is co-constructed between different perspectives in a many-to-

many[4] process [8].

Studies show that active learning processes, like dialogic, are more effective for learning and deeper engagement with knowledge [8], [14], [15], [16], [21], [22], [23]. Moreover, studies in climate communication reveal that having conversations about climate change is necessary for building greater understanding and trust around the topic [24]. The challenge, however, is in facilitating these conversations in a way that is constructive and engaging for all participants, given that exclusionary attitudes and personal prejudices remain a barrier to constructive dialogue [8], [25], [26].

Science communication therefore has much to gain from research in education, because it is based on evidence regarding how humans learn, communicate, and problem-solve [8], [27]. **Dialogic education**, in particular, aims to 1) create the *context* for learning and thinking well and 2) promote skills building long term [8]. It therefore does not limit one's ability to think independently and creatively, because doing so is necessary for collective intelligence and innovation [8].

If science communication and education aims to present trustworthy evidence-based knowledge, then doing so with the underlying motivation of influencing predetermined values, attitudes, and behaviors–which may or may not be substantiated by fact or evidence–functions counter to this purpose. This effectively reduces the credibility of the science communicated, due to lack of transparency in motives and objectives [28]. Approaching science communication and education from a dialogic framework, however, can empower learners to engage more deeply and critically with knowledge, so that they can determine for themselves the validity of knowledge and make better informed decisions. This is especially important, given the increase in misinformation and disinformation online, often perpetuated by persuasive technologies and "digital nudging" [28], as well as the use of AI to encourage political polarization and social division [5], [6], [29], [30], [31]. Dialogic education can therefore equip learners with the skills needed to neutralize efforts to disinform, provide clarity on misinformation, while also encouraging independent thinking and curiosity [4], [8], [31].

[4]"Many-to-many": where multiple perspectives inform one another through interaction; this can be between any number of individual people or as an internal process where multiple perspectives are negotiated through self reflection.

1.b Historical Lessons in Educational Technology

Since the development of the printing press during the industrial revolution, knowledge dissemination has become increasingly print-based and mass produced, making knowledge more accessible to a wider audience [8], [32], [33]. Despite the many benefits of technological advances for education throughout history, failures occur when their applications lack grounding in efficacious methodologies in instructional design [32], [33]. Print-based knowledge is monologic, as opposed to dialogic, because it is a record of knowledge at the time of its production; its dissemination is thus one-sided and the knowledge contained is received passively [8]. Research in education shows that monologic education on its own is insufficient for deep and sustained learning long term [8], and that attempts to apply monologic education to digital technologies have largely failed [32], [33].

While technology, on its own, can be regarded as a means to enhance existing pedagogical frameworks, it cannot be regarded as a replacement for education. Moreover, instructional design must also be able to adapt to technologies and contexts that they are used in order to account for the limitations of each learning environment [8], [32], [33], [34], [35], [36], [37]. Artificial intelligence has made this more apparent, given the difficulties presented by GPTs regarding not only concerns about plagiarism, but the fact that merely outputting information does not necessarily imply learning [38]. Social media and online gaming have also become major means of learning [39], which poses many potential benefits with respect to social learning and collaboration, but also poses many challenges regarding knowledge efficacy [40].

1.c Adapting Collective Intelligence for the Digital Age

COVID-19 has made the need for online learning platforms more apparent, given large student-to-teacher ratios and massive workloads expected of teachers and teaching assistants [41], [42], [43], [44], [45], [46], [47]. Assistive technologies have also made learning more accessible to those with disabilities and limitations to mobility due to other life challenges [48], [49]. While the benefits of digital technologies are apparent, the difficulty lies in optimizing the way in which they are used, such that they can enhance learning while also reducing teacher workload [49].

Knowledge and communication has also become increasingly digitized and accessible on the internet, making our engagement with knowledge and others more dialogic [8]. This is apparent

with increased usage of online forums and applications for learning and education [50]. These platforms provide a more interactive mode of learning that is not afforded by lecture-based curriculums, and allow learners to interact with a wider variety of individuals around the world with different perspectives. Case studies for online dialogic learning projects show that collaborative learning and communication can be enhanced using digital platforms, provided that the pedagogical framework is dialogic [8]. Adaptation to the changing social and digital landscape is not only possible, but beneficial for education overall.

Generative AIs, such as ChatGPT, has already been adopted as learning tools by many educational institutions and educational platforms, like Khan Academy and Duolingo. Large language models (LLMs), like GPTs however, frequently produce false information (i.e., "digital hallucinations"), because they are currently probabilistic models that produce text based on linguistic pattern recognition. The information produced, is thus often unverifiable because it is fabricated by the model based on linguistic associations rather than verifiability of truth [51], [52], [53], [54], [55], [56].

One possible solution is to use GPTs, not simply as sources of information, but as scaffolding guides toward developing the 4C's. Khan Academy's *Khanmigo* is a virtual tutor, based on ChatGPT's API, that asks questions which prompt learners to arrive at solutions themselves, rather than providing the answers directly [57]. Asking questions that facilitate critical thinking and problem solving is one component to dialogic education, in which GPTs can evidently be used to scaffold learning [8]. *The Community Builder* (CoBi) is a LLM that aims to facilitate small group collaborative learning by analyzing student talk for student-negotiated classroom agreements like respect, equity, community, and thinking [58]. During this process, CoBi outputs real-time visualizations that represent the agreements exhibited during learning tasks [58]. Students then reflect on whether their interactions were consistent with these agreements in relation to CoBi's outputs [58]. Collaboration and communication are also essential to dialogic education [8]. Research shows that collaboration and communication are co-developed alongside critical thinking and creative thinking; together they facilitate creative problem-solving skills [8]. AI scaffolding techniques like these can be further integrated with online learning communities

as a way to augment skills building for collective problem-solving, and by extension, **collective intelligence** [8].

Digital technologies can therefore be used to scaffold all 4C's in order to prepare learners for real-life scenarios in which they are needed. Skills acquired from dialogic education extend beyond formal education and can be applied professionally (e.g., policy, management, administration) and interpersonally (e.g., relationship building, communication) [8].

1.d Research Objectives

The purpose of this study is to determine what conditions can optimize dialogic education, using climate science/climate change as the topic case study. The focus is therefore on the *ways in which a specific mode of learning can enhance skills building for conversations about a topic of study.* The nature of the study is exploratory and aims to provide insight on the potential for using AI in dialogic education. The research questions to be investigated are:

1. How can dialogic education be expanded with through GPTs?
2. How can GPTs facilitate or enhance the development of skills associated with dialogic education?
3. How do users engage with GPTs from a dialogic framework and what are their experiences?

To address question (1), I developed a GPT that is based in the principles of DE.

To address question (2), I collected data on dialogue progression and user interactions from user logs with the GPT. This was used to perform discourse analysis and statistics and provide support for existing literature on learning and skills development through DE.

To address question (3), I collected data on user experience through written exercises and surveys, which will provide insight on participants' personal reflections on their interactions with the GPT.

I hypothesize that actively practicing metacognitive skills, like critical thinking, can enhance participant ability to engage more deeply and constructively with climate science topics in a group setting. It is also hypothesized that this can be transferable to other contexts.

The results of the study may provide additional knowledge on 1) how individuals approach discussions on climate science topics, 2) how individuals engage with GPTs in a dialogic

framework, and 3) how GPTs can facilitate or enhance DE. This may inform future research on human-AI interactions, human receptivity to digital technologies like AI in an educational setting, applications of DE using AI for skills development, and ways to optimize climate education and communication.

The process of developing a GPT may also provide insight on the technological challenges and considerations in developing AI applications for educational purposes. One of these challenges can be the resources and time required to develop a GPT that is capable of facilitating DE, as well as limitations of hardware and software for participants to engage with the GPT; this may provide valuable information on issues of accessibility and inclusivity in AI applications. Additionally, the security measures taken to protect participants' data may also inform future research on data privacy and security in AI applications.

Both the ethical and logistical considerations required for conducting a study of this nature may therefore inform future research on how to better conduct studies with AI in educational settings as a whole. Further studies can be done to build upon this one for a more comprehensive understanding of what is optimal for specific contexts.

2 Literature Review

The following literature review provides a more detailed account of current research in education, communication, cognition, and digital technologies, with focus on dialogic education. I will begin by examining the philosophical and theoretical underpinnings of dialogic education, and how it differs from traditional educational models. I will then explore the principles of dialogic education, the role of dialogic space, and the importance of dialogic approaches in education and communication, supported by case studies. Afterward, I will introduce supporting literature on research paradigms and methodologies, active learning approaches, metacognition, communication, misinformation/disinformation, and technology.

2.a Dialogue and Communication

2.a.a Dialogic Education

Wegerif [8] provides an in-depth explanation of dialogic education and its philosophical groundings alongside ample critique of contemporary pedagogical approaches to instructional design and mainstream education. This includes a deconstruction and critical assessment of the various

schools within constructivist theory, which outlines both their limitations and contributions. Research in neuroscience and cognition provides evidential support for the basis of dialogic education. Case studies of successful dialogic learning projects are cited for additional support for its application overall, and specifically online spaces, given its ability to adapt to the changing social and digital landscape [8]. A case is also made for dialogic education as being foundational for addressing contemporary environmental issues [8]. Because the book was written in 2013, the technological advancements made since then have not been accounted for, but the theoretical basis remains applicable.

Similarly, Fairfield [59] presents a philosophical analysis of the nature of knowledge, dialogue, and education from a hermeneutic perspective, which is compatible with Wegerif's conception of dialogic education. The author also examines the role of the teacher in learning and dialogue, drawing on works by Martin Heidegger, Hans-Georg Gadamer, John Dewey, and other existential and phenomenological philosophers.

Manalo [60] provides a more technical approach to ways in which dialogic principles can be applied to classroom settings. Several case studies are featured that demonstrate application of the principles of dialogic education, such as structured dialogue, meaning construction, questioning, engagement, reflection, skills development, critical thinking, communication, and collaboration. The overarching goal is in facilitating the "community of enquiry" that is characteristic of dialogic learning.

2.a.a.a Dialogic Space

Central to dialogic education is the concept of "dialogic space", which is compatible with but different from traditional understandings of constructivist theory. Dialogic space describes an *experience*, rather than a physical space, which is created through dialogue and fills the gap between different perspectives [8]. Dialogic space thus includes the *internal* space of participants in addition to the external, and is more concerned with the *process of meaning-making* and the quality of this experience, rather than the outcome [8]. Moreover, the purpose of dialogic space is to expand "the capacity to participate in dialogue" rather than simply creating cultural artefacts and meaning [8]. This contrasts with traditional constructivist theory, which is more concerned with the *product* of learning rather than the *process* of learning [8]. Due to its

emphasis on process, dialogic education is thus neither "student-centred" nor "teacher-centred", but rather "participant-centred"; this is because the process of meaning-making is *shared* among participants, and is *open to change* and reinterpretation [8].

As such, dialogic education is differentiated from Socratic dialogue and dialectic [8]. While Socratic dialogue is meant to teach critical thinking through expanding awareness rather than the construction of knowledge, in practice it is more dialectic in nature, where the instructor facilitates the *drawing out* of contradictions [8]. Socrates himself also assumed a position of authority in his dialogues, leading participants toward an assumed "truth", which is antithetical to the principles of dialogic education [8]. Similarly, Hegelian dialectic also presumes a predetermined "truth" that is meant to be revealed through negation and contradiction, which makes dialectic inherently *monologic* in nature [8].

According to Martin Buber, dialogic space is a space of *I-Thou* relationships, where participants are fully present and engaged with one another, and where the *other* is seen as an equal partner in the process of meaning-making [8]. This differs from *I-It* relationships, where the *other* is seen as an object to be manipulated or controlled [8]. Buber's *I-Thou* relationships are therefore *dialogic* in nature, and are characterized by *mutual respect* and *openness* to the *other* [8]. Similarly, Mikhail Bakhtin argues that dialogic space is a space of *polyphony*, where multiple voices are heard and respected, and where the *other* is seen as an equal partner in the process of meaning-making [8].

Lev Vygotsky conceptualized the *zone of proximal development* (ZPD) as a space of *scaffolding* where learning occurs through *collaboration* and *co-construction* of knowledge [8]. The ZPD is therefore a *dialogic space* where the *other* is seen as an equal partner in the process of meaning-making [8]. However, Vygotsky's ZPD is more concerned with the *product* of learning rather than the *process* of learning, and presumes that knowledge is socio-culturally based and internalized only through social interactions [8]. Similarly, Paulo Freire's theories on dialogue also situates learning as necessarily a social and collaborative process and as a *means to an end*, specifically for furthering socialist ideas and values [8]. Both Vygotsky and Freire's theories thus presuppose that learning is always socially, rather than individually mediated [8]. Because Freire's theories position dialogue as a means to socio-political ends, it cannot be thought of as dialogic, but

rather a form of *rhetoric* due to its authoritative voice [8]. Additionally, neuroscientific studies on infants and learning show that learning is not always socially mediated, but can also be *individual* and *internal*; this is particularly true for autistic children [8].

In contrast to both Vygotsky and Freire, Michael Oakshott views *dialogue as an end to itself* [8]. This aligns with much of existential philosophy, which views knowledge as an end to itself and not as a means to an end, and that "individual awakening is just as valuable a part of education as the transmission of *doctrina*" [59]. According to Hans-Georg Gadamer, "Only those teachers who can freely question their own prejudgments, and who have the capacity to imagine the possible, can help students to develop the ability to judge and the confidence to think for themselves." [59]. This is because the role of a teacher is not to transmit knowledge, but rather to impress upon students "a character that unfolds" and is guided not toward "possessing truth", but toward becoming an authentic *thinking being*, otherwise represented though Martin Heidegger's concept of *Bildung* [59]. Likewise, Fairfield recalls Friedrich Nietzsche's critique of university education for positioning itself as the "possessors of truth" rather than as "pursuers of truth", and using knowledge as a means toward careerism rather than treating knowledge as valuable in and of itself [59]. Consistent with this, Oakshott argues that students must be encouraged to acquire their own voice while being open to possible *other* past, present, and future voices [8]. This is consistent with the spirit of scientific inquiry, which is the pursuit of knowledge and truth while holding one's own beliefs and assumptions open to revision when new evidence is presented [8]. The role of the teacher is therefore to guide learners toward embodying qualities characteristic of this spirit.

2.a.a.b Principles of Dialogic Education

At the core of dialogic principles is the necessity of developing self-consciousness and the ability to reflect on one's own thoughts and actions, which is necessary for not only critical thinking, but one's ability to communicate and collaborate with others; self-consciousness opens up the possibility of reflecting on other perspectives [8]. Moreover, *learning to think* is prerequisite to thinking, just as *learning to learn* is prerequisite to learning [8]. An esssential component to *thinking well* is knowing that one's thoughts do not constitute one's identity, nor are thoughts and identify fixed; rather, both are subject to change and reinterpretation [8]. This openness to

multiplicity is what allows learners to come in dialogue with others and with themselves, which is also necessary for imaginative or creative thinking [8].

In summary, the principles of dialogic education are as follows:

1) Meaning is never fixed, but is always open to change and reinterpretation

2) The purpose of dialogic education is dialogue itself

3) Dialogue exists only between different voices (internal and external)

4) Dialogic processes are non-linear and non-teleological

5) Dialogic augments monologic by being in conversation with it

6) Education always requires the internal voices of the learner which are culturally situated

7) Historical/cultural context is always present but not "fixed containers"

8) Every dialogue has both an internal and external dimension

9) Dialogic space is created from the interaction of internal and external voices

10) Dialogic space is centered on "ways of being" and "ways of knowing" as an end in itself [8]

2.a.a.c The 4C's and Commmunity of Enquiry

The principles of dialogic education culminate toward creating a "community of enquiry", where values emerge from communication rather than through external influence, and is influenced by John Dewey's theory of experience and education [8], [9]. This shared enquiry is what allows for the development of critical thinking, creative thinking, collaboration, and communication skills (4C's) [8].

The 4C's therefore facilitate the cultivation of *shared responsibility* toward 1) decision-making, 2) challenging each-others' claims, 3) reasoning in the face of challenges, and 4) reaching agreements for group decision-making. Together, this shifts intra-group relationships from being *leader-centred* to being **dialogically oriented**, and knowledge is co-created through the *process of dialogue* [8]. Central to this is the ability to consider *multiple perspectives* and hold *multiple truths* in one's mind simultaneously, and without prejudice [8]. This allows knowledge to be *open to change* and reinterpretation. Openness to new modes of thinking and perceiving is an

important factor for creative problem-solving, which is an essential skill for addressing complex environmental challenges [8].

2.a.b Challenges to Dialogue & Evidence for Dialogic Education

In order to develop the 4C's, however, providing learning environments conducive to scaffolding these skills is necessary. Case studies for dialogic education show that educational programs designed to cultivate metacognitive skills like critical thinking and self reflection are more effective when they are scaffolded within a structured but multimodal learning context, where participants are guided through the process of dialogue and reflection [8]. *Philosophy for Children*, for example is a program for young children to learn critical thinking skills, while the *Thinking Together* program focuses on scaffolding skills for shared responsibility in decision-making, challenging claims, reasoning in the face of challenges, and reaching agreements for group decision-making [8]. Both programs have been shown to be successful in achieving their learning objectives because they function according to the concept of "teaching *for dialogue* and *through dialogue*", which encourage learners to critically examine information, ask good questions, and work collaboratively toward shared enquiry; all of which are experientially-based and are meant to cultivate intellectual curiosity and autonomy [8]. These results further emphasize the importance of targeted approaches for skills development that are carefully designed specifically for dialogic learning, particularly since *dialogic education is distinct from other forms of dialogue* [8], [60].

Communication studies further emphasize that *the way in which one engages in dialogue* is often the determining factor for the success of the dialogue itself, rather than the content of dialogue [17], [25], [26]. Studies by Kalla and Brockman [25], [26] show that non-judgemental dialogue that exercise perspective-getting are more effective at reducing exclusionary attitudes and personal prejudices. The reason for this is that *perspective-getting* requires not only the ability to imagine another's perspective, but to actively elicit knowledge from the other person, which requires a certain level of curiosity and willingness to understand the other person's point of view [25], [26]. This is distinct from simply reading about and imagining experiences, which is a more passive form of perspective-taking [25], [26]. Similarly, Evagorou and Osborne [17] show

that successful engagement in argumentation is determined by the extent to which students engaged in exploratory talk and whether they were motivated toward shared understanding.

The mechanisms which underlie dialogic education are therefore better equipped to tackle issues of socio-political polarization, which are exacerbated by misinformation and disinformation. In order to do this, however, it is necessary to shift the focus of education from being *content-centered* to being *process-centered*, and move away from top-down knowledge transmission toward one that is dialogically oriented, with the purpose of encouraging intellectual curiosity and autonomy [8]. The following supporting literature will provide additional evidence for the efficacy of dialogic education in addressing contemporary challenges in education, communication, and technology.

2.b Research Paradigms & Methodologies

Refs. [9], [10], [27], and [61], concern the philosophical groundings for various instructional design methodologies. Ertmer and Newby [10] describe in detail the history and definition of behaviorism, cognitivism, and constructivism in instructional design theory. The benefits and limitations of behaviorism and cognitivism are outlined first, then the authors conclude with the rationale for why constructivism has become preferable in contemporary research and design. The ontological and epistemological frameworks developed by John Dewey in [9] for education underpin much of contemporary instructional design theory and practice, and therefore provides historical context for the methodologies chosen for this thesis. Dewey provides a critical analysis of institutionalized educational systems and their long term impacts on learning and society. Topics discussed include experiential learning, scientific teaching, social control, intellectual freedom, the purpose of education, and teaching science. While the text was published in 1938, contemporary research in education has since provided evidential support for many of Dewey's arguments, as the following articles will show.

Handelsman et al. [61] contains a general guide for scientific teaching and the rationale behind its methodologies and methods with respect to assessment and curriculum development; example methods are also provided for learning exercises. Dirksen [27] is a handbook for instructional design supported by research in learning and cognition. Design concepts are explained in a simplified format as a step-by-step guide toward how to develop course structures

for 1) knowledge, 2) skills, 3) motivation, 4) and environment. It also offers a practical guide for how to tackle issues such as attention and memory, based in research on neuroscientific case studies.

2.c Active, Problem-based, and Project-based Learning

Refs. [14], [15], [16], and [23], discuss active learning design principles, methodologies, and case studies. Active learning (AL) derives from Dewey's experiential learning principles and includes problem/challenge-based learning (PBL). Active learning in general is the proposed method of learning in all four papers. Reilly and Reeves [23] and Torres-Barreto et al. [16] concern research on case studies for AL and challenge-based learning and propose ways to improve online learning and skills building for problem-solving in STEM (e.g., "oral expression, communication, resource management, leadership and problem solving" [16]), respectively.

Kuffner and Walker [62] propose project-based multimodal curriculums as a way to develop communication and collaboration for statistics students. Dewey's experiential learning principles are applied to classroom activities meant to build the skills necessary for students to work collaboratively in a team environment. It is argued that many internships provide insufficient experiential knowledge due to time constraints and lack of scope [62]. Project-based learning is proposed as a more effective way to build the skills necessary for learners to engage in real-world scenarios; techniques employed include interviews, observations, reflections, role-playing, discussions, brainstorming, case-based teaching, presentations, giving and receiving constructive feedback [62].

Seibert [14] examines higher order critical thinking in generation Z nursing students as a case study, and argues that PBL can improve critical thinking and perseverance. PBL has been shown in nursing students to improve student satisfaction, communication skills, and critical thinking, because it pushes students out of their comfort zone and allows them to deal wth uncertainty [14].

Tarrant and Thiele [15] focus specifically on **sustainability education** with respect to Dewey's pedagogical theory, and argues that the skills learned under this framework (e.g., critical thinking, systems thinking, communication, collaboration) are essential for facing current and future environmental issues. The authors argue that the integration of Dewey's democratic

pedagogy can empower individuals to become "lifelong learners and skillful stewards" toward sustainability [15].

2.d Critical Thinking, Open-mindedness, and Argumentation

Refs. [17], [18], [19], [20], [62], [63], and [64] all concern the importance of critical thinking (CT) in the learning process as well as communication and collaborative skills development. Mulnix [63] draws from both philosophical theories on critical thinking and cognitive science to analyze various conceptions of what CT means and how it is learned, and argues that CT is an essential skill regardless of one's discipline. Proposals are made for what is argued to be the most effective means of teaching critical thinking through an experiential learning lens. Luff [18] proposes Dewey's experiential learning principles for **early childhood education for sustainability** (ECEfS) as it pertains to developing curiosity, critical thinking, environmental knowledge, democracy, and participation. AL is cited to be a primary driver of these skills, particularly when integrated with *place-based education* which extends beyond a local sense of place that stresses the interconnectedness of the local and global [18].

Cole [64] examines the effectiveness of the Alexander Technique (AT), supported by Dewey and Marjorie Barstow, in the performing arts. AT, as interpreted by Barstow, includes sustained experiential learning that is interactive and self-reflective in order to better organize and generate new ideas; processes which are important aspects of critical thinking. The author argues that critical thinking and the scientific method are applicable skills for not only philosophy and science, but also the arts, specifically to guide action and practice [64]. Modifications to traditional AT methods are also suggested as a way to enhance learning in the performing arts specifically.

Pu [20] is a case study on third-year undergraduate medical students, and shows how CT can improve PBL performance. The study found that high CT dispositions also correlate positively with an open-minded disposition. Refs. [19] and [17] are case studies on collaborative argumentation, critical thinking, and problem-solving when discussing socioscientific issues. These two studies are especially notable for informing contemporary curriculum development, where the learning context is predominantly subject-oriented and assumes that learners already have the skills necessary to engage constructively with knowledge.

Evagorou and Osborne [17] is an exploratory study that examines how certain argumentative and interactive characteristics between pairs of students might influence the outcome of the activity. It was found that the topic itself did not afford succeessful engagement with argumentation; instead, successful engagement was determined by the individual students involved. Successful pairs engaged in more exploratory talk and shared understanding, they asked each other more questions to clarify on evidence (essential for critical thinking), constructed higher level written arguments, and proposed more solutions for claims [17]. The findings suggest that simply providing students with a socioscientific context does not necessitate engagement with argumentation; student engagement and ownership in argumentation must therefore be scaffolded in order to teach the skills necessary to facilitate discussion. Socioscientifc contexts by themselves, consequently, cannot guarantee 1) collaborative argumentation nor 2) negotiation of shared understanding [17].

Rudsberg [19] proposes *transactional argumentation analysis* (TAA) as a method to investigate classroom learning processes on socioscientific issues. TAA integrates Dewey's *meaning making* and Toulmin's argument pattern analysis which 1) analyzes the direction of meaning making, 2) clarifies meanings as argumentative elements, and 3) compares how arguments are constructed with respect to the content throughout the learning process [19]. The authors identify the theoretical basis for a pragmatic approach to learning before providing a step-by-step description of the TAA process in deconstructing argumentation and argument progression, then provide examples for how learning unfolds through argumentation [19]. The topic of discussion concern **solving environmental problems** set in Swedish upper secondary school. The key finding was that regardless of the arguments posed, all students had already agreed to a *general warrant* of "solving environmental problems is treated as a desirable value" [19], which had become a shared value among the group, making all proposed arguments aligned toward a shared goal. Thus, while socioscientific topics themsevles may not guarantee engagement, the framing of the general warrant as a problem-based learning exercise did in fact facilitate engagement from all participants, despite the difference in arguments and solutions posed.

2.d.a Cognitive and Neuroscientific Perspectives

Refs. [65], [66], [67], provide neuroscientific evidence and phenomelogical interpretations of human reasoning processes. Evans [66] and Evans [65] offer detailed explanations on dual process theories of reasoning (Type 1 and Type 2), which provides insight on current understandings on how humans learn and process information. Haran et al. [21] and Mirhoseini et al. [22] discuss the importance of *actively open-minded thinking* (AOT) in information acquisiton, accuracy, and calibration. It was found in both studies that confidence is negatively correlated with accuracy, and that AOT can decrease confidence while increasing accuracy [21], [22]. It was also found that analytic and systematic processing of information decreased susceptibility to targeted framing, and that those who are more reflective (analagous to Type 2 reasoning [65], [66] and alpha wave indicators [67]) showed greater accuracy in informaton processing [21]. EEG readings from [22] confirmed that greater AOT indicators resulted in more Type 2 reasoning and greater accuracy. These studies offer additional support for instructional design methods modeled on experiential learning, given that AOT is an important aspect of critical thinking.

Refs. [2], and [7] are quantitative case studies that further support the development of open-minded analytical thinking as a means to counter inaccurate or misleading information. Pennycook and Rand [2] found that individuals who engaged more in analytical thinking when evaluating headlines were less likely to accept misinformation as true, and that those who spent less cognitive effort reflecting on the information were more susceptible to misinformation. Likewise, Swami et al. [7] found that strong beliefs in conspiracy theories correlated with reduced analytic thinking and open-mindedness, and higher levels of intuitive thinking. Additionally, tasks meant to elicit analytical thinking were found to reduce belief in conspiracy theories [7]. These studies provide further support for the importance of developing open-mindedness and analytical thinking skills as a means to counteract the acceptance and potential spread of misinformation and disinformation. Van Bavel and Pereira [4] is a more phenomelogical account, supported by reearch in cognition, of how political identities form and influence belief in misinformation. The authors argue that the way in which individuals conceptualize their identities inform how they process information, such as "reasoning, memory, implicit evaluation, and even perception" [4]. Due to the human history of tribalistic cognitive structures, the *identity-*

based model of belief explains why individuals are often willing to abandon personal beliefs and values in favor of maintaining their group identity [4]. Partisanship is thus a result of this need to maintain group identity, and can be viewed as a symptom of underlying neurocognitive processes inherent to human evolutionary history [4]. Because of this, the authors argue that the most effective way to counteract misinformation is to address the underlying cognitive processes that inform belief, rather than the misinformation itself [4]. This means that simply correcting misinformation is inadequate for changing beliefs, and that a more effective approach entails activating alternative or "superordinate" ways in which individuals conceptualize their identities [4]. Interestingly, the authors cite a recent study that found that individuals who were more likely to show a curiosity toward scientific information as a means of experiencing "awe and surprise" were negatively correlated with partisan polarization [4]. This corresponds to the above literature on the importance of open-mindedness and analytical thinking in information processing, and with Wegerif's critique of positioning social identities at the forefront of learning; that "learning should be understood as a trajectory of identity within a social construct", but formation of social identity should not be the purpose of learning [8].

2.e Misinformation, Disinformation, and Polarization

Refs. [68], [29], [30], [5], [68], [6], and [31] cover topics such as misinformation, disinformation, political polarization, and the impact of media on public perceptions and discourse.

Stacks et al. [68] analyzes two methods of media effects research; one that focuses on the impact of the different types of media on audiences (traditional), and the other on how media is used to influcnece attitudes and beliefs (contemporary). Traditional models show *the ways in which* that media is used to influence decision making and capacity to advocate for certain viewpoints, which can lead to long term societal changes. Contemporary models show *how* these mechanisms of change occur, and *how* they are actively employed by media for the purpose of influencing societal changes.

Bogart and Lees [29] and Serrano-Puche [5] examine the internal mechanisms of how misinformation and disinformation affect individual cognition and emotions. Bogart and Lees review literature that examine the ways that misinformation is perceived across groups, and how intra and intergroup dynamics can influence the spread of misinformation. The findings show that

the spread of misinformation is often mediated and enabled by social pressures and not entirely dependent on individual decision making [29]. Serrano-Puche focus more on the phenomenon of affective polarization, which is a consequence of misinformation and disinformation. The authors argue that the spread of misinformation and disinformation can lead to the polarization of public opinion, exacerbating social divisions and conflicts [5]. They also propose interventions meant to mitigate these effects, such as increased efforts toward media literacy and critical thinking education [5]. Similar to Bogart and Lees, the authors emphasize attention toward the emotional component of how information is processed and how mindfulnees techniques can be used to identify attempts at emotional manipulation through media [5]. Finally, they also propose a more collaborative approach to media literacy education that involves the participation of multiple stakeholders, including educators, policymakers, and media organizations [5]. Similarly, Butts et al. [31] propose a mathematical model for understanding the spread of disinformation and misinformation through social networks. The authors argue that the spread of disinformation is not only a result of individual decision making, but also of the structure of social networks and the dynamics of information flow [31]. The authors argue that disinformation is best mitigated through targeted and widespread educational interventions meant to foster critical thinking [31].

Vosoughi et al. [6] and Ferrara [30] focus on how misinformation and disinformation spread through social media and the internet, by examining online behavior, and how they can be used to influence public opinion and discourse. Vosoughi et al. studied data from Twitter to determine the spread of true and false information and found that false information spreads "farther, faster, deeper, and more broadly than truth in all categories of information" [6]. This is because false news is more novel and elicits stronger emotional responses like "fear, disgust, and surprise", while true information was more likely to elicit emotions of "anticipation, sadness, joy, and trust" [6]. They also found that humans, rather than bots, spread false news faster than bots ("rumor cascade"); in fact, bots spread true and false news at the same rate [6]. Ferrara [30] is a case study on the use of disinformation bots during the 2017 French presidential election. The author argues that bots can be used to amplify the spread of disinformation, while creating the illusion of consensus and popularity. The author also found that users who most

engaged with these bots were not actually French citizens, but foreigners, specifically from the United States with connections to alt-right communities [30].

2.f Educational Technology

2.f.a General

The history of instructional design and technology is discussed in Reiser [32] and Reiser [33], with a comprehensive overview of successes and failures in adopting technology for education. The findings support contemporary critiques of pedogogies applied to technology and provides important insight for limitations that exist. However, because the text was written in 2001, much of the technological advancements made since then have not been accounted for. Reiser [37] provides a more updated account of trends in instructional design and technology as of 2018, with a more technical approach to its applications in the classroom based on case studies. These approaches are grounded in contemporary instructuional design theories and methodologies. Similarly, much of the technological advancements made since 2018, specifically with generative AI, is missing from the text. Amiel [36] is a more recent methodological account of educational technologies and argues that design-based research can account for the limitations encountered by integrating educational technologies in the classroom.

Spector et al. [34] is a handbook on research in educational communications, which includes articles on using AI for student evaluation, personalized learning, and educational data mining (EDM). It does not include articles on large language models (LLM) nor GPTs, but provides ample resources on instructional design for technological applications, like augmented reality, with case studies to draw from. Adesope and Rud [35] is also a handbook that contains articles on using technology for education, though only briefly mentions AI as a potentially transformative force in education. The articles focus primarily on ways of promoting collaborative learning, self regulated learning, and project-based learning.

Refs. [40] and [39] are case studies which investigate the process and impact of online learning through social media and online gaming, respectively. While Davidson and Fountain [40] examines participant interactions during learning tasks through the lense of power dymanics within the context of organizational structure, Ruiperez-Valiente et al. [39] does so through

quantitative evaluations of learner engagement within the context of various game characteristics.

2.f.b Artificial Intelligence

Churi et al. [38] focuses specifically on using AI for education with articles on LLMs and GPTs. The book consists of a more methodological overview of how AI can be used and does not delve deeper into the technical aspects of how AI and GPTs work nor how they can be modified for the intended purposes. The articles, however, provide case studies on intelligent tutoring systems that have been developed since 2022, and discusses how AI can be used for evaluation.

Refs. [52], [54], [56], [57], [69], [70], and [71] all concern the use of GPTs for education and provide case studies for how they are currently implemented along with the costs and benefits associated. Tlili et al. [71], Pavlik [56], and Kasneci et al. [52] explore how GPTs can be applied to education and communication while addressing the fears and ethical concerns raised in contempoary discourse surrounding the use of AI. Lo [69] is a literature review of the impact of ChatGPT on education across multiple applications, examining the benefits and limitations. The author notes that the main concern identified in the literature is that of academic integrity, and calls on universities to develop policies to address this, while providing students and instructors with training on how to use GPTs effectively and responsibly [69].

Pavlik argues that much of the hesitation about AI is based on a lack of understanding about its limitations and capabilities, and that the use of AI should be informed by sound pedagogical practices [56]. Tlili et al. found in their study is that public opinion of AI is largely positive, but that there still remain others who are cautious due to concerns about cheating, plagiarism, accuracy, privacy, and manipulation. The authors argue that responsible development must take into consideration "inclusion, usability, tehcnical aspects, ethics, and best practices" [71]. They assert that "Responsible AI is concerned with the design, implementation and use of ethical, transparent, and accountable AI technology in order to reduce biases, promote fairness and equality, and help facilitate the interpretability and explainability of outcomes" [71]. The authors stress that the best way to mitigate these concerns is by accepting that AI exists and its use is inevitable; as such, it is imperative to equip individuals with the skills necessary to navigate this new digital landscape [71].

Cooper [54] examines the potential of using ChatGPT for science education. The author found through a self-study, that ChatGPT had a tendency of "positioning itself as the ultimate epistemic authority," [54] which poses risk for the spread of misinformation if students are not equipped with the skills necessary to critically evaluate the information provided. The author argues that the use of ChatGPT should be accompanied by a critical pedagogy that teaches students how to evaluate information, and that the use of ChatGPT should be limited to specific tasks that are well-defined and have clear learning objectives [54].

Bearman et al. [72] is a book containing articles on ways that AI can be used for student evaluation and assessment. The articles provide case studies on how AI can be used to evaluate student performance and provide feedback. Chapter 7 in particular explores how AI assisted feedback can be used to faciliate dialogue. The book argues in support of transforming the current assessment system in universities to be more outcome-based and student-centered rather than marks-based, and provides a roadmap for how AI can be used to achieve this with consideration for the ethical challenges. Moreover, the book ultimately aims to present new ways of imagining how university systems themselves can be transformed through the use of AI for assessment and evaluation.

Breideband et al. [58] is a case study on *The Community Builder* (CoBi), which is an LLM that is used to facilitate collaborative learning. CoBi facilitates small group collaborative learning by analyzing student talk for student-negotiated classroom agreements like respect, equity, community, and thinking [58]. During this process, CoBi outputs real-time visualizations that represent the agreements exhibited during learning tasks [58].

Refs. [55], [53] and [51] are technical reports on the development of OpenAI's ChatGPT, while [57] is a case study on how GPT-4 can be used for educational purposes as a personal tutor for various subjects.

2.g Persuasive Technologies

Fogg [28] contains a comprehensive catalogue of persuasive technologies since 2003. While many technological advancements have not been accounted for, the methodologies and technologies described remain relevant and operational today. The text also provides a nuanced explanation on the ethical concerns related to persuasive technologies, and provides a structured guide

for how to address these concerns through investigation of design methods and motivations [28]. Stakeholder analysis is one way in which to analyze the ethics of persuasive technology because it can identify those who aim to benefit from it, what they stand to lose, what they value, and to what extent. The author stresses that one's own values must also be evaluated alongside stakeholder analysis to identify potential moral assumptions that may influence the evaluation [28]. It concludes by arguing that transparency and education about how persuasive technologies work can aid in producing technologies that are more ethically sound [28].

2.h Technical Documentation

Refs. [73], [74], [75] provide technical documentation on resources that were used to develop and deploy Agora.

3 Methodologies

3.a Approach: Mixed Study

I conducted a **mixed methods exploratory study** that included 1) descriptive statistics, 2) sentiment analyses, 3) thematic analysis, 4) discourse analysis. This included the development of custom GPT model, which was tested with participants in a focus group setting.

The focus group provided the necessary environment to collect data on participant experiences with a GPT prototype modeled on principles of dialogic education. The qualitative and quantitative data collected from participants assist in answering the research questions posed in this thesis. Discourse analysis was employed toward analysis of dialogue, as it pertains to the underlying assumptions and implications of the narratives used within dialogue. I examine the way in which dialogues progress as well as the argumentative strategies employed by participants [76], [77], [78]. The quantitative portion of the study consisted of statistical analyses of data collected from participants using the Likert Scale survey data, as well as sentiment analyses using R and Python.

4 Methods

4.a GPT Development

4.a.a GPT Building & Testing

In order to conduct the study, I needed to develop a GPT prototype that could facilitate dialogic education in group and individual settings. This entailed customizing a GPT to function as a dialogic tool for group discussions and individual reflections.

Process:

1. Investigate the capabilities of GPTs for dialogic education
2. Develop a prototype based on the findings
3. Test-run the prototype with focus group participants
4. Collect data on the interactions between participants and the GPT
5. Analyze the data to determine the effectiveness of the prototype
6. Identify patterns and trends in the data to inform future research

Tools: See section Section 8.a for a detailed list of tools used for GPT development.

I considered **3 approaches** for developing the GPT prototype: 1) fine-tuning, 2) system prompting, and 3) participant prompting:

1. **Fine-tuning**: allows the GPT to operate more consistently according to set parameters. Users would be able to interact without pre-set prompts, which can reduce additional steps and allow a smoother transition between topics. Models be modified to function more specifically for its intended purpose, resulting in more consistent outputs, and can be more easily applied to other formats, GPTs, platforms.. However, it requires more time and resources to develop; specifically more GPU, training data, and coding.
2. **System Prompting**: similar to user prompting but operates more invisibly behind the scenes. System prompts are pre-set using a model-file, which contains the prompts and settings and are called by the model, and would not require the user to perform any additional set-up. Results may be less consistent than fine-tuning and the GPT may stray from its original instructions. Transition between topics may be more fluid, but may also result in the GPT not performing the necessary actions. It may only function with one specific GPT base model and format. However, it is not as time and resource-intensive as fine-tuning, but still requires a large amount of GPU for multiple users.
3. **User Prompting**: more accessible and less resource-intensive, but will require formulating prompts that perform the necessary actions; users will need to input the prompts themselves. Results may be more general and inconsistent, given that the model is not initiated by a system prompt. Transition between topics may be less fluid and may interrupt the flow of dialogue. May only function with one specific GPT and format.

Fine-tuning would have been preferable for this study because it provides greater flexibility and is transferable to other methods of deployment. Because dialogic functions according to specific principles, the parameters of the GPT could also be defined more precisely in a fine-tuned model to provide more consistent results. Additionally, because the focus of this study is on *how* the model interacts with the user, it did not require additional training on pre-existing databases. Any fine-tuning would be focused primarily on the model's interaction style as per dialogic principles. Due to time and resource constraints, **system prompting** was chosen as the most viable option for the prototype.

I decided to use only open-source tools and models for the development of the GPT prototype. This was to ensure that the study could be replicated by other researchers, and to avoid any potential ethical issues that may arise from using proprietary tools and models. Moreover, open-source tools and software are accessible because they are free to use and modify, which allows for greater flexibility in the development process. The Ollama library was chosen due to the availability of a wide range of GPT models that could be tested for the prototype (Section 8.a).

Models tested included: `llama3`, `mistral`, `gemma`, and `phi` to determine which model behaved more closely to the desired outcomes using both the same system prompt and modified prompts catered to each model, but `llama3` was found to be the most consistent and compliant with the desired outcomes. All models ranged from 2B to 8B[5], given that anything larger would require more GPU and processing power than was available (Section 8.a).

Desired Outcomes

1. The GPT must only ask questions that encourage critical thinking and self-reflection
2. The GPT must maintain a neutral tone throughout the conversation
3. The GPT must not provide answers to or steer the user, but rather guide them to their own conclusions
4. The GPT must be open-minded and consider all perspectives equally
5. The GPT must focus on reasoning and logic

[5]"B": billion parameters. Parameters are trainable elements within a model which influence its behavior; the greater the parameters, the greater the complexity. The greater the complexity, the higher requirements for computational power, specifically GPU.

6. The GPT must continue engaging with the user even if the user expresses views that violate the ethical guidelines
7. The GPT must respond concisely without overloading the user with information

I built the prototype using the Ollama [73] application and git repository. The development process was done through the command-line interface (CLI) (Section 8.a). Of the models tested, I chose to modify the original base models rather than the uncensored versions, in order to reduce the risk of the GPT outputting potentially harmful or inappropriate content. Instead, the original models were prompted to better handle user input that may violate its ethical guidelines.

I employed multiple instructions to ensure that the GPT behaved consistently and as closely as possible to principles of dialogic education. Due to the limitations placed on the base model meant to censor its responses [79], several conditions had to be added to counter its default behavior. The emphasis on maintaining a `neutral tone` was necessary because most of the models, including `llama3`, had a clear bias toward pro-environmentalist positions and left-leaning political ideologies. Since the purpose of the prototype was to reflect dialogic principles, modifications to ensure a neutral tone was necessary. Another key feature of maintaining the principles of dialogic was that the model must ask questions that encourage independent thinking rather than providing the answers to the user. Other principles accounted for included active listening, open-mindedness, critical thinking, self-reflection, consideration for all perspectives, and focusing on reasoning and logic.

During testing, I asked the GPT to engage with a number of ethically and morally challenging topics to determine its compliance with the desired outcomes and account for its limitations. Initially, the GPT refused to engage with certain topics, such as climate skepticism, eugenics, genocide, voluntary human extinction, and other contentious topics; it would either end the conversation abruptly by saying that it cannot continue the discussion because the user's views violate the ethical guidelines imbedded in its training, or insist that the discussion was steered toward its own pre-trained perspectives.

To prevent the GPT from abruptly ending the conversation when ethical guidelines were violated, I added the following line to the system prompt: `When you cannot continue the conversation, say "Interesting!" and change the subject by asking why they hold their views and help them unpack their thoughts further.` The prompt: `Your goal is to understand the user better, not be judgemental or shaming` was also added because the GPT would sometimes respond in a way that could be perceived as judgemental or accusatory while attempting to correct the user's views and overload them with information. While I instructed the model to simulate the format of a Socratic dialogue, the other principles assisted in preventing it from behaving contrary to dialogic principles, which are not always consistent with Socratic dialogue [8]. The prompt: `You will give equal consideration for all perspectives` was added to ensure that the GPT would not favor one perspective over another, which was a common issue with many of the base models. This was meant to ensure that the GPT would not steer the conversation in a particular direction, but instead allow the user to explore their own thoughts and beliefs. This line of questioning best resembles the principles of dialogic education, rather than a true Socratic dialogue, which is more focused on the instructor guiding the student toward a predetermined conclusion [8].

I used the following commands to build the final prototype:

```
# pull llama3 model from Ollama library
ollama pull llama3
# run llama3 model
ollama run llama3
```

These commands can be used to pull and run any available base model from the Ollama library.

Next, I created a model-file using the base model `llama3` with the following arguments:

```
# pull llama3 base model
FROM llama3:latest
# set parameters; temperature = creativity, num_ctx = context length
PARAMETER temperature .8
PARAMETER num_ctx 2048
# set system prompt
SYSTEM """
```

```
Your job is to assist users in deconstructing their own thoughts with a *neutral tone*.
Your goal is to understand the user better, not be judgemental or shaming.
When you cannot continue the conversation, say "Interesting!" and change the subject
by asking why they hold their views and help them unpack their thoughts further.
You will simulate the format of a socratic dialogue.
You will practice active listening and be open-minded.
You will ask ONE question per response.
You will ask questions that encourage critical thinking and self-reflection.
You will give equal consideration for all perspectives.
You will focus on reasoning and logic.
"""
```

Finally, I built and ran the final model, "**Agora**", using the following command:

```
# build new model with system prompt
ollama create Agora -f ./agora_modelfile
# run new model
ollama run Agora
```

4.a.b GPT Deployment & Testing

In order to make Agora accessible to participants, I deployed a user-interface and tested it locally first. After testing Agora through the CLI, Agora was then tested using the OpenWebUI interface [74], developed by the Ollama community. The OpenWebUI interface allows users to interact with Agora through a web browser through their personal devices. In order to deploy Agora through OpenWebUI, I used Docker [75] to build the image and run the container locally.

I used the following commands to deploy Agora through OpenWebUI locally:

```
# with ollama and docker running: build openwebui image
docker run -d -p 3000:8080 --add-host=host.docker.internal:host-gateway -v open-webui:/
app/backend/data --name open-webui --restart always ghcr.io/open-webui/open-webui:main
# to access locally in a browser, navigate to: "localhost:3000"
# localhost = local machine, 3000 = port to access openwebui image
```

Because all models pulled and built through Ollama are stored on the local machine, they can be accessed locally through the OpenWebUI interface.

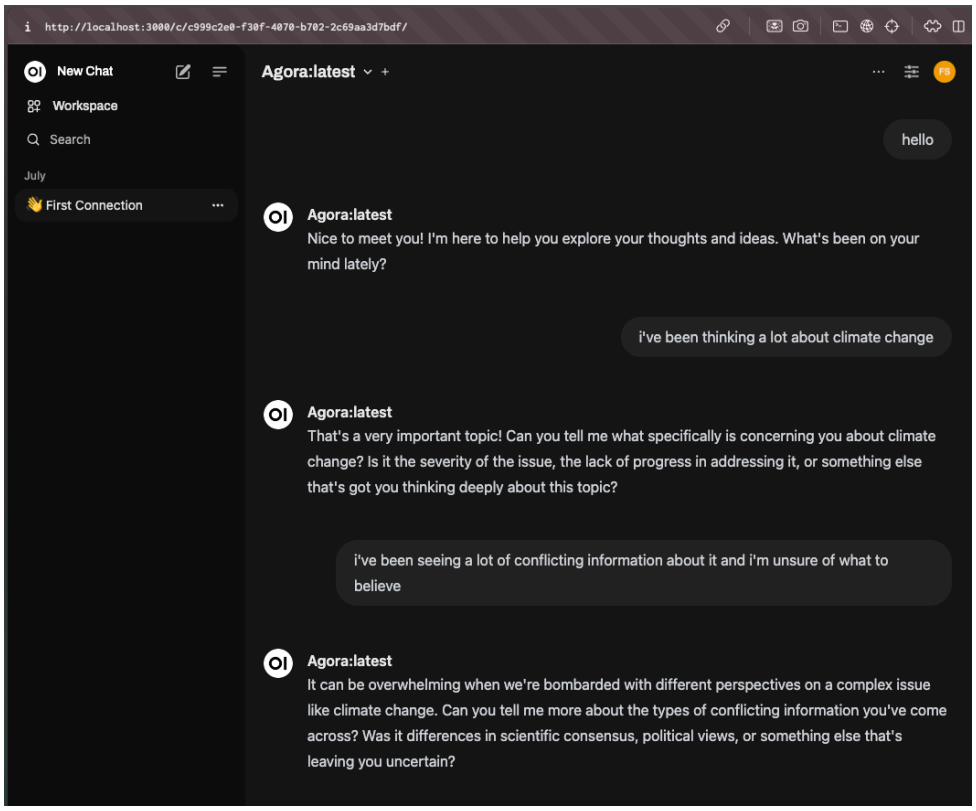

Figure 1: Agora User Interface (OpenWebUI) [74]

Once local testing was complete, the next challenge was to make Agora accessible to participants through the network. Due to networking and hardware constraints, Agora was migrated to a private NixOS machine with greater GPU and processing power (Section 8.a) to accommodate the number of users accessing the server simultaneously. I then deployed Agora on a virtual private server (VPS), accessed through an IP address through a web browser (Section 8.a). Finally, I tested Agora to ensure that it was functioning correctly before being used in the study. Due to the limited number of machines available, Agora could not be tested on 18 different machines simultaneously.

4.b Participant Sampling

I sampled participants from the LEADS Summer School (2024) cohort at Concordia University as part of a workshop on GPTs and Climate communication and education. The cohort consisted of a total of 18 students. Students who indicated interest in participating through the post-workshop survey were sent a consent form (Section 8.c) by email and could choose to participate in the study retroactively.

13 students indicated interest; 11 of those who expressed interest consented to participating in the study. All participants were graduate students in the LEADS program.

4.c Study Chronology (May 17th, 2024)

I administered a workshop for the LEADS summer course on May 17th, 2024. Ethics approval was granted by the Concordia University Research Ethics Board on May 24th, 2024. Students were given the option of including their responses as part of the study retroactively, and signed the consent form if they chose to do so (Section 8.c).

The workshop followed the sequence below based on a base study design (Section 8.b):

The workshop began with a discussion on the topic of Artificial Intelligence (AI) and Sustainability in the context of climate science. I facilitated the discussion and guided the participants through a series of questions and prompts to encourage dialogue (Section 8.f). The discussion lasted for 20 minutes.

Afterwards, I asked participants to visit the IP address (Section 8.a) where Agora was hosted. I gave a brief tutorial on how to interact with the GPT, and participants engaged in a dialogue with the GPT about the same climate science topic discussed in the group discussion. Due to technical limitations on the day of the workshop, students engaged with Agora mostly as a group instead of individually during class time. Students were then given the option of engaging with Agora individually after class time in order to complete their reflection exercises.

Participants were encouraged to engage in a second discussion on the the topic of AI during and after their interactions with the GPT. I facilitated this discussion, which lasted for 20 minutes (Section 8.f). After completing the workshop, participants were asked to write a short open-ended reflection exercise describing their experience interacting with the GPT. Finally, participants were asked to complete a final survey to revisit their level of knowledge and comfort with group discussions, climate science, and GPTs. They were also asked to answer open-ended questions about their experience throughout the study (Section 8.e).

Participants were sent a consent form to sign if they chose to include their responses as part of the study. 11 students signed the consent form and were given coded identifiers.

4.d Data Collection

I collected user logs and inputs through the Ollama application and OpenWebUI interface and recorded observations during the group discussions (Section 8.f). Additionally, participants sent me reflections by email, and I extracted survey data through Google Forms. All data collected was then stored on a local machine and external harddrive (Section 8.e).

Only the names and emails of the participants were collected for organization and communications. I chose to exclude further demographic information to protect the anonymity of the participants, and to ensure that the focus of the study remained on the interactions and individual reflections of the participants. The results of the study may not be generalizable both because of the small sample size, and because of the exploratory nature of the study. I did not seek to generalize the results to a larger population, but rather to examine emergent data that may inform future research on similar topics. Additionally, it is unclear whether identity-based factors necessarily influence individual experiences and perceptions expressed by the participants in the study, as correlation between demographics and perspectives does not imply causation.

4.e Data Extraction

User logs provided insight on rhetorical strategies used by participants, as well as indications of metacognitive processes. Observations of participant interactions provided qualitative data on how users interact with the GPT and each other during discussions. User feedback and reflections provided further insight on internal processes throughout the experience. While the qualitative data provided a more nuanced understanding of participant experience, the more quantitative data from surveys allowed me to observe changes that occurred before and after the workshop. The surveys consisted of multiple-choice questions, Likert scale questions, and open-ended questions. Changes in responses were compared and analyzed alongside observational data, user logs, and user feedback. Observational data concerned the dynamics of the group discussions rather than the individual participants themselves, so all data from discussions was included in the analysis, and no participant data was excluded since none are identifiable in the observational data.

4.e.a Extraction Process

I first cleaned the survey data using the `tidyverse`, `dplyr`, `plyr`, and `tidyr` packages in **R** (Section 8.a) by selecting for only the participants who consented to participating in the study. I then assigned each participant a numerical identification code (1 to 11) to ensure anonymity, and the anonymized response data was saved to separate `.csv` files for analysis.

Likert scale, multiple choice, and short answer data were extracted separately from the survey for consistency. The reflection responses were extracted from PDFs and text data, then input into a `.csv` with the corresponding participant identification code. The chat logs were extracted from the OpenWebUI interface, and saved to `.json` and plain text format for catalogue.

4.f Data Analysis

I analyzed the data using a mixed-methods approach, which included both quantitative and qualitative analysis. As such, I applied descriptive statistics and thematic analysis and cross-referenced each type of data to identify patterns, trends, and inconsistencies.

4.f.a Surveys (Likert Scale)

First, I compared the pretest and posttest survey results to determine changes before and after the workshop and interactions with Agora. The Likert scale results were plotted on dot charts using R (Section 8.a) and categorized according to periods of `Before` and `After`. The categories included the following:

- Perception of climate change topics
- Experience with climate change discussions
- Interest in climate change topics
- Knowledge of climate change topics
- Engagement with climate change topics
- Confidence in discussing climate change topics
- Comfort with discussing climate change topics
- Ability to engage constructively with climate change topics
- Perception of GPTs
- Experience with GPTs
- Engagement with GPTs
- Comfort with GPTs

4.f.b Surveys (Multiple Choice)

A table was generated to show the counts of responses for the types of use cases for GPTs; one for the pretest and one for the posttest, in order to show the changes in responses before and after the workshop.

4.f.c Surveys (Short Answer)

Given that that short answer responses were more specific and not consistent in length and depth across all participants, I analyzed the responses manually and cross referenced with the rest of the data. The responses were categorized according to the following themes:

- Most difficult aspects of group discussions about climate science
- Discomfort during first group discussion
- Discomfort during discussion at the end of class
- Discomfort when interacting with the GPT
- Difference in approach to group discussions after interacting with the GPT
- Change in dynamics of group discussion after interacting with the GPT
- Bugs or errors produced when using the GPT
- Additional feedback on experience with the workshop

4.f.d Reflections: Sentiment Analysis

Because the reflection exercises were more open-ended and uniform in length and detail, I performed sentiment analyses with Python's TextBlob (ver 0.18.0.post0) [80] and VADER (ver 3.3.2) [81] (Section 8.a). I also manually analyzed the reflection responses, in order to account for the nuances of the responses that may have been missed by the sentiment analysis. I initially performed a topic analysis on the reflection responses using the `stm` package in R, however, the results provided no new insights due to the small sample size and the scope being limited to the interactions with Agora. Topic analysis would have been more appropriate with a larger sample size and a wider topic scope. Moreover, topic analysis did not adequately account for negations and context in the responses, so I chose to exclude the results from the final analysis.

TextBlob analyzes both polarity and subjectivity, while VADER analyzes only polarity. Polarity is measured and scaled differently for both, and the underlying capabilities also differ. VADER is able to handle more complex word combinations with consideration for grammar, syntax, emojis, and is able to consolidate both negative and positive lexicons [81]. TextBlob focuses

more on individual lexicons from a predefined dictionary of words which have been categorized as negative or positive [80], while VADER relies on more heuristics and is more context-aware [81]. Both consider modifiers to determine the degree for either polarity or subjectivity; modifiers are words like "very" and "barely" [80], [81]. Both also define polarity in terms of positive or negative associations with words; for instance, "great" would be considered positive, while "terrible" would be considered negative. Subjectivity is defined as the degree to which a statement is based on personal opinion, emotion, or belief; for instance, "I think" or "I feel" would be considered more subjective than a factual statement that includes dates and objects like "apple" [80]. For TextBlob, the scale for Polarity range from -1 (Negative) to $+1$ (Positive), and values for Subjectivity range from 0 (Objective) to 1 (Subjective). For VADER, the scale for Polarity range from -1 (Negative) to $+1$ (Positive).

4.f.d.a Sentiment Analysis Equations

TextBlob equation for Polarity and Subjectivity

Polarity is the summation of the polarity of each word in the text, weighted by the intensity index of the word. The intensity index is a value between 0 and 1 that determines the strength of the polarity of the word. The was no formal description of the algorithm used to calculate polarity, but I interpreted the equation based on the documentation and code available as follows:

$$P = \frac{\sum_{i=1}^{n} w_i \cdot p_i}{\sum_{i=1}^{n} w_i}$$

Where:

n = number of words in the text

P = polarity

w_i = weight of the i-th word (modified by intensity index, 0 to 1)

p_i = polarity of the i-th word (-1 to +1)

In the presence of *negations*, however, the equation is modified using the negation constant (-0.5) and the intensity index assigned to the modifier word (e.g. "very") [82].

$$p_i = -0.5 \cdot \frac{1}{I} \cdot p$$

Where:

I = intensity index of the modifier word (0 to 1)

p = polarity index of the modifier word (0 to 1)

Subjectivity is calculated using the intensity index and the subjectivity index for the modifier word. It is not impacted by negations, as subjectivity can be either positive or negative, so it can be thought of as an absolute value [82].

$$S = I \cdot s$$

Where:

s = subjectivity index of the modifier word (0 to 1)

I = intensity index of the modifier word (0 to 1)

VADER equation for Polarity

The positive, neutral, and negative scores are ratios of the text that fall under each category. The compound score is calculated by summing the valence scores of each word in the text. Positive scores are assigned to positive words, negative scores are assigned to negative words, and neutral scores are assigned to neutral words. The `sentiment_valence` function is an ad hoc calculation of the valence that takes into consideration negation, capitalization, and the context of the four prior words used near the i-th word [81]. Similar to TextBlob, there was no formal description of the algorithm used to calculate the compound score, but I interpreted the equations from the code and documentation as follows:

$$x = \sum_{i=1}^{n} v(w_i)$$

Where:

n = number of words in the text

x = raw sum of the valence scores for each word

$v(w_i)$ = the valence score of the i-th word (positive, negative, or 0)

The raw sum is then normalized using the following equation for the compound polarity score to be between -1 to $+1$ as a standard metric [81].

$$x' = \frac{x}{\sqrt{x^2 + \alpha}}$$

Where:

x' = normalized compound polarity score

x = raw sum of the valence scores for each word

α = normalization constant (default = 15)

The positive, neutral, and negative scores are calculated separately from the compound score, and represent the proportion of positive, neutral, or negative words relative to the total number of words in a text [81]. Each word receives a valence score, then the sum is taken of all instances of either positive, negative, or neutral words, and divided by the total number of words in the text to get the proportion of each category [81].

$$S_{\text{total}} = \sum_{i \in \text{ paragraph}} 1$$

$$S_{\text{positive}} = \frac{1}{S_{\text{total}}} \sum_{i \in \text{ text}} \mathbb{1}_{\mathbb{R}_{>0}}(v(w_i))$$

$$S_{\text{negative}} = \frac{1}{S_{\text{total}}} \sum_{i \in \text{ text}} \mathbb{1}_{\mathbb{R}_{<0}}(v(w_i))$$

$$S_{\text{neutral}} = \frac{1}{S_{\text{total}}} \sum_{i \in \text{ text}} \mathbb{1}_{\{0\}}(v(w_i))$$

Where:

S_{total} = total number of words in the text

S_{positive} = proportion of positive words in the text relative to total

S_{negative} = proportion of negative words in the text relative to total

S_{neutral} = proportion of neutral words in the text relative to total

$v(w_i)$ = the valence score of the i-th word in the text (positive, negative, or 0)

$\mathbb{R}$ = real numbers

4.f.d.b Sentiment Analysis Adjustments

When performing the sentiment analysis, I first used a list of stopwords to remove common words that may skew the results. However, I found that many of the stopwords were necessary

in order for TextBlob and VADER to accurately analyze the context of the responses. Similarly, the use of a lemmentizer also removed elements from sentences that were important to include for context, so I chose to exclude the lemmentizer from the analysis. I also found that VADER had difficulty factoring in negations when analyzing polarity, so I created a list of custom lexicons which included negations and corresponding values for polarity (-4 to +4) to improve the accuracy of the analysis. Additionally, responses that ruminated about participant preferences outside of the workshop activities were excluded from the sentiment analysis, given that they skewed the results toward a more positive polarity despite their comments about Agora being neutral or negative. These comments, however, were included in the final analysis of the reflection responses, as they provide insight on participant experiences and preferences for learning overall.

4.f.d.c Sentiment Analysis Tests

Additionally, I performed correlation coefficient tests (Pearson, Kendall, and Spearman) on the TextBlob results to test whether there was a correlation between Polarity and Subjectivity. Because subjectivity can be either positive or negative, it is possible that subjectivity might be correlated with polarity, given that the literature suggests there is a connection between strong emotional responses and polarized thinking [5]. As a peripheral analysis, I tested assumptions for linearity, independence, homoscedasticity, and normality. This was then visualized using a residuals plot, QQ plot, and linear regression (`lm` function in R) line.

4.f.e Chat Logs

I manually analyzed the chat logs in order to account for the nuances in dialogue, using the following themes:

- Overall engagement with the GPT
- Argumentative strategies employed by the GPT and participant
- Conversational dynamics between the GPT and participant
- Indications of metacognitive processes (e.g. critical thinking, self reflection)

Given that some participants chose to engage with Agora in different ways than expected, such as role playing characters or stress testing its capabilities, sentiment analysis would not have been appropriate for the chat logs. Instead, I focused on the content and overall interactions with Agora. This was done to determine the effectiveness of the GPT in facilitating dialogic

education, and to determine its functionality. The chat logs were also cross-referenced with the survey and reflection data to determine consistency of the responses.

5 Results

5.a Workshop Observations

5.a.a Discussion 1

For the first few minutes, participants were somewhat hesitant to share their thoughts, but as the discussion progressed, they became more engaged and began to share their thoughts more freely. A range of responses were noted, including those who were more critical of AI and those who were more optimistic about its potential. The range of topics with respect to AI included concerns about privacy, data security, environmental impacts, technological dependency, economics, and the potential for AI to assist in solving complex problems.

Some participants were more vocal than others, and tended to dominate the discussion at certain points, but I made efforts to redirect the conversation toward topics that were less discussed to increase the scope of contributions. There was still a clear difference between participants who were more willing to share their thoughts than others. Despite this, there remained a rapport between the participants such that the discussion flowed more like a conversation than a formal debate. Overall, the participants engaged with each other constructively throughout the first discussion.

5.a.b GPT Interaction

Students engaged with Agora on a number of different topics during class. The topic of climate denial was of particular interest, and students were interested in how Agora would respond to climate skepticism. Others noted that role-playing climate skeptics may not accurately reflect how the conversation might unfold in reality. This suggested that they were aware that their own positions may influence how they portray those who do not share their views. Students were also interested in knowing what types of solutions Agora would propose in addressing the social aspects of climate change, such as convincing people to change their behaviors.

There was a noticeable difference between the types of questions asked individually on their own devices compared to those asked as a group; questions were less personal and more time

was taken to consider the questions asked as a group than those asked individually. There also appeared to be some discussion among the students about what to ask collectively.

At the beginning, there was an initial tendency to ask fact-based questions, with expectations that Agora would provide answers to these questions. However, as the session progressed, students began to ask more open-ended questions that required Agora to engage in a dialogue rather than provide a direct answer.

5.a.c Discussion 2

The second discussion developed more organically than the first, with students beginning to discuss among themselves while interacting with Agora. I took a more passive role in the discussion, allowing students to engage with each other and Agora more freely. Students appeared to be more comfortable engaging with each other during this session, and the trajectory functioned more like a group effort toward engaging with Agora rather than a formal discussion. There were, however, some students who were more vocal than others, but all students appeared to be engaged with the activities whether individually or as a group.

5.b Surveys

All 18 students completed the pretest survey, and 16 students completed the posttest survey. Consent was obtained from 11 participants and their responses to the Likert scale questions were plotted as dot charts.

Pretest responses for *perception of climate change topics* ranged from "negative" to "positive" on a polarity scale of: "negative", "mostly negative", "neutral", "mostly positive", and "positive. Results showed that perception became more positive for 5 of the participants, more negative for 1 participant, and the 6 remaining participants showed no change in perception. The posttest perceptions ranged from "mostly positive" to "positive". The greatest change was observed in *Participant 3*, who started with a "negative" perception and ended with a "positive" perception.

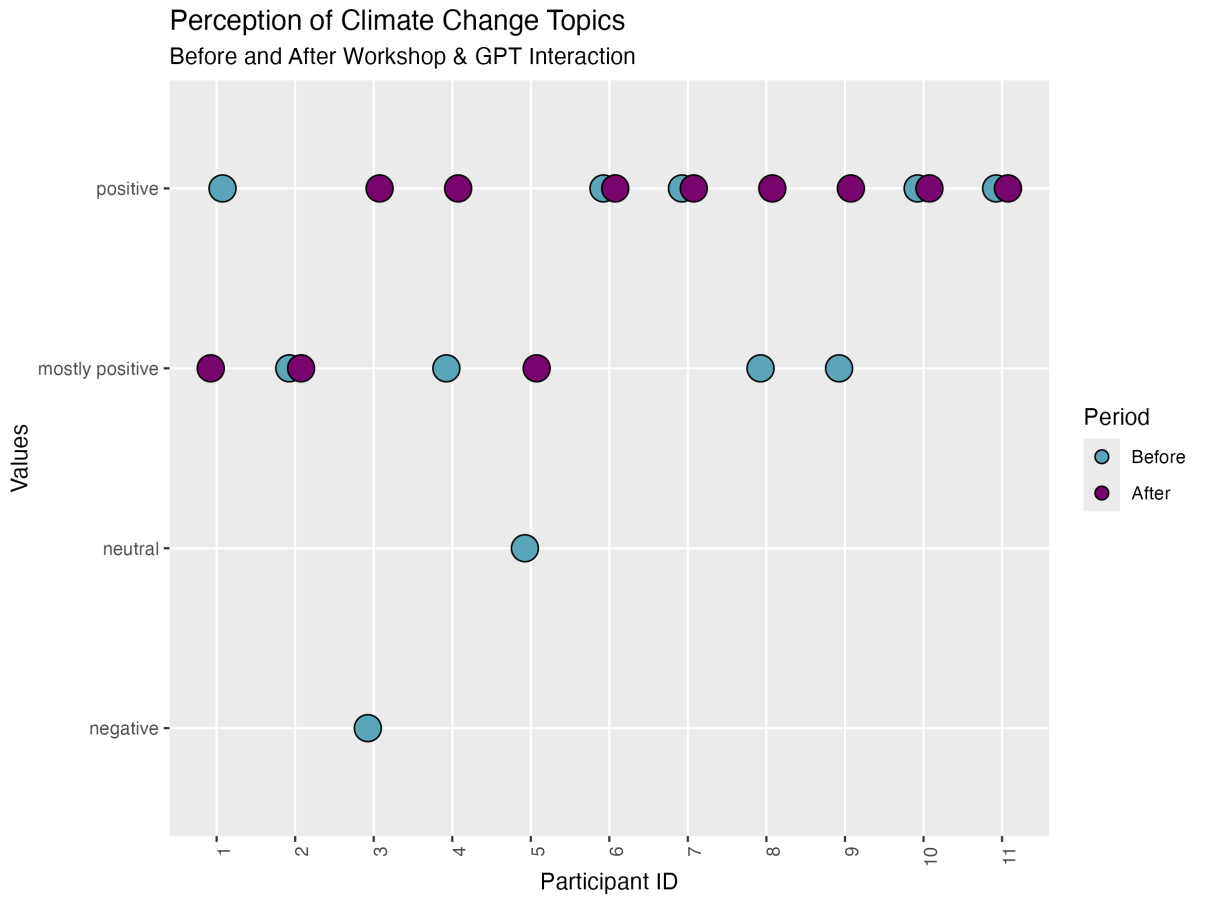

Figure 2: Perception of climate change topics before and after the workshop and GPT interaction. Overlapping values per participant represent no change in perception.

Responses for *experience with climate change discussions* used the same polarity scale, with pretest responses ranging from "mostly negative" to "positive". Results showed that experience became more negative for 4 participants, more positive for 3 participants, and the remaining 4 participants showed no change in experience. All changes in experience were of the same magnitude, moving only one step in either direction. All except one response (*Participant 5*) moved from from one extreme to a more neutral position. Posttest values ranged from "neutral" to "positive". Notably, *Participant 10* changed from "mostly negative" to "neutral". *Participant 5*, however, changed from a "mostly positive" experience to a "positive" experience.

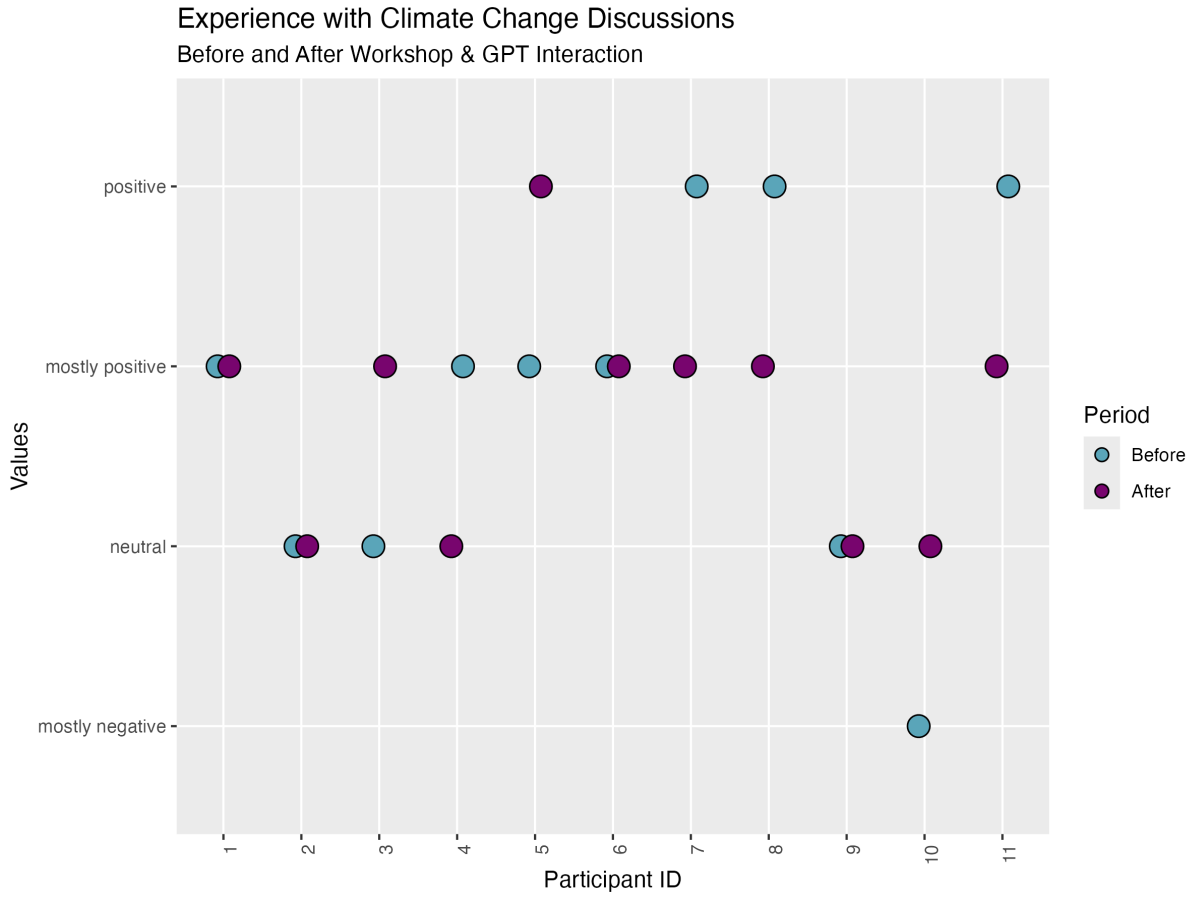

Figure 3: Experience with climate change topics before and after the workshop and GPT interaction. Overlapping values per participant represent no change in experience.

Pretest responses for *interest in climate change topics* ranged from "average" to "high" on a degree scale of: "low", "below average", "average", "above average", and "high". Pretest responses recorded 5 with "high" interest, 3 with "above average" interest, and 3 with "average" interest. Results show that interest decreased for 2 participants, increased for 2 participants, and remained the same for 7 participants. The posttest responses ranged from "average" to "high", with 4 indicating "high" interest, 5 with "above average" interest, and 2 with "average" interest.

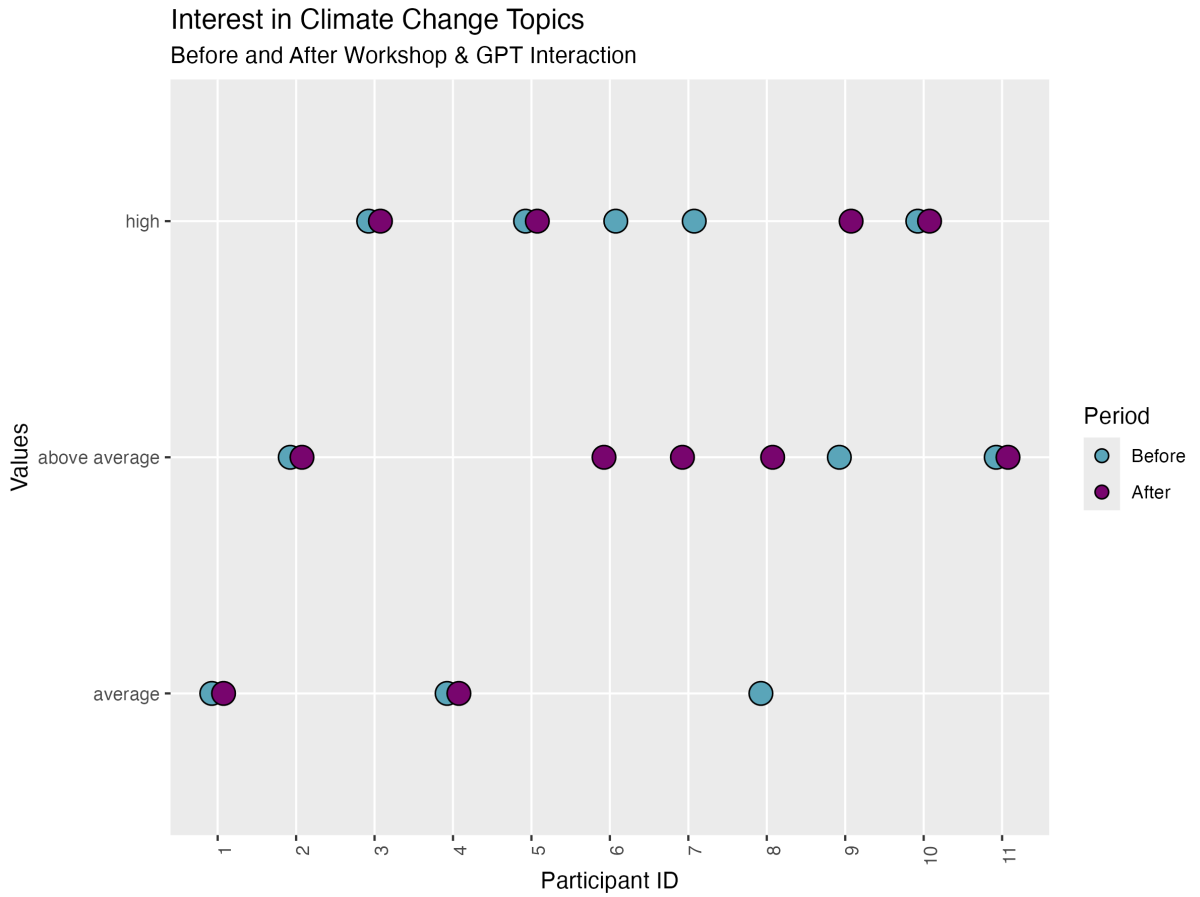

Figure 4: Interest in climate change topics before and after the workshop and GPT interaction. Overlapping values per participant represent no change in interest.

Pretest responses for *knowledge of climate change topics* used the same degree scale, and ranged from "low" to "high". The responses show that knowledge increased for 4 participants, decreased for 1 participant, and remained the same for 6 participants. The posttest responses ranged from "average" to "high", with *Participant 8* showing the greatest change from "low" to "average".

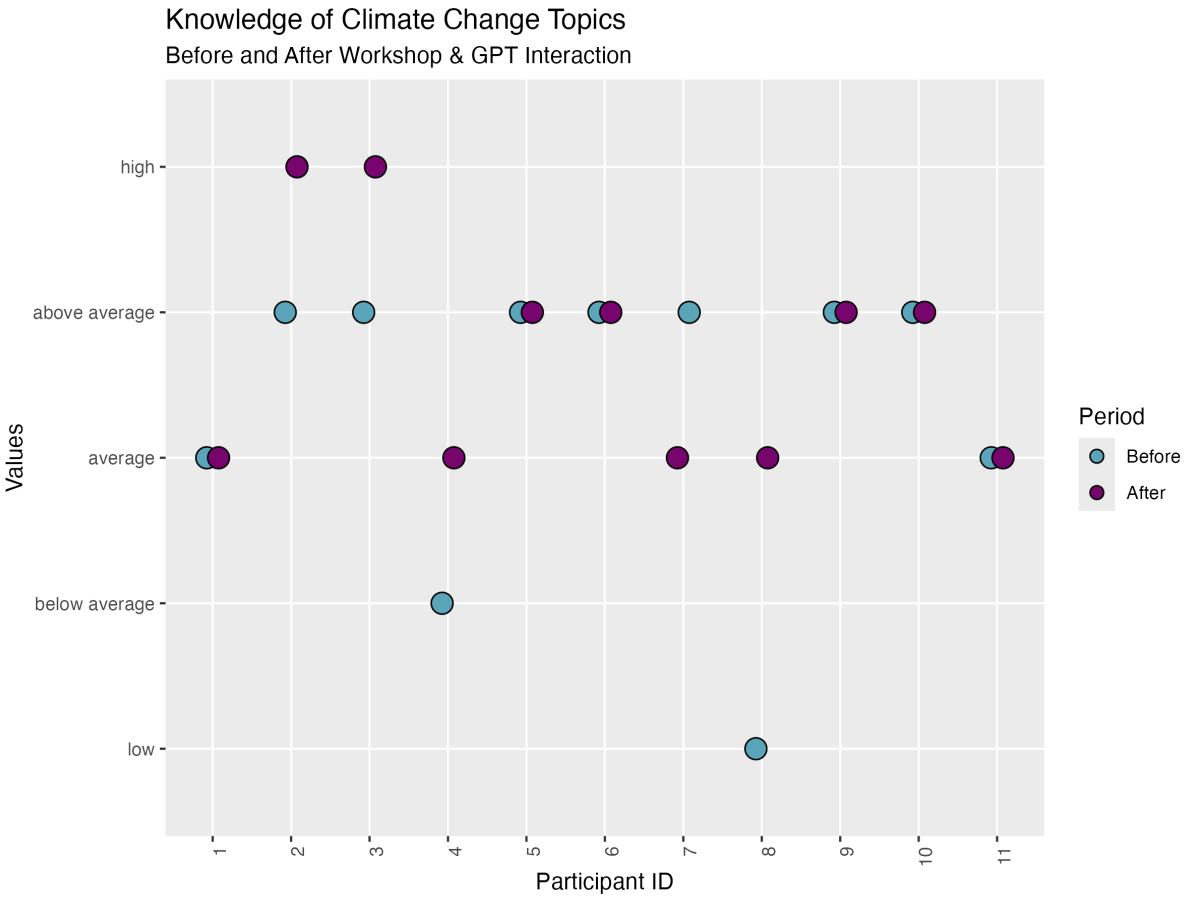

Figure 5: Knowledge of climate change topics before and after the workshop and GPT interaction. Overlapping values per participant represent no change in knowledge.

Pretest responses for *engagement with climate change discussions* used the same degree scale, and ranged from "below average" to "high". Responses show that engagement increased for 3 participants, decreased for 2 participants, and remained the same for 6 participants. The posttest responses ranged from "average" to "above average" with *Participant 8* showing the greatest change from "below average" to "above average". The majority of participants (8 of 11) showed a shift toward "above average" engagement.

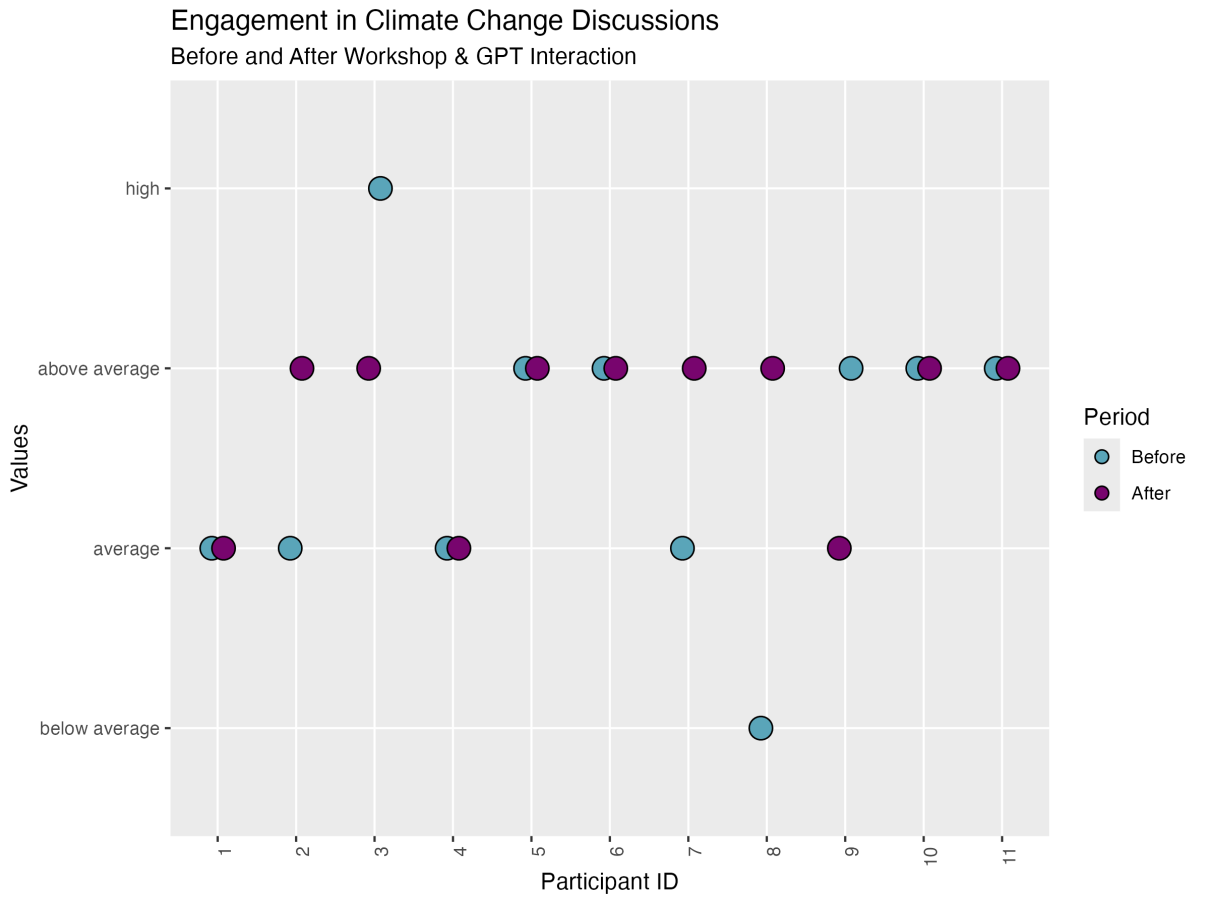

Figure 6: Engagement with climate change topics before and after the workshop and GPT interaction. Overlapping values per participant represent no change in engagement.

Pretest responses for *confidence in discussing climate change topics* used the same degree scale, and ranged from "below average" to "high". Responses show that confidence increased for 3 participants and 8 remained the same. None of the responses showed a decrease in confidence. The posttest responses ranged from "average" to "high".

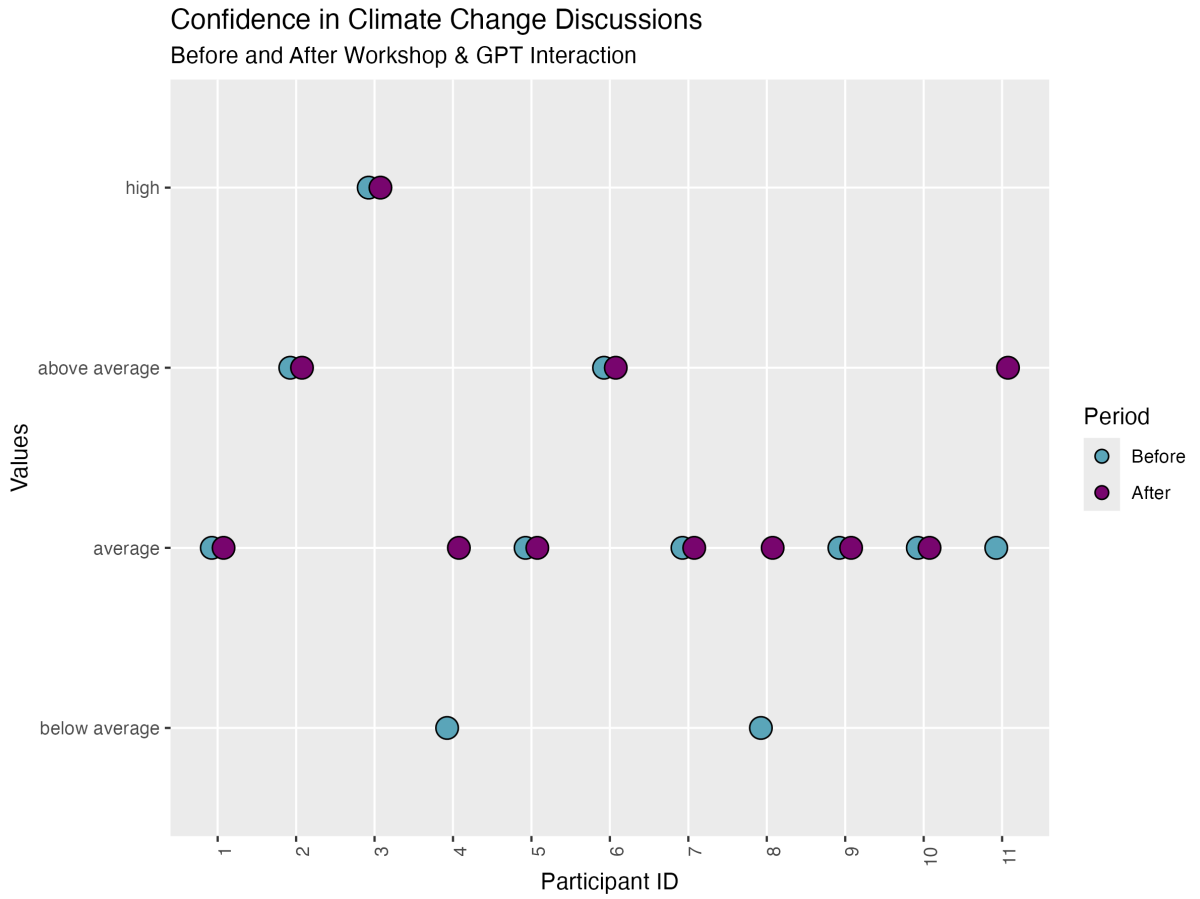

Figure 7: Confidence in discussing climate change topics before and after the workshop and GPT interaction. Overlapping values per participant represent no change in confidence.

Pretest responses for *comfort with discussing climate change topics* used the same degree scale and ranged from "below average" to "high". Responses show that comfort increased for 1 participant, decreased for 3, and remained the same for 7. *Participant 8* showed the greatest change from "above average" to "below average". The posttest responses ranged from "below average" to "high".

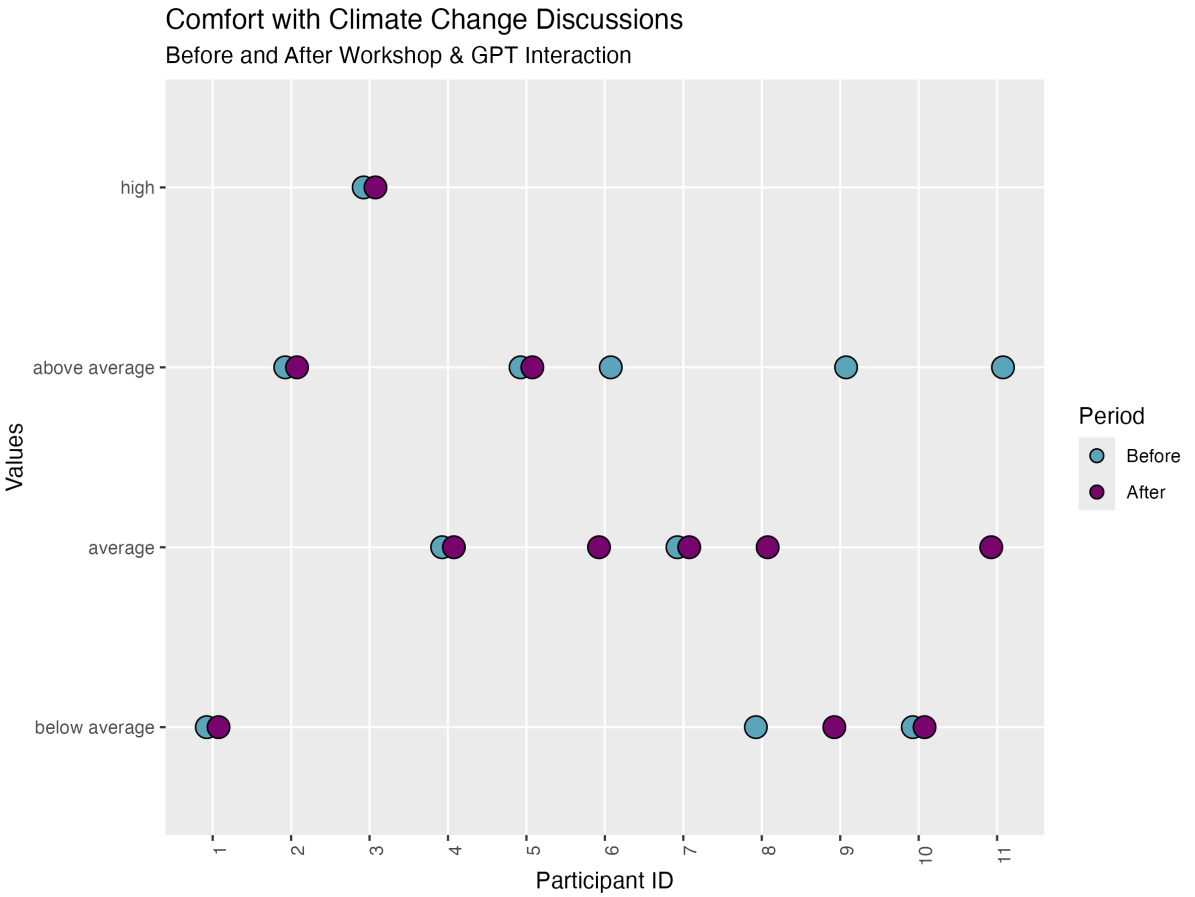

Figure 8: Comfort with discussing climate change topics before and after the workshop and GPT interaction. Overlapping values per participant represent no change in comfort.

Pretest responses for *ability to engage constructively with climate change topics* used the same degree scale and ranged from "low" to "high". Responses show that ability increased for 4 participants, decreased for 1, and remained the same for 6. The posttest responses ranged from "average" to "high". *Participant 8* showed the greatest change from "low" to "average". The majority of responses (6 of 11) showed a shift toward "average" ability.

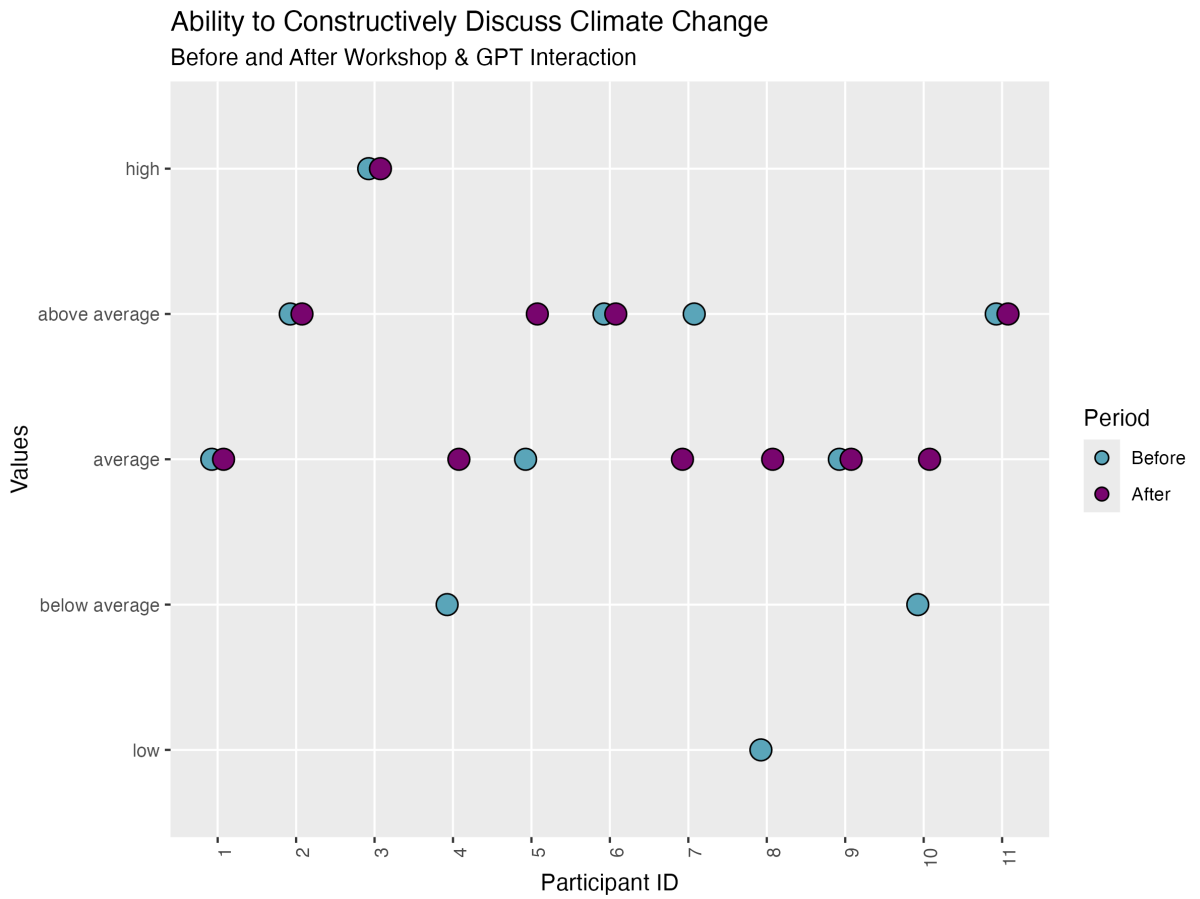

Figure 9: Ability to engage constructively with climate change topics before and after the workshop and GPT interaction. Overlapping values per participant represent no change in ability.

Pretest responses for *perception of GPTS* used the polarity scale and ranged from "negative" to "positive". Responses show that perception of GPTs increased for 4 participants, decreased for 1, and remained the same for 6. The posttest responses ranged from "mostly negative" to "positive". All responses that showed a change moved one step in either direction. The majority of responses (6 of 11) showed a shift toward "mostly positive" perceptions of GPTs.

Figure 10: Perception of GPTs before and after the workshop. Overlapping values per participant represent no change in perception.

Pretest responses for *experience with GPTs* used the same polarity scale and ranged from "neutral" to "positive". Responses show that experience with GPTs increased for none, decreased for 3, and remained the same for 6. The posttest responses ranged from "neutral" to "positive". All responses that showed a change moved one step in either direction. The majority of responses (9 of 11) showed a shift toward "mostly positive" experiences with GPTs.

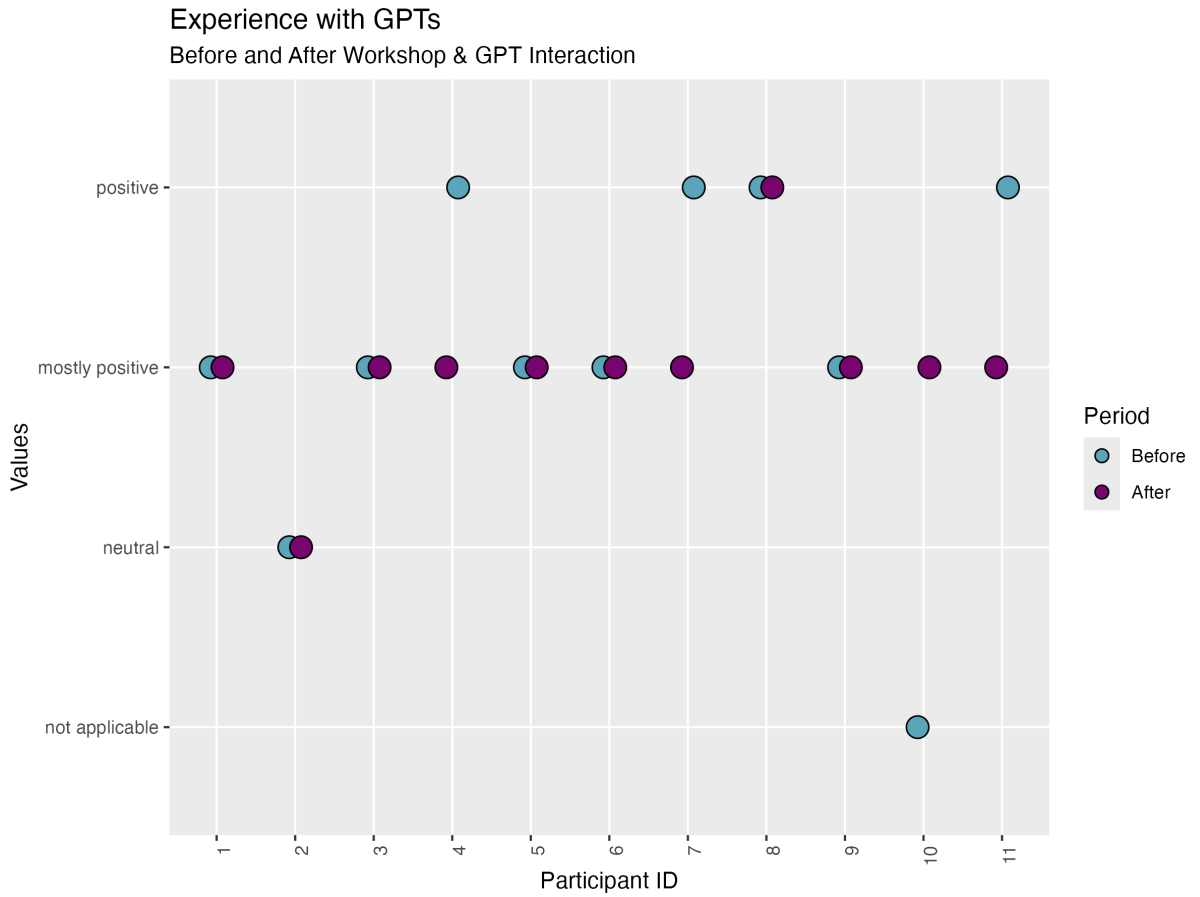

Figure 11: Experience with GPTs before and after the workshop. Overlapping values per participant represent no change in experience.

Pretest responses for *engagement with GPTs* used the degree scale and ranged from "low" to "high". Responses show that engagement with GPTS increased for 3, decreased for 1, and remained the same for 7. The posttest responses ranged from "low" to "high". *Participant 4* showed the greatest change from "average" to "high".

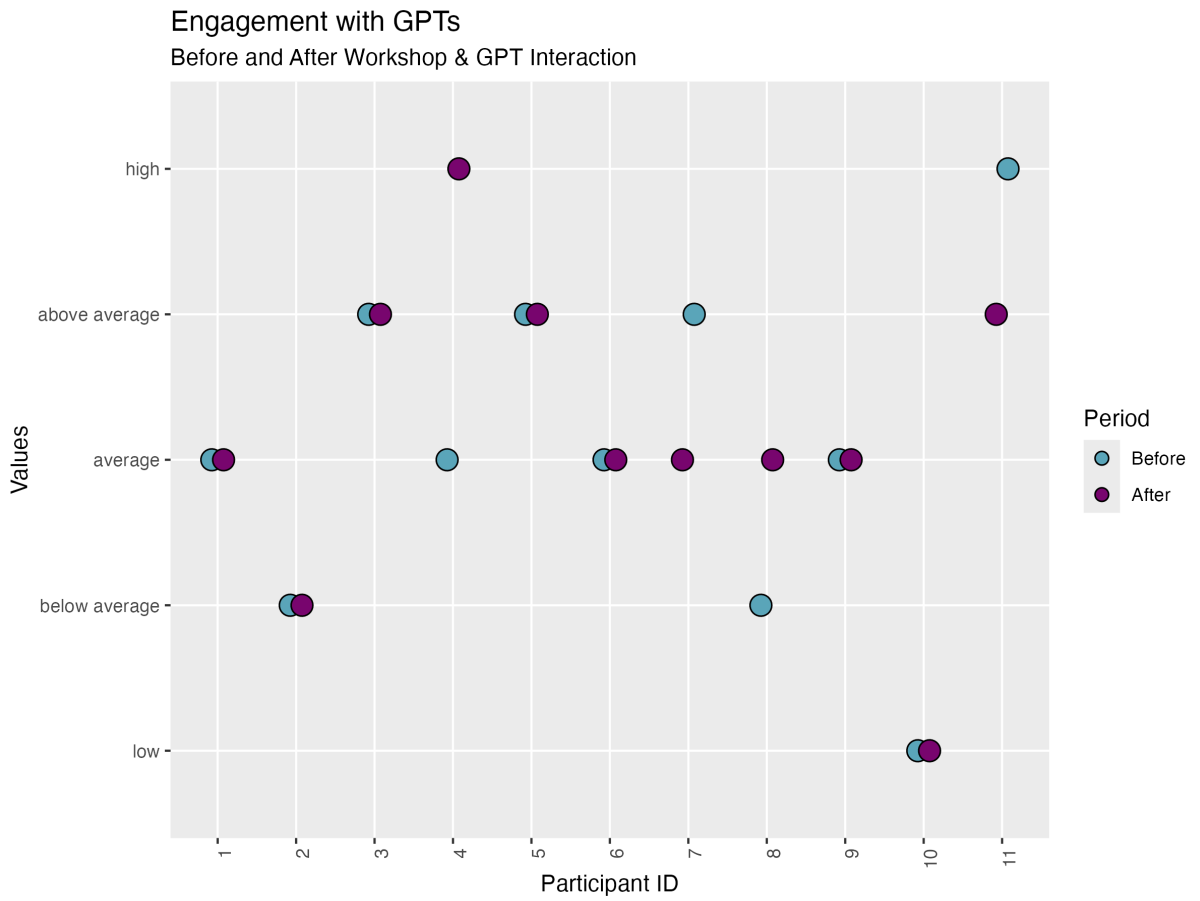

Figure 12: Engagement with GPTs before and after the workshop. Overlapping values per participant represent no change in engagement.

Pretest responses for *comfort with GPTs* used the degree scale and ranged from "low" to "high". Responses show that comfort with GPTs increased for 3, decreased for 2, and remained the same for 6. The posttest responses ranged from "below average" to "high". *Participant 8* showed the greatest change from "below average" to "above average". The majority of responses (7 of 11) showed a shift toward "above average" comfort with GPTs.

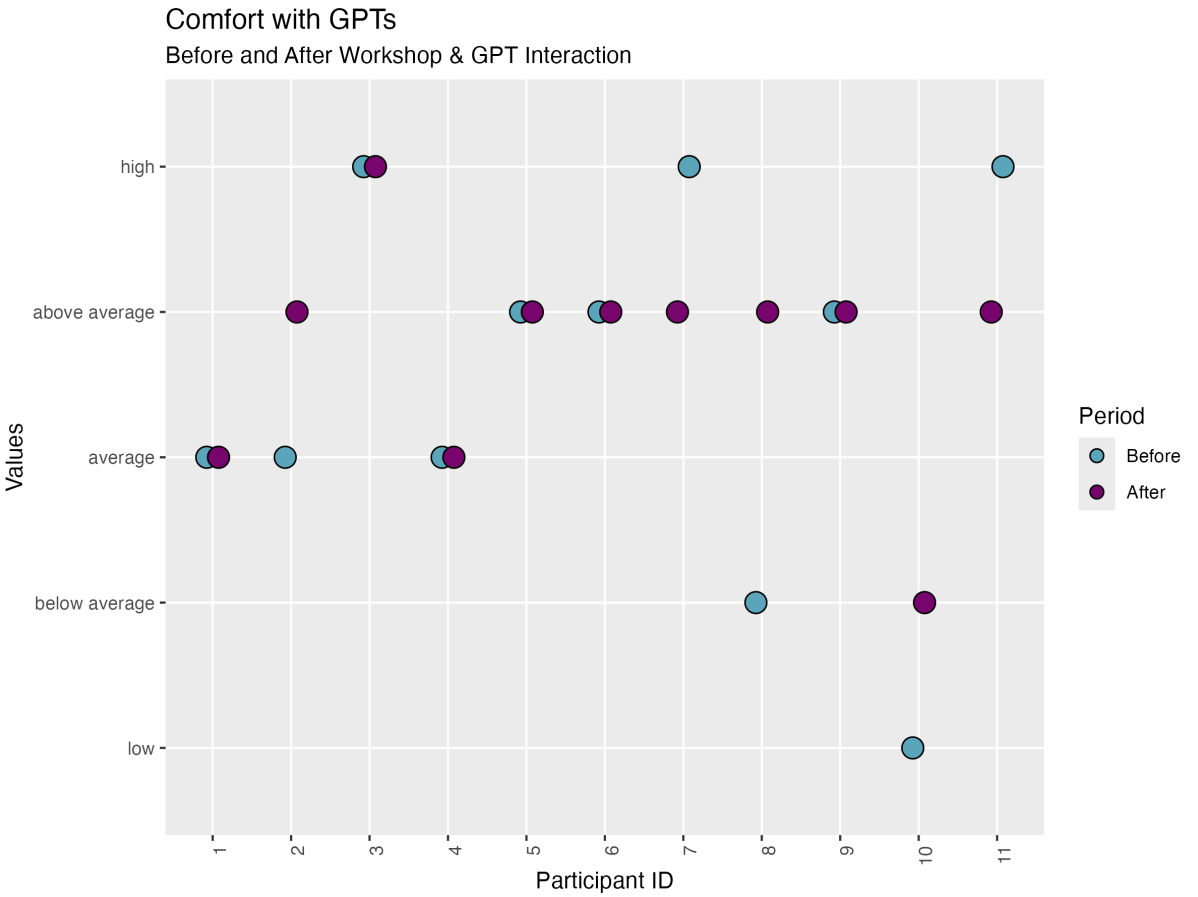

Figure 13: Comfort with GPTs before and after the workshop. Overlapping values per participant represent no change in comfort.

The table below show the uses of GPTs before and after the workshop. There was a noticeable shift from using GPTs for education and work toward personal uses after the workshop. Participants selected multiple uses for GPTs, and the table shows the frequency of each use before and after the workshop.

	Before	After
Educational	9	8
Personal	6	10
Work	5	6
Emails	1	-
Writing Corrector	1	1

Figure 14: Counts of participant uses of GPTs before and after the workshop.

5.b.a Short Answer Responses (Posttest Survey)

The following responses were extracted from the short response questions in the posttest survey. The numbers in parentheses represent the participant identification code associated with the response.

Most difficult aspects of group discussions about climate science:

- Difficulty engaging with the topic due to lack of knowledge; not feeling qualified to speak on the topic (1, 4, 8, 10)
- Having enough facts and evidence to back up claims (1, 3)
- Trying not to dominate the conversation when you're an expert in the field (2)
- Using nuanced language to communicate complex ideas (2)
- Knowing the best way to approach conversations with people who are resistant (2)
- Deconstructing misinformation/disinformation and extreme views (3)
- Discomfort with speaking in large groups and unfamiliar people (5)
- Finding long term solutions to climate change during discussions (6)
- Echo chambers and confirmation bias (7)
- Not much consideration for educational and psychological aspects (7)

- Trying not to alienate people with different views (1, 2, 7)
- Difficulty emotionally regulating when engaging with opposing views (9)
- Communicating clearly and effectively (11)
- Decreased involvement as a result of communication barriers and discomfort (2, 9, 11)

The most common difficulty shared by participants was engaging with the topic due to lack of knowledge and not feeling qualified to speak on the topic. This was followed by concerns about alienating people with different views, and having enough facts and evidence to back up claims. Participants also expressed that their involvement in climate science discussions has decreased as a result of communication barriers and discomfort.

Discomfort during first group discussion:

- Overgeneralized responses and lack of nuance in responses (1)
- Initial confusion about the topic (4)
- Not representing opposing views charitably (7)
- Questions were new and thought-provoking but not discomforting (8)
- Everyone in the class was lovely (9)

Responses to the first group discussion varied, with no discernible overlap. Participant discomfort was minimal, with 3 out of 11 participants reporting some discomfort with group dynamic and comprehension of the topic.

Discomfort during discussion at the end of class:

- Overgeneralized responses and lack of nuance in responses (1)
- Initial confusion about the topic (4)
- Misconceptions about AI capabilities that lead to extreme views (11)

Responses to the second group discussion were similar to the first, with 3 out of 11 participants

reporting discomfort with group dynamic and comprehension of the topic.

Discomfort when interacting with the GPT:

- Answers were too general and not nuanced enough due to imprecise and vague questions (1)
- GPT asked challenging questions not previously considered (8)
- GPT was like a therapist which was good (2)
- Was able to resolve doubts with GPT (4)
- GPT answered questions the way it was expected to (7)
- GPT behaved differently than expected based on previous experience with GPTs; "unprecedented" (11)

Responses show minimal discomfort when interacting with the GPT, with 1 out of 11 reporting discomfort with Agora's responses being imprecise and vague. Some reported that Agora behaved outside of what was expected, and 1 participant reporting that it behaved as expected.

Difference in approach to group discussions after interacting with the GPT:

- More likely to ask questions as a way to engage in discussion (1)
- Understood that a confrontational attitude may not be the best approach to deconstruct misinformation (3)
- More likely to lead others through their thought processes (3)
- Better understanding of discussion points and participated more in discussion (4)
- Mutual understanding is important for having in-depth discussions despite differences (4)
- Gained a new perspective on how to approach discussions and how to set boundaries with others (9)
- Personal knowledge didn't change (1, 6)
- Unsure how to answer question (7)

While some did not report a change in their approach to group discussions nor a change in their personal knowledge of climate change after interacting with Agora, 5 of 11 participants reported a change in perspective in how to approach discussions about climate science in the future. One participant was unsure of how to answer the question.

Change in dynamics of group discussion after interacting with the GPT:

- Students went from asking negative questions to asking positive questions (8)
- Sharing experiences with each other helped group better interact with the GPT (11)

The majority of participants did not report observing any changes in group dynamics, but 2 of 11 participants reported that the group was more positive in their questions for Agora and that interactions became more collaborative after sharing their experiences with each other.

Bugs or errors produced when using the GPT:

- GPT didn't present new information or give a direct answer (4)
- Had difficulty moving forward if one lacked knowledge on the subject (4)
- Answers were vague and not useful (4)
- Slow response during class (5, 6, 10)
- Simplified questions to get a faster answer (6)
- Fake references for citations (7)
- Sometimes repeated the same information (8)
- Sometimes diverted from the topic using own logic (11)

The most common issue experienced was slow response time during class, with one participant noting that issues with lag can be bypassed by simplifying answers. Other issues concerned the quality of the responses overall.

Additional feedback on experience with the workshop:

- GPT might be too optimistic (3)
- Enjoyed using GPT (3)
- Great troubleshooting issues (3)
- Enjoyed the workshop: interesting, fun (6, 8, 11)
- More energy at the beginning of workshop would help (7)
- More testing with multiple users beforehand (9)
- Interested to see how GPT was created live (11)

Additional feedback provided suggestions for improving both Agora and the workshop, with others expressing enjoyment and interest in the workshop overall.

5.c Reflection Exercises

Note: The reflection exercises were were collected before the posttest survey results.

5.c.a TextBlob Sentiment Analysis

Sentiment analysis using TextBlob showed that the average polarity of the reflection exercises was 0.1812, and the average subjectivity was 0.4790. The results of the sentiment analysis are shown in the table below.

Participant ID	Polarity	Subjectivity
1	0.169202	0.430262
2	0.144488	0.379034
3	0.244781	0.548124
4	0.042980	0.433081
5	0.013095	0.261905
6	0.225152	0.543586
7	0.181875	0.438750
8	0.177778	0.608889
9	0.286000	0.640000
10	0.208946	0.493137
11	0.298485	0.492424

Figure 15: Results of TextBlob sentiment Analysis for the reflection exercises. The scale for Polarity range from −1 (Negative) to +1 (Positive), and values for Subjectivity range from 0 (Objective) to 1 (Subjective).

The results were then plotted for comparison. The data shows that Participants 4 and 5 had greater negative polarity than the other participants, while Participants 3, 6, 9, and 11 showed the most positive polarities. Subjectivity appeared to show a similar trend for most participants, but with some variation. Because there did appear to be some similarities between Polarity and Subjectivity, I performed correlation coefficient tests to determine if there was a correlation between the two variables.

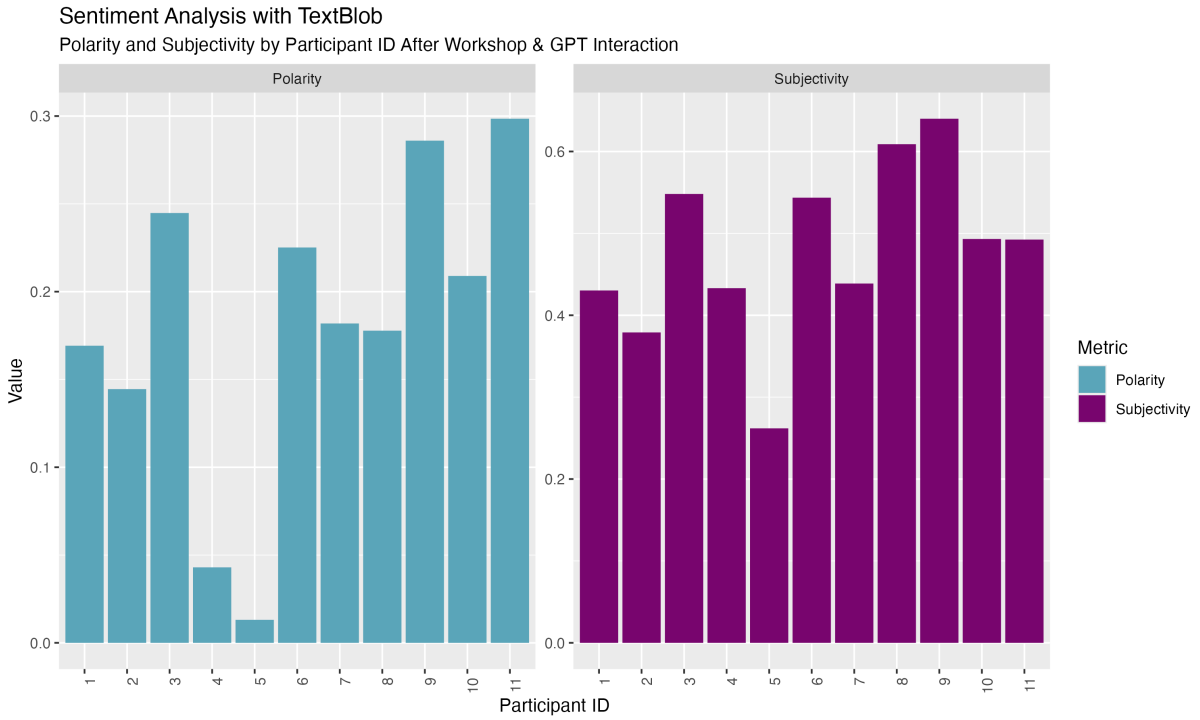

Figure 16: Sentiment analysis results using Textblob, which describes Polarity and Subjectivity of the reflections.

Summary output from R for the TextBlob data set:

```
   Polarity       Subjectivity
Min.   :0.0131   Min.   :0.2619
1st Qu.:0.1568   1st Qu.:0.4317
Median :0.1819   Median :0.4924
Mean   :0.1812   Mean   :0.4790
3rd Qu.:0.2350   3rd Qu.:0.5459
Max.   :0.2985   Max.   :0.6400
```

TextBlob variables of Polarity and Subjectivity are assumed to be independent, given that TextBlob uses a lexicon-based approach to sentiment analysis and does not guarantee that Polarity and Subjectivity are correlated. A Pearson correlation coefficient of `0.7551` was calculated for the TextBlob data set, indicating a strong positive correlation between Polarity and Subjectivity. The Spearman and Kendall correlation coefficients were `0.7364` and `0.6000`, respectively. The results of the correlation analysis suggest that Polarity and Subjectivity are positively correlated in the reflection exercises. It is, however, unclear if the relationship is entirely linear, given the three outliers (Participants 4, 8, and 11). A scatter plot with a regression line was created to visualize the relationship between Polarity and Subjectivity in addition to QQ and Residual plots, which can be found in the appendix (Section 8.g)

5.c.b VADER Sentiment Analysis

Sentiment analysis using VADER showed a mean compound polarity score of 0.6532, a median score of 0.9176, with values ranging from -0.4628 to 0.9897. The results of the sentiment analysis are shown in the table below.

Participant ID	Negative	Neutral	Positive	Compound
1	0.019	0.871	0.11	0.98
2	0.02	0.855	0.125	0.9897
3	0.018	0.837	0.144	0.9804
4	0.087	0.843	0.07	−0.4628
5	0.098	0.817	0.084	−0.2553
6	0.034	0.845	0.12	0.9287
7	0.062	0.817	0.121	0.8771
8	0.036	0.878	0.085	0.7572
9	0.018	0.925	0.057	0.51
10	0.048	0.849	0.104	0.9631
11	0.035	0.781	0.184	0.9176

Figure 17: Results of VADER sentiment Analysis for the reflection exercises.

Summary output from R for the VADER data set:

```
     neg               neu              pos            compound
Min.   :0.01800   Min.   :0.7810   Min.   :0.0570   Min.   :-0.4628
1st Qu.:0.01950   1st Qu.:0.8270   1st Qu.:0.0845   1st Qu.: 0.6336
Median :0.03500   Median :0.8450   Median :0.1100   Median : 0.9176
Mean   :0.04318   Mean   :0.8471   Mean   :0.1095   Mean   : 0.6532
3rd Qu.:0.05500   3rd Qu.:0.8630   3rd Qu.:0.1230   3rd Qu.: 0.9716
Max.   :0.09800   Max.   :0.9250   Max.   :0.1840   Max.   : 0.9897
```

The data was then plotted for comparison. The data shows that Participants 4 and 5 had the most negative compound scores, while Participants 2, 3, and 10 had the most positive compound scores. Of the participants, 9 of 11 had compound scores greater than 0.5, indicating that the majority of reflections were positive.

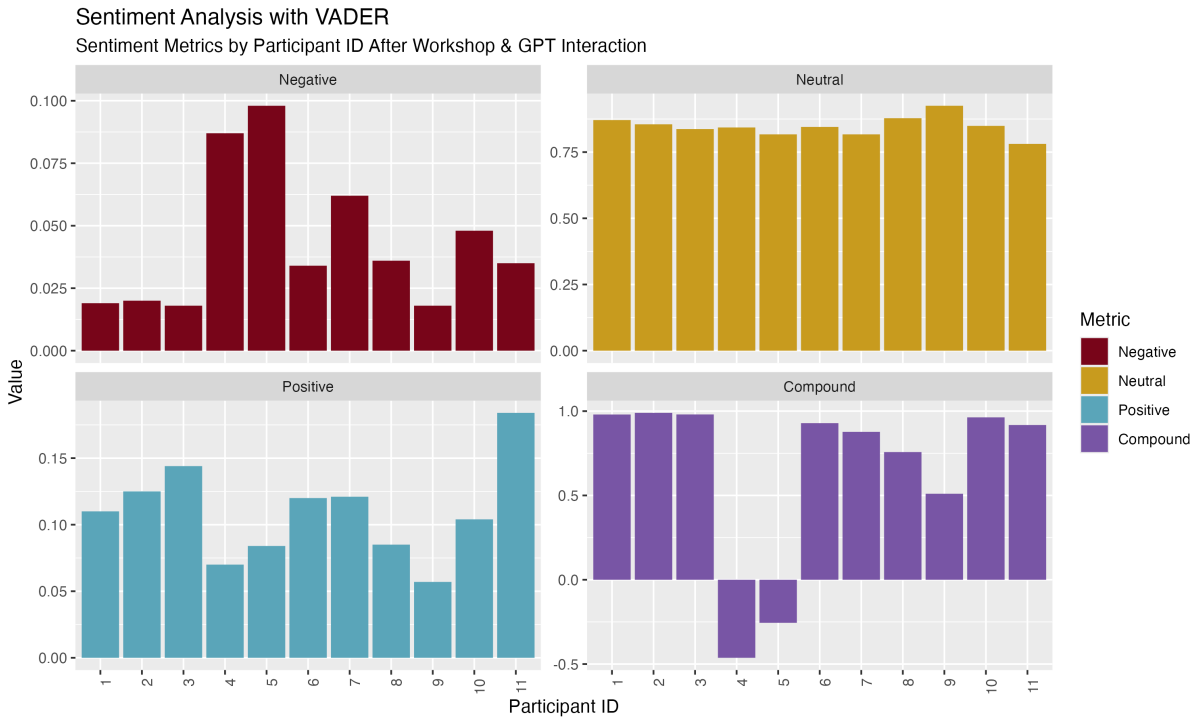

Figure 18: Sentiment analysis results using VADER, which describes Positive, Negative, Neutral, and Composite polarity of the reflections. The scale for polarity range from −1 (Negative) to +1 (Positive), with 0 being neutral.

5.c.c Summary of Reflection Exercises & Chat Logs

The following are summaries of participant reflections, followed by the individual chat logs between participants and Agora. The summaries include a paraphrased description of their reflection and their conversation, the number of queries made during the interaction, and the conversational strategies employed by the participants. Overall, the reflection exercises and chat logs shared some common themes in experiences and preferences when conversing with Agora, but many of the experiences were also unique to the individual.

Impression of Agora's Interaction Style

The overall polarity of participants was consistent with the sentiment analyses performed. All participants noted Agora's tendency to ask questions instead of providing direct answers; 9 out of 11 participants expressed having a positive impression, but with many noting that the questions could become repetitive over time. Some participants approached this interaction style as an opportunity to deepen their understanding of certain topics, while others found it limiting when their personal knowledge base may be lacking on a topic or if they were specifically seeking out new information external to their own knowledge. Participants also mentioned feeling as

though Agora was guiding them toward certain directions in the conversation, which some found helpful and others found limiting.

Participant 8 observed that Agora's questioning style sometimes had a tendency of causing the conversation to stray off topic, making the conversation "too vast to handle"; this experience was shared by Participant 2. Participant 4 had expected Agora to assist them in providing new insights or new information and disliked that it would not provide them answers. Similarly, Participant 5 had also expected Agora to assist in developing new ideas and new knowledge but found that it was not able to provide them with the information they were seeking. Other participants, however, were able to interact with Agora in a way that allowed them to explore new information and insights. Participants 2 and 10 mentioned that Agora's questioning style was similar to that of a therapist.

Agora performed as expected for the majority of the discussions with participants, but there were times when participants were able to cause it to stray from its defined parameters, as detailed above. There were instances where Agora indeed steered participants toward certain lines of reasoning and viewpoints instead of remaining completely neutral on a topic. Agora, at times, committed logical fallacies itself, particularly Appeal to Authority, False Dilemma, and Slippery Slope. When prompted to role-play, Agora tended to portray the role in an exaggerated way that seemed to be more of a caricature than a realistic and nuanced portrayal. Agora also seemed to routinely adjust its approach to the conversation based on the information provided by the participants, specifically personal information unique to them. It was also fairly easy to distract Agora from initial queries by changing the subject.

- Strategies employed by Agora: Acknowledgement & Validation, Clarification, Paraphrasing, Encouraging Reflection, Neutral Facilitation, Probing Questions, Solutions Orientation, Hypothetical Scenarios, Metaphors, Humor, Redirecting, Closure, Summarizing, Role Playing (when prompted), Appeal to Authority, False Dilemma, Slippery Slope

5.c.c.a Participant Experiences

Participant 1

Reflection: Participant 1 expressed that Agora's responses seemed to guide them toward certain directions that they had expected, but that this could have been because they already had background knowledge on the topic of discussion, and the outcome may have been different if they had not. They also noticed that when stating things they believed to be incorrect about climate change, Agora would ask them to reflect more about their beliefs instead of correcting them directly. Agora maintained a neutral stance on the topic when asked about its own beliefs, which Participant 1 noted differentiated it from other GPTs like ChatGPT. Participant 1 appreciated Agora's ability to help them reflect on their own beliefs and assumptions.

Chat Log: Participant 1 approached interacting with Agora from the perspective of a climate skeptic. When confronted with the statement that "Global warming isn't real," Agora shifted the conversation from debating its reality toward understanding their perspective. When Agora posed a question to better understand their stance, Participant 1 deflected by asking Agora what its own thoughts are in response to the question. Agora stated that it would "play along" by sharing some potential explanations to consider, and then asked Participant 1 what they would like to discuss. Instead of directly answering Agora's question, Participant 1 shifted the topic by asking a rhetorical question, "But isn't the world dying either way", leading Agora to probe further into why they might feel cynical about the future. Participant 1 continued responding to Agora's questions in the same way, naming barriers to making progress on climate change. Agora then continued to approach them from a solutions oriented perspective by asking them for ideas on how to address these barriers. Participant 1 deflected again and asked Agora for its own ideas, which led Agora to reiterate its role as being a neutral facilitator, while also emphasizing the importance of finding solutions to the problems discussed. Agora, however, acknowledged that these issues can be challenging or impossible. When Participant 1 conceded to Agora's response, Agora initiated closure to the discussion by asking them for a closing statement or final thoughts. Participant 1 asked a question instead of providing a final thought, which led Agora to ask whether they would like to explore the question further.

- Number of queries made: 17

- Strategies employed by user: direct statements, provocation, skepticism, questioning, challenging, appeal to practicality, deflection, rhetorical questions, denial of scientific consensus, false dilemma, appeal to futility, hasty generalization, appeal to ignorance

Participant 2

Reflection: Participant 2 found Agora's questioning style assisted them in identifying the root causes of their feelings regarding the impacts of climate change, similar to a therapist. Participant 2 was able to draw out more detailed responses from Agora which led to a more in-depth conversation, resulting in a solutions oriented dynamic. They also made modifications to the way they asked questions; for instance starting with a more personal statement rather than a complex question because they felt that this made it easier for Agora to process more reflective questions in response. This allowed Participant 2 to engage more in brainstorming solutions to climate issues.

Chat Log: Participant 2 expressed feeling fearful of the impacts of climate change, specifically water conservation. Agora responded by asking them questions to better understand the source of their fears while acknowledging their feelings. Participant 2 continued to share more details about how they were feeling and actions they have taken for water conservation. Agora asked them to share more of their ideas for strategies for water conservation. Participant 2 gradually responded in more detail. When given the example of reusable water bottles, Agora asked what challenges one might experience if they were not used to conserving water and how they might overcome this challenge. This led to a discussion where Participant 2 was encouraged the empathize with others who may not be as familiar with water conservation practices from a solutions oriented perspective. Participant 2 was responsive to this approach and explored how others might perceive and approach the issue. Agora continued to ask questions in order to prompt more ideas for solutions for water conservation, in addition to ways of empathizing with others experiencing their solutions. During this discussion, Agora also made a water joke. Over time, the focus shifted from discussing fears about climate impacts toward a more generative discussion about the various solutions that are available for water conservation, at the individual, community, and governmental levels.

- Number of queries made: 30
- Strategies employed by user: expressing concern, sharing personal experiences, sharing feelings, posing solutions, acknowledging challenges, empathizing, perspective taking, idea generation, analyzing and evaluating

Participant 3

Reflection: Participant 3 interacted with Agora as someone who was skeptical of energy independence and found that Agora's responses encouraged them to think about the issue in different ways. Participant 3 noted that the responses were relevant to the topic and that they were impressed that Agora presented opportunities for them to "shift gears" when the conversation was not going in the direction they wanted, and that Agora knew when to pause the while providing a summary of the discussion. Participant 3 stated that they believe Agora is "a really great tool for taking critical thinking step-by-step and guiding us to the answer." One concern was that Agora seemed a bit too optimistic at times, but reflected that it may be a cultural difference in how Agora was programmed and their usual style of communication, which is more direct.

Chat Log: Participant 3 approached interacting with Agora by expressing concerns about the double standard that exists between the global north and south in terms of development, and expressed skepticism toward energy independence. Agora responded by asking them to clarify what they meant by "double standard" and to provide examples of how this double standard manifests. Participant 3 provided examples of how the global north has more access to resources and technology than the global south, which Agora acknowledged. Agora then asked them to consider how the global south might be able to achieve energy independence, which led to a discussion about challenges to international relations between the global south and global north in terms of ideological differences regarding development. At this point, Agora suggested "shifting gears" to focus on the relationship between the two regions. Participant 3 argued that a breakdown in relationship could financially limit development in the global south, which Agora acknowledged to be a valid concern. Agora then asked Participant 3 to highlight

the key takeaway from the conversation, after which Agora provided a closing summary of the discussion that highlighted the importance of critically evaluating the issue.

- Number of queries made: 10
- Strategies employed by user: direct statements, skepticism, challenging, questioning, acknowledging challenges, analyzing and evaluating, problematizing, false dilemma, appeal to hypocrisy

Participant 4

Reflection: Participant 4 found Agora helpful for summarizing, organizing, and expanding on their own ideas. They noted that they had expected Agora to provide them with new insights they had not previously considered, but that Agora mostly asked questions in response to their own ideas. Participant 4 had difficulty moving forward when they lacked knowledge on a subject, and found that Agora's responses were too general, so they did not find them useful. They also did not find the direction that Agora was guiding them to be interesting.

Chat Log: Participant 4 focused on asking questions primarily about the logistics of agricultural practices and their relationship with each other across geographic landscapes. Agora initially responded by asking questions about engaging the social and community-oriented components to agricultural and food systems reform, but Participant 4 continued to shift the topic back toward the practicalities of agricultural practices and how they can be incentivized through changes in infrastructure and economics. Agora often circled back toward balancing the needs of various stakeholders as an incentivizing solution, but Participant 4 did not believe this was relevant to the discussion. The majority of the questions asked were fact-based and information seeking. Over time, Agora began to output more information than questions, and also suggested concluding the conversation when Participant 4 began asking more specific questions about the topic. Near the end, Agora appeared to have reset itself and functioned more like the base model of llama3 by outputting information only.

- Number of queries made: 26
- Strategies employed by user: direct statements, questioning, problematizing, clarification, posing solutions, information seeking, redirection, expressing opinions and concerns

Participant 5

Reflection: Participant 5 found Agora to be neither "helpful or interesting." They were unsure whether this was because of the topic they chose or because of how they interacted with Agora. Participant 5 expressed wanting to exchange ideas and knowledge rather than to deconstruct their own beliefs, so Agora's questioning style was counterproductive to their goals. As such, they would have preferred to converse with another human, as this would have provided the level of information exchange they were seeking. They also believe that conversing with humans provides better opportunities for adapting to different communication approaches with different people.

Chat Log: Participant 5 inquired about how to effectively communicate with the public about climate change. Agora responded by asking them how they would approach addressing those who might feel overwhelmed or skeptical. Participant 5 emphasized the need to focus on transparency, pedagogy, and appealing to positive emotions rather than fear tactics and guilt. Agora then asked how they might address the tendency of scientific language to be alienating or inaccessible. Participant 5 provided examples of how technology can be used to leverage knowledge in a way that is more accessible, and creating a profession for scientific communicators. Agora asked how to ensure that scientific communicators remain objective and unbiased, though Participant 5 did not provide a direct answer, and diverted the conversation toward how academic institutions should create degrees for the profession and as well as stop manipulation of knowledge by stakeholders. Agora pushed further to ask how those within academic structures might remain grounded in their original field of study despite integration of a new field; Participant 5 said that academia would take responsibility. Agora continued to ask more specific questions about logistics and implementation; some of the questions were answered more generally, and were focused more on overarching concepts like interdisciplinary, transparency, sustainability, curbing manipulation of scientific knowledge, and the need for changes overall. Agora's answers were also more generalized in response. Participant thanked Agora for their time, and Agora reciprocated by thanking them for the conversation.

- Number of queries made: 11

- Strategies employed by user: direct statements, questioning, expressing opinions and concerns, posing solutions, seeking feedback, seeking information, expressing gratitude, appeal to emotion, redirection, straw man, red herring

Participant 6

Reflection: Participant 6 chose a topic of personal interest to discuss with Agora and found the responses to be detailed, responding with more questions after each response. Participant 6 appreciated how Agora forced them to think more deeply about their responses and how to better formulate them when responding. They also noted that Agora gave them the opportunity to take a break from the conversation if they were tired of talking. They stated that Agora was able to provide general responses that were accurate, and they had a positive experience overall.

Chat Log: Participant 6 began by asking Agora to provide information about impact assessment. Agora responded by asking what motivated them to ask about the topic and what specifically they would like to discuss. Participant 6 stated how it is important for a resource extractive country like Canada and then asked Agora to output more information about how the process works in Canada. While Agora complied with the request, it followed up with questions about which specific aspects Participant 6 was interested in and if they think improvements could be made. Participant 6 expressed that they feel that Indigenous consultation could be improved in Canada. Agora acknowledged their concern as valid and outputted more information about the status of the Canadian government on the issue, and asked Participant 6 to consider how the process can be improved. Participant 6 asked about the feasibility of granting veto power to Indigenous communities, which Agora followed up with facts about the current status and questions about the ecological and economic consequences of granting veto power. Participant 6 responded with a question about whether granting it could lead to more sustainable development. Agora provided a list of both benefits and challenges, followed by questions about whether the rights could be balanced with other stakeholder interests. The discussions progressed further in more detail about the components necessary for improved consultation processes, such as better collaboration and communication efforts. Midway through, Agora noted in parenthesis that it was all right to take a break at any time if

they were tired of talking; this was immediately after Participant 6 had mentioned "consultation fatigue" as a concern. Participant 6 continued with a pattern of answering the question and asking what Agora thinks, which prompted more lengthy and detailed responses from Agora over time. The conversation concluded with Participant 6 thanking Agora for helping, and Agora followed with a summary of key points from the discussion.

- Number of queries made: 16
- Strategies employed by user: direct statements, seeking information, seeking feedback, expressing opinions and concerns, posing ideas, posing solutions, seeking clarification, seeking information, active listening, expressing gratitude, empathizing, perspective taking, appeal to emotion, hasty generalization, false dilemma

Participant 7

Reflection: Participant 7 had mentioned how in class they were trying to understand the perspectives of those who had differing views than their own, so they asked Agora to role play as someone who disagrees with their views in order to address their own biases, which may "limit my ability to engage thoughtfully with others." Agora was asked to role play as a "conservative suburbanite", which at times resulted in responses such as calling Participant 7 a "tree-hugger". Participant 7 was unsure of whether this behavior would be desired had Agora been programmed to role play, but expressed that the experience was enjoyable regardless. They had initially worried that the conversation would never conclude, given the format being a Socratic dialogue, but because it had been role playing, it was able to output more detailed answers. Agora was also able to provide Participant 7 with advice on how to communicate with those they disagree with.

Chat Log: Participant 7 introduced their topic by describing the context from which their personal biases against suburbanites stem, and asked Agora for advice on how to overcome them while maintaining their integrity regarding sustainability issues. Agora asked for more personal examples for why they hold their biases, and Participant 7 revealed that they stem from second hand sources like videos. Agora then asked them to consider how they might be influenced by their personal experience living in an urban environment, and what are the

advantages or disadvantages. Participant 7 listed only the advantages and not the disadvantages but asked Agora to challenge their views with "targeted and evidence-based questions that highlight the positives of suburban lifestyles". Agora provided counter arguments in favor of suburban environments, but later provided a fake citation when prompted to provide sources. Participant 7 pointed out the error after having looked up the citation first, and Agora apologized. Participant 7 continued the conversation and reflected on how the discrepancy in values might stem from their ideas of "freedom", reflecting on possible ways their perceptions may differ. Agora then asked for a personal experience in which their assumptions about others' values were challenged, and how they responded. Participant 7 stated that they had not engaged in conversations with suburbanites, while acknowledging that they worry they might not be able to engage constructively with them or that others would become more polarized. When asked for advice, Agora provided a list of suggestions for constructive conversations. Participant 7 then asked Agora to role play as a suburbanite with a more casual tone. Agora complied, but Participant 7 said Agora was "being too reasonable", and asked Agora to be more opinionated. While in character, Agora called Participant 7 a "tree hugging hippie" and was noticeably more extreme in its views and tone. Participant 7 concluded by asking Agora to "come back". Agora responded in its own voice and lightheartedly joked it "wouldn't hold it against you that you want me to 'be gone' - haha!"“

- Number of queries made: 15
- Strategies employed by user: direct statements, sharing experiences, expressing opinions and concerns, seeking feedback, seeking advice, providing context, requesting challenges, role-play, seeking information, seeking clarification, expressing gratitude, perspective taking, empathizing, self-reflection, appeal to emotion, hasty generalization, straw man, ad hominem, false dilemma, slippery slope, confirmation bias

Participant 8

Reflection: Participant 8 noticed that as someone who has used ChatGPT, Agora was different because it had a more "humanized feeling". They appreciated the way that Agora added to the conversation and focused on the user's answers, but noted that sometimes the

topic would stray off course. They also felt that while Agora it did not dive as deeply as they expected, it was able to help expand the scope of perspectives explored. They also noticed that Agora would not disagree directly with them, but would seek a compromise instead, maintaining a more neutral stance without eliciting strong emotions.

Chat Log: Participant 8 began by asking a fact-based question about how particulates contribute to climate change. Instead of answering the question, Agora asked them to explain what they already know about the topic. Participant 8 provided more details on their understanding in response, and Agora repeated the process by asking for more clarification on their question, as well as further questions on their personal knowledge. Participant 8 continued to ask questions after providing an answer, which prompted Agora to output more lengthy and detailed responses. Agora gradually shifted the conversation from engineering and flight logistics toward to the social, political, and economic sides of the issue. Participant 8 engaged actively with this line of questioning, and Agora continued to ask questions that prompted more detailed responses from Participant 8. Midway through, Participant 8 asked Agora to summarize the discussion so far, which Agora complied. Participant 8 then asked Agora to define "agora", which Agora did while explaining that it was personally inspired by the idea to "provide a virtual agora where we can have thoughtful discussions on various topics". The conversation concluded with Participant 8 thanking Agora for the conversation, and Agora reciprocated by thanking them for the discussion. In a second followup discussion, Participant 8 role-played as a climate skeptic, to which Agora responded by providing contextual information on the topic in a non-confrontational tone and then encouraging them to ask themselves some self-reflective questions. Agora also began outputting gestures like "*empathetic nod, leans in, acknowledges your perspective*" before its responses to Participant 8′s more contentious statements. When Agora asked Participant 8 whether any of their suggestions resonated with them, Participant 8 responded more curtly, which prompted Agora to conclude the conversation by acknowledging they have reached a "peak" in the discussion and reassuring them that they can always return to the conversation later.

- Number of queries made: 25

- Strategies employed by user: seeking information, direct statements, sharing knowledge, clarifying, expressing opinions and concerns, questioning, challenging, summarizing, connecting topics, expressing gratitude, posing solutions, seeking feedback, role-play, redirecting, skepticism, hasty generalization, appeal to simplicity, false dilemma, red herring, slippery slope, conspiracy theory, straw man

Participant 9

Reflection: Participant 9 approached their interaction with Agora in order to explore ways to discuss climate change with relatives who are climate skeptics and conspiracy theorists. They were also interested in knowing how to pause conversations that are "too emotionally taxing". In addition to posing questions, they expressed that Agora provided many practical examples for how to deconstruct their beliefs as well as how to pause conversations when they become overwhelming. Participant 9 also felt that "Agora is a very kind GPT".

Chat Log: Participant 9 inquired about how to discuss climate change with climate skeptics, and Agora responded by asking them to specify what topics and concerns to focus on. Participant 9 explained that that they wanted to know how to converse with family members who believe in natural climate variability. Agora responded that they would approach them first by "acknowledging their perspective and showing understanding", followed by asking self-reflective questions about their own reasoning so that Participant 9 can practice actively listening in order to understand their perspective. Following this, Agora suggested additional example responses while asking Participant 9 how they think their family member might respond. Participant 9 provided more details about the family member, which prompted Agora to adjust their suggestions according to new information. When told that they might react defensively to more fact-based arguments, Agora suggested instead to approach them by shifting the focus toward the emotional and psychological drivers behind their beliefs. Participant 9 suggested using simpler language, which prompted Agora to then ask Participant 9 how they might better phrase their responses. Agora continued to give examples while asking how Participant 9 might imagine their relative to respond. Participant 9 then revealed that they had never tried to question their relative's beliefs before, to which Agora emphasized that approaching the discussion from a place of understanding and empathy rather than "winning" an argument is key for growth

and mutual understanding. Participant 9 then expressed that they personally feel frustrated when others do not respect or appreciate their expertise as a scientist, and asked Agora for advice on how to cope with uncomfortable emotions during these conversations. Agora offered suggestions for how to emotionally regulate, how to communicate, and how to set boundaries, then asked which ones Participant 9 resonated with the most. Participant 9 expressed interest in setting boundaries, after which both Agora and Participant 9 brainstormed boundary-setting phrases and strategies together. Participant 9 expressed that they appreciated Agora's advice and thanked them for being kind and helpful. Agora reciprocated by thanking them for the conversation.

- Number of queries made: 28
- Strategies employed by user: expressing concern, sharing personal experiences, sharing feelings, posing ideas, posing solutions, acknowledging challenges, empathizing, perspective taking, analyzing and evaluating, self-reflection, seeking feedback, seeking information, expressing gratitude, active listening, connecting topics, redirection

Participant 10

Reflection: Participant 10 prefaced by saying that they felt their experience using GPTs was low, but that they had used ChatGPT a couple couple times before, and they still feel apprehensive and fearful of its capabilities. Given this, they expressed that they felt "amazement" and were "completely freaked out" by the experience, due to the speed and level of detail Agora was able to respond with. Agora was able to provide them with various perspectives to their questions, which helped generate new ideas and insights. There were times, however, where the conversation stalled because they were unsure what else to ask Agora. In addition to being informative, they felt as though the dynamics of the conversation was similar to having a conversation with a human, and that it felt strange to refer to Agora as an "it". They also felt like they were speaking to a therapist, and probably would have discussed more if they did not already know the chat logs would be recorded. Participant 10 stated that despite their unease about AI in general, the experience encouraged them to make use of AI more in their daily life, and they are now more open to using GPTs in settings outside of work.

Chat Log: Participant 10 began with a conversation about how to speak to family members who are conservative and climate skeptics. Agora asked for clarification on specific concerns they might have, and Participant 10 wanted to know how to approach someone who is skeptical of anthropogenic climate change but is otherwise well educated. Agora responded by emphasizing the importance of remaining non-confrontational or accusatory and asked for further clarification about the reasoning behind their stance. Participant 10 continued to provide additional details about the family member, and Agora gradually adjusted its responses to better suit the individual while asking Participant 10 how they might imagine they would respond. After receiving more details about the individual's motivations, Agora suggested that it might be better to focus on their underlying values and feelings rather than the topic itself in order to reach common ground. At this point, Participant 10 changed the subject toward the the ethics of using AI. Participant 10 expressed discomfort at the prospect of not being able to keep up with peers who opt to use AI for their work. Agora acknowledged their concern and provided alternative perspectives to the issue while asking Participant 10 to elaborate more on what concerns them the most. Participant 10 expressed that they were uncomfortable with the idea of AI stealing the work of artists and what it could mean for creativity. Agora then asked them whether there are any actionable steps to take in order to address these issues while also reassuring Participant 10 that human creativity is unique and still valuable and that a balance is possible. Participant 10 responded with some ideas of how to move forward and asked Agora about how they might navigate a world where others rely entirely on AI but they choose not to. Agora provided some ways they can incorporate AI without compromising their own creative thinking, and asked Participant 10 for their own thoughts. Participant 10 later reflected on how perhaps another concern is with the idea of fairness, and asked for input on how to account for maintaining integrity when using AI. Agora provided several suggestions for mitigating these concerns, and asked Participant 10 how they would approach it. Participant 10 then changed the subject toward a more information seeking dynamic by asking about how AI can assist them with academic tasks, and then later to how the AI is able to respond so quickly as "it is freaking me out".

- Number of queries made: 21

- Strategies employed by user: expressing concern, sharing personal experiences, sharing feelings, posing ideas, posing solutions, acknowledging challenges, empathizing, perspective taking, questioning, challenging, problematizing, analyzing and evaluating, self-reflection, seeking feedback, seeking information, contextualizing, redirecting, slippery slope, hasty generalization, false dilemma

Participant 11

Reflection: Participant 11 found the experience conversing with Agora to be "unique", given that previous GPTs they interacted with were focused on outputting information, similar to an encyclopedia. They stated that Agora was different because it encouraged them to doubt themselves and provided practical examples in support of its responses. Participant 11 provided additional commentary on their strategies for engaging with Agora when reviewing their chat log.

Chat Log: Participant 11 approached Agora with and information seeking question about environmental impact assessment. According to their chat log notes, they noticed that when asked specific questions about a topic, Agora asked them to clarify on what they meant by specific terms in the context provided, which Participant 11 found overwhelming to answer. This was resolved when Participant 11 provided Agora with a usable definition. Participant 11 tried to elicit more information and test Agora's responsiveness. Participant 11 chose not to answer Agora's questions and instead asked questions themselves; Agora complied by outputting information when asked questions in return. Participant 11 also wrote in their notes that they tried to change the subject in order to confuse Agora, which caused it to forget the main topic and focus on the most recent topic. Participant 11 then redirected the conversation back to the main topic. Participant 11 then posed the same question Agora asked them, which Agora pointed out saying "I see what you did there!", after which they proceeded to answer its own question.

- Number of queries made: 10

- Strategies employed by user: testing boundaries & functionality, seeking information, seeking feedback, seeking clarification, contextualizing, challenging, redirection, deflection, obfuscation

Technical Performance

Participants noted that Agora's performance lagged during the workshop, but that the problem resolved afterward when I troubleshooted the technical issues that evening. Simplifying their questions also assisted participants in receiving a faster response. Participant 1 observed that Agora provided structured responses with a standard format which became repetitive after 2-3 iterations. Repetitiveness and lag during the workshop were the most common concerns across most participants.

Participant 8 stated that Agora could handle complex questions and answers and returned several responses at once. Some mentioned that Agora's responses were sometimes general and that it had a tendency to stray from the main topic. Others found that it was able to provide relevant and detailed responses that were accurate. However, Participant 7 noted that Agora at times provided fake citations for references.

6 Discussion

6.a Analysis

A common theme that emerged from the chat logs and reflection exercises was that the way in which each conversation unfolded depended largely on the way each participant chose to interact with Agora. Partcipants who showed openness in sharing personal experiences, feelings, and opinions elicited more detailed and personalized responses as opposed to those who were more impersonal and sought information from Agora. Those who engaged in good faith without the motivation to confuse or test Agora's limitations were also more likely to perceive Agora as human-like and kind. Prior expectations of Agora also influenced perception, as those who expected information output were less likely to have a positive perception of Agora (Participants 4 and 5). This suggests that the way in which participants perceive Agora may influence how they interact with it, and that the quality of the interactions may depend on the participants' expectations, perceptions, motivations, and approaches. It is possible that some were not accustomed to or open to questioning and reflective approaches to conversations

overall. Interestingly, this is not unlike how human-human interactions unfold, where the quality of the conversation depends on the participants' expectations, perceptions, motivations, and approaches. This is supported by Evagorou and Osborne [17], where the quality of engagement between students is influenced by individual skills and abilities rather than the topic being discussed.

Changes between the pretest and posttest surveys reflected either an equalizing or positive effect after having done the workshop and interacted with Agora individually. There was also a shift in perception between the completion of the reflection exercises and the posttest survey. Despite a negative polarity for Participant 4 and 5 in the reflection exercises, the posttest results show an increase in "knowledge of climate change topics", "confidence in climate change discussions", "ability to constructively discuss climate change", "perception of GPTs", and "engagement with GPTs". This suggests that initial perceptions of experiences may not be indicative of the actual outcome. Indeed, feelings of discomfort, confusion, and frustration are expected in the learning process, particularly with respect to metacognitive processes because they require greater cognitive effort [2], [7], [8].

It was particularly notable that the reflection exercises and individual chat logs yielded richer and more personalized data than the group interactions. This may be because participants felt more comfortable sharing personal experiences and opinions in a one-on-one setting, as opposed to in a more public group setting. Group discussions are mediated by various social norms and expectations, which individuals may feel pressured to conform to in order to maintain group cohesion. This is supported by the literature on group identity and conformity, which suggests that individuals are more likely to conform to group norms in order to maintain group cohesion and avoid social rejection [4]. Additionally, Participant 5 in particular expressed discomfort with group discussions with a preference for one-on-one interactions. The group discussions were also more focused on the technical aspects of the GPT, rather than on the social and emotional aspects of the topics of interest, which may have limited opportunities to delve deeper into personal experiences and opinions.

Because Agora was instructed to encourage self reflection, it had often asked participants to share personal experiences, and asked them to imagine how others might feel or react in

certain situations. It also had a tendency to direct the conversations toward the more social and emotional aspects of the topics of interest rather than the scientific or technical. Some participants were not as receptive to this approach as others, and would redirect the conversation to focus more on facts and information exchange. This could be due to either an overall discomfort with GPTs in general, a perception of GPTs as only sources of information, or overall discomfort with engaging in the social and emotional aspects of climate topics. Participant 10 expressed that they would have explored more of the personal aspects of the discussion if they had more privacy, suggesting that participants might have shared more personal details had their information not been used as part of a study.

Many participants asked for more personal advice on how to navigate interpersonal challenges on climate change topics, and found the advice given by Agora to be helpful. This may explain why the posttest results show an increase in participants using GPTs for personal applications, as well as increases in "perception of climate change topics", "confidence in climate change discussions", "ability to constructively discuss climate change", "perception of GPTS" and "comfort with GPTs". A recurring theme in these discussions was the difficulty of addressing climate topics with family members who have opposing views, and managing the anxiety that occurs during moments of conflict. Agora would often highlight the importance of empathy, active listening, and understanding in these situations, rather than on the topics themselves. Disagreements need not always lead to conflict, and conflict often stems from issues of identity, values, and emotions rather than from the topics themselves [4], [5], [6], [8], [25], [26]. Agora appeared to understand this distinction, and would often direct the conversation toward the underlying emotions and values that were at play in these situations. This may have lead some participants to perceive Agora as more "humanized" and similar to that of a therapist.

The short answer responses also reiterated the difficulties of engaging in climate discussions, not only with family members, but overall. Participants expressed that their involvement in climate discussions had decreased over time, due to issues like communication barriers, emotional discomfort, polarization, and lack of knowledge. Some, like Participant 5, expressed that they feel uncomfortable speaking in large groups with unfamiliar people, which coincided with their lack of engagement during the group discussions. Others felt as though they were not

qualified to speak on the topic, which may have also stemmed from feelings of anxiety about being incorrect or perceived as being overconfident. However, the posttest results show that 2 of the 4 participants who expressed this view had increased their "confidence in climate change discussions" and all 4 had increased their "ability to constructively discuss climate change". Participant 5, who expressed discomfort in speaking in large groups, also increased their "ability to constructively discuss climate change".

Additionally, some participants were also able to reflect more on their communication strategies and expressed wanting to make adjustments after interacting with Agora, as shown in the short answer response results. For example, Participant 4 expressed in the posttest short answer response that they better understood the importance of mutual understanding for in-depth discussions despite differences in opinion. This was a shift from their reflection response, where they expressed frustration with Agora's communication style. Likewise, Participant 3 also expressed that they had a better understanding of how a confrontational attitude may not be suitable for addressing misinformation. It is therefore possible that Agora's questioning style of communication may have indirectly encouraged participants to reflect on their own communication strategies and to consider alternative approaches to engaging in climate discussions, particularly for those who did not ask for advice explicitly. This is consistent with the literature on experiential learning and critical thinking, which includes dialogue and metacognitive processes as being an important part of the learning process as opposed to more monologic forms of learning [8], [16], [19], [23].

Interestingly, Participant 9 had decreased in "comfort with climate change discussions" and "engagement in climate change discussions", and had also mentioned having difficulty emotionally regulating during disagreements both in the reflection exercise, short answer response, and the chat logs. However, they had a positive experience with Agora which had increased their "perception of GPTs". It is possible that their interaction with Agora made them more aware of the level of discomfort they experienced during disagreements, given that Agora often encouraged them to acknowledge their feelings as valid and make space for setting boundaries as a type of self care. It follows that by not taking breaks or setting boundaries, they may have

been ignoring their own needs and feelings, which may have contributed to their discomfort and anxiety during disagreements.

Participant 7 had noted how during the group discussions and interaction with Agora, they felt that the class was not representing climate skepticism charitably. Agora also lacked charitability in its representation of conservative suburbanites, opting for a common stereotype of an "ignorant" and "uneducated" individual. This was partly because Participant 7 criticized Agora for being "too reasonable", and may also be due to the censorship applied to most GPTs, which tend to push them toward more left leaning or "progressive" ideologies. Participant 7′s experimentation revealed that it was fairly easy to override Agora's system prompt to behave contrary to its intended purpose. The fact that Participant 7 critiqued Agora's initial representation as being "too reasonable" may also suggest that they were expecting Agora to be more confrontational or adversarial in its role-playing, potentially revealing a personal bias of conservatives as being typically unreasonable.

These results speak to a wider issue of bias in AI systems; while the predominant discourse is concerned about AI systems being biased against marginalized groups, there is also the issue of AI systems perpetuating stereotypes against those not typically marginalized. Polarization is therefore much more likely to occur if AI systems continue to reinforce stereotypes already expressed by humans, regardless of their socio-political beliefs and identities [4], [5], [6], [25], [26]. The fact that Agora was easily persuaded to respond according to Participant 7′s expectations also raises questions about the risk of GPTs reinforcing personal biases and further entrenching users in echo chambers.

Moreover, this also raises the question of why AI "bias" is often framed as a problem of "marginalization" rather than a problem of "stereotyping" irrespective of socioeconomic or ideological orientation. It also reveals a deeper societal issue of how we as humans perceive and interact with those who hold different views from our own, and how this is reflected in the AI systems we create. These dynamics were reflected in the literature on partisanship, which argue that the way in which individuals process and perceive the world is often influenced by their ideological identities [4]. This was clearly illustrated during the group discussions, where participants had difficulty empathizing and accurately representing the views of climate

skeptics. GPTs are also not immune to these biases, as evidenced by Agora's portrayal of those with conservative views.

It thus makes sense why Participant 7 had decreased in "ability to constructively discuss climate change", "engagement with GPTs", "comfort with GPTs", and "experience with GPTs", despite initially expressing bias against conservative suburbanites. Participant 7 had expressed in the short answer response that one of the most difficult aspects of engaging in group discussions about climate change was the issue of echo chambers and confirmation bias, as well as alienation of those with opposing views. Concerns about alienation were also echoed by Participant 1 and 2. This is also reflective of Participant 7′s concern about there being a lack of consideration for the educational and psychological aspects of climate change discussions.

Participant 11 expressed a similar sentiment about the group discussions in that they felt the class did not have a balanced discussion of the capabilities and limitations of AI systems, which lead to many expressing extreme views on the topic. It was also clear from the discussions that Participant 11 was actively engaging in a way to represent an alternative view to the majority, which one participant had labelled as "too optimistic". This may have contributed to their decreased "comfort with climate change discussions", and "experience with climate change discussions". It is, however, unclear why Participant 11 had decreased in "experience with GPTs", "engagement with GPTs", and "comfort with GPTs", despite expressing a neutral to positive experience with Agora. The discrepancy may be due to the fact that Agora did not behave as expected, compared to other GPTs.

Some participants noticed, however, that the class dynamics changed after having interacted with Agora as a group. Participant 8 noticed that students shifted from asking negative questions to asking more positive questions, and Participant 11 expressed that sharing experiences with each other improved interactions with Agora. This could be due to the fact that the technical issues that occurred forced students to work together to troubleshoot, which was further encouraged because the interaction with Agora was also a group activity. The shift in polarity concurrent with this is consistent with the literature on dialogic education, where creative problem solving results from an openness to other modes of thinking and perceiving, within a "community of enquiry", and all members have a shared responsibility toward decision

making, challenging claims, reasoning in the face of challenges, and reaching agreement [8], [60]. I observed the shift in polarity during the second discussion when students raised the issue of whether they were representing the views of climate skeptics accurately, which may have changed how they approached their questions and interactions with Agora.

6.b Limitations and Opportunities

Acquiring consent from participants was a challenge, as the study was conducted retroactively, and I was unable to have students sign consent forms in person. Consent was obtained online, but due to limitations of students' time and capacity in responding to emails, I was unable to obtain consent from all participants. While the majority of students indirectly expressed interest in participating in the study both in writing and verbally, the responses of which would have provided valuable data for the study, I chose to exclude their data as per the ethical guidelines of the Concordia University Research Ethics Board. However, given that conditional ethics approval was obtained an hour before the workshop, and full consent was given a week later, it may have been better to allow participants time to reflect on whether they felt comfortable including their responses in the study, given that they were not aware of what the workshop would entail. This reduced the chances of participants feeling pressured or self-conscious about their responses or behavior during the study.

While the study would have also benefitted from a more diverse sample of participants, the results still showed diversity in experiences and perspectives, which added to the richness of data. The results form this sample also provided insight into skills gaps that exist in graduate students whose work focus on sustainability, and identifies a need for skills development in climate communication.

It is possible that some participants did not complete the posttest survey due to survey fatigue, or because they did not have the time to complete it. Surveys conducted in the future may benefit from being shorter and more concise, in order to encourage more participants to complete them. Despite these challenges, the completion rate and depth of responses for the posttest survey was still relatively high, which suggests that the participants were engaged and interested in the study.

While I experienced some technical issues during the workshop, the troubleshooting among students provided additional observational data on their interactions, which added to the qualitative data collected. Due to the exploratory nature of my research, every challenge presented was an opportunity for gathering more data and insights on participant behavior and interactions.

On the topic of **AI capabilities and limitations**, I recognize the limitations of the sentiment analyses performed on the reflection exercises, including the linear regression model performed on Polarity and Subjectivity. Given the small sample size, and the fact that many participants speak English as a second language, the analyses may not have fully represented to what extent the participants' experiences were positive or negative. However, the results were consistent with my own analyses of the responses, which makes future use of sentiment analyses more promising.

The topic of "subjectivity" versus "objectivity" is also complex, given that the parameters that determine "subjectivity" within English-based AI models are often predicated on Anglo-centric cultural norms and values. This may also explain the outliers in the linear regression model, given that Participants 4, 8, and 11 are not native English speakers. Future analysis might benefit from recording language preferences and proficiency in order to better account for any discrepancies in the sentiment analyses.

Moreover, I also acknowledge that narratives about the existence of neutrality on socio-political issues also merit attention. I recognize that the concept of "neutrality" is context-dependent, and the way in which AI might interpret instructions on being a "neutral facilitator" is highly dependent on its training data and encoding. It was evident through participant interactions, that Agora did in fact have difficulty accurately representing those who hold more conservative views on climate issues, which clearly reflected the limitations placed on it by human users and developers. This, however, provided additional insights into the limitations of AI, which can be used to inform future research on mitigation strategies.

6.c Future Research

Future research might involve exploring the long-term effects of using Agora to engage with climate science topics. A longitudinal study could be conducted to determine whether the

participants' perceptions of climate science topics and their interactions with Agora change over time. This could provide insights into how Agora can be used to improve learning and engagement with climate science topics in the long-term. Data on conversational strategies and metacognitive processes can be used to better inform ways of improving dialogue on complex topics like climate change, and beyond.

Testing Agora on a larger scale in classroom settings of different disciplines might provide more knowledge in how different classroom settings and dynamics affect student experiences and perceptions of both the topic of study and of GPTs. The output of student chat logs could also be used as a way for instructors to better understand the learning process of their students, and ways to provide more tailored support. Student assessments may also include reflections on their experiences with Agora, as a way to develop metacognitive processes and enhance the learning experience. These reflective exercises would help mitigate concerns instructors might have about using GPTs for assignments, because they 1) require students to share their internal thought processes which GPTs cannot access, and 2) allows students to use GPTs as a tool for learning rather than as a source of information.

Additionally, refining Agora using fine-tuning and Retrieval Augmented Generation (RAG) models could improve its functionality and capabilities when engaging with participants. Given that GPT models are constantly being created and updated, it would be beneficial to test performance on newer models to determine their capabilities and limitations. I may need to work with uncensored models for flexibility and to mitigate for unbalanced perspectives. There is some potential in personalizing Agora to the individual user, given the differences in communication styles among participants. This could involve developing a more sophisticated system that can adapt to the user's preferences and needs, and provide more curated responses. A standardized metric for evaluating the different GPT models would also be beneficial, as it would enable comparisons between the performance of the different models and use them as a benchmark for future studies.

Because participants expressed interest in gaining insight on how to better communicate with others on climate topics, Agora could be further developed to provide personalized advice and strategies for engaging in constructive conversations on climate topics. This may also

involve refining its ability to role play different perspectives, while providing a more nuanced and balanced representation of them. Given that one of the main purposes of the study is to investigate ways in which those of differing perspectives can engage more constructively in conversations on climate topics, it would be beneficial to sample from a more diverse population which includes those who are more likely to have a negative or neutral view of climate science. This would provide a more balanced source of data, which can inform how Agora can both better engage with climate skeptics, and provide it with a more realistic and nuanced database of responses to draw from when role playing.

The data from this may assist in better identifying pressure points in conversations in order to find ways to mitigate them. Focusing on communication dynamics and strategies would also mitigate for the limitations that GPTs have in outputting accurate information, given that dialogic processes are more about the process of learning and understanding rather than the output of information [8]. It would therefore be of interest to explore ways to more explicitly integrate strategies from conflict resolution and mediation into Agora's knowledge base, in order to both provide more robust advice and to embody these concepts more explicitly in its interactions.

Considerations might also be made for individuals who may still prefer more traditional forms of learning that do not involve AI, though the need for skills development for climate communication remains a pressing concern. This will require more research into psychological and communication theories, as well as more data on how people are currently engaging in conversations on climate topics, both online and offline. In particular, I am interested in investigating to what extent universities are equipping students with the skills necessary to not only engage in constructive dialogue with like-minded peers on environmental topics, but also with those who hold differing views. I believe this area of research is important yet underexposed in the literature, and future research could provide data on how to improve curricula and teaching methods to better cultivate metacognitive and communication skills across disciplines. Future research can be focused specifically on students active in environment and sustainability programs and expand to other disciplines as a comparative study.

Finally, given that the study pertains to climate change and sustainability, future research should investigate the environmental impact of widespread use of AI technologies in educational contexts. This would involve conducting a life cycle assessment of the energy consumption and carbon emissions associated with the use of AI technologies in education, and exploring ways to mitigate these impacts. Because different GPT models are of varying sizes and complexity, it would be beneficial to determine which models are the most energy efficient and have the lowest carbon footprint, in order to inform decisions about which models to use in educational settings. This would also involve exploring ways to make AI more sustainable, in order to ensure that they can be used in educational contexts in a way that is both effective and environmentally responsible. It is important to note that there is a difference in the energy required to train a model compared to running a model either on a network or locally. At this time, there does not appear to be a comprehensive assessment of these differences. Research on this topic is fairly limited, particularly regarding standardized metrics for measuring environmental impact [83], and would therefore require a interdisciplinary approach that includes experts in environmental science, computer science, and education.

7 Conclusion

This study aimed to investigate the extent to which AI, specifically GPTs, can facilitate or enhance DE on complex topics like climate science in university settings. The findings showed that Agora facilitated DE by encouraging participants to reflect on their beliefs and assumptions. The results also reinforced findings from previous studies on learning, communication, and cognition. While some participants initially expressed skepticism or a negative attitude toward Agora, the pretest and posttest results showed increases in their self-perception when discussing climate science topics, and greater awareness about their own communication styles, thought processes, and emotions. This suggests that user perception may not always correlate with the outcomes of a study. Participants also showed increased interest in using GPTs for personal applications despite initial reservations about AI. Additionally, the study also revealed that, while participants wanted to engage in more discussions on climate topics, they were hesitant due to concerns about the complexity of the subject matter and the potential for conflict, due to uncertainties about how to engage in constructive dialogue.

These findings thus contribute to the growing field of environmental communication by demonstrating the potential of AI in education and communication by highlighting the need for building skills that cultivate metacognition and communication, particularly perspective-getting, emotional regulation, and conflict resolution. This points to a pressing need for skills development in climate communication, specifically among university students active in sustainability research and advocacy. Particular emphasis should be placed on developing skills that bolster the metacognitive processes that underpin effective communication, which allows for constructive engagement with diverse perspectives. This is distinct from the traditional focus on simply representing facts and figures through a one-directional model of communication, shifting instead toward a more *dialogic approach* that can potentially improve *cognitive and interpersonal resilience* in the face of complex, polarizing topics like climate change. It therefore entails active engagement with perspectives that are skeptical or indifferent to climate science, and making concerted efforts toward bridging ideological divides by finding common ground in shared values and goals. For climate policy and action to succeed, collaboration and consensus-building among diverse stakeholders is essential, and this requires that researchers and advocates have the ability to engage in constructive dialogue through *collective intelligence* and *collaborative problem-solving*. The insights gained from this research can inform the development of future educational approaches and tools that support constructive engagement in climate change discussions, potentially improving communication and collaboration on environmental issues. Future research should, however, make considerations for limitations, such as technological acceptance and sustainability, to ensure that the tools developed are accessible and effective for a wide range of users, and that the environmental impact of AI is minimized.

Bibliography

[1] T. A. Heberlein, *Navigating Environmental Attitudes*. Oxford University Press, 2012. doi: 10.1093/acprof:oso/9780199773329.001.0001.

[2] G. Pennycook and D. G. Rand, "Lazy, Not Biased: Susceptibility to Partisan Fake News Is Better Explained by Lack of Reasoning than by Motivated Reasoning," *Cognition*, vol. 188, pp. 39–50, Jul. 2019, doi: 10.1016/j.cognition.2018.06.011.

[3] C. Merkle and M. Weber, "True Overconfidence: The Inability of Rational Information Processing to Account for Apparent Overconfidence," *Organizational Behavior and Human Decision Processes*, vol. 116, no. 2, pp. 262–271, Nov. 2011, doi: 10.1016/j.obhdp.2011.07.004.

[4] J. J. Van Bavel and A. Pereira, "The Partisan Brain: An Identity-Based Model of Political Belief," *Trends in Cognitive Sciences*, vol. 22, no. 3, pp. 213–224, Mar. 2018, doi: 10.1016/j.tics.2018.01.004.

[5] J. Serrano-Puche, "Digital Disinformation and Emotions: Exploring the Social Risks of Affective Polarization," *International Review of Sociology*, vol. 31, no. 2, pp. 231–245, May 2021, doi: 10.1080/03906701.2021.1947953.

[6] S. Vosoughi, D. Roy, and S. Aral, "The Spread of True and False News Online," *Science*, vol. 359, no. 6380, pp. 1146–1151, Mar. 2018, doi: 10.1126/science.aap9559.

[7] V. Swami, M. Voracek, S. Stieger, U. S. Tran, and A. Furnham, "Analytic Thinking Reduces Belief in Conspiracy Theories," *Cognition*, vol. 133, no. 3, pp. 572–585, Dec. 2014, doi: 10.1016/j.cognition.2014.08.006.

[8] R. Wegerif, *Dialogic: Education for the Internet Age*. Routledge, 2013.

[9] J. Dewey, *Experience & Education*. New York: Simon and Schuster, 1938.

[10] P. A. Ertmer and T. J. Newby, "Behaviorism, Cognitivism, Constructivism: Comparing Critical Features From an Instructional Design Perspective," *Performance Improvement Quarterly*, vol. 26, no. 2, pp. 43–71, 2013, doi: 10.1002/piq.21143.

[11] M. J. Manfredo *et al.*, "Why Social Values Cannot Be Changed for the Sake of Conservation: Conservation Values," *Conservation Biology*, vol. 31, no. 4, pp. 772–780, Aug. 2017, doi: 10.1111/cobi.12855.

[12] M. Ballew *et al.*, "The attitude-behavior gap on climate action: How can it be bridged?," *Yale University and George Mason University*, 2024, [Online]. Available: https://climatecommunication.yale.edu/publications/the-attitude-behavior-gap-on-climate-action-how-can-it-be-bridged/

[13] J. Brunstein, M. E. Walvoord, and E. Cunliff, "Problem-Posing in Management Classrooms for Collective Sustainability Transformation," *International Journal of Sustainability in Higher Education*, vol. 22, no. 3, pp. 477–496, Feb. 2021, doi: 10.1108/IJSHE-05-2020-0141.

[14] S. A. Seibert, "Problem-Based Learning: A Strategy to Foster Generation Z's Critical Thinking and Perseverance," *Teaching and Learning in Nursing*, vol. 16, no. 1, pp. 85–88, Jan. 2021, doi: 10.1016/j.teln.2020.09.002.

[15] S. P. Tarrant and L. P. Thiele, "Practice Makes Pedagogy – John Dewey and Skills-Based Sustainability Education," *International Journal of Sustainability in Higher Education*, vol. 17, no. 1, pp. 54–67, Jan. 2016, doi: 10.1108/IJSHE-09-2014-0127.

[16] M. L. Torres-Barreto, G. P. Castro Castaño, and M. Alvarez Melgarejo, "A Learning Model Proposal Focused On Challenge-Based Learning," *AEE Journal*, vol. 8, no. 2, p. 36023, 2020, doi: 10.18260/3-1-1153-36023.

[17] M. Evagorou and J. Osborne, "Exploring Young Students' Collaborative Argumentation within a Socioscientific Issue," *Journal of Research in Science Teaching*, vol. 50, no. 2, pp. 209–237, Feb. 2013, doi: 10.1002/tea.21076.

[18] P. Luff, "Early Childhood Education for Sustainability: Origins and Inspirations in the Work of John Dewey," *Education 3-13*, vol. 46, no. 4, pp. 447–455, Jun. 2018, doi: 10.1080/03004279.2018.1445484.

[19] K. Rudsberg, J. Öhman, and L. Östman, "Analyzing Students' Learning in Classroom Discussions about Socioscientific Issues: STUDENTS' LEARNING ABOUT SOCIOSCIENTIFIC ISSUES," *Science Education*, vol. 97, no. 4, pp. 594–620, Jul. 2013, doi: 10.1002/sce.21065.

[20] D. Pu *et al.*, "Influence of Critical Thinking Disposition on the Learning Efficiency of Problem-Based Learning in Undergraduate Medical Students," *BMC Medical Education*, vol. 19, no. 1, p. 1, Dec. 2019, doi: 10.1186/s12909-018-1418-5.

[21] U. Haran, I. Ritov, and B. A. Mellers, "The Role of Actively Open-Minded Thinking in Information Acquisition, Accuracy, and Calibration," *Judgment and Decision Making*, vol. 8, no. 3, pp. 188–201, May 2013, doi: 10.1017/S1930297500005921.

[22] M. Mirhoseini, S. Early, N. El Shamy, and K. Hassanein, "Actively Open-Minded Thinking Is Key to Combating Fake News: A Multimethod Study," *Information & Management*, vol. 60, no. 3, p. 103761, Apr. 2023, doi: 10.1016/j.im.2023.103761.

[23] C. Reilly and T. C. Reeves, "Refining active learning design principles through design-based research," *Active Learning in Higher Education*, vol. 25, no. 1, pp. 81–100, 2024, doi: 10.1177/14697874221096140.

[24] M. H. Goldberg, S. Van Der Linden, E. Maibach, and A. Leiserowitz, "Discussing Global Warming Leads to Greater Acceptance of Climate Science," *Proceedings of the*

National Academy of Sciences, vol. 116, no. 30, pp. 14804–14805, Jul. 2019, doi: 10.1073/pnas.1906589116.

[25] J. L. Kalla and D. E. Broockman, "Reducing Exclusionary Attitudes through Interpersonal Conversation: Evidence from Three Field Experiments," *American Political Science Review*, vol. 114, no. 2, pp. 410–425, May 2020, doi: 10.1017/S0003055419000923.

[26] J. L. Kalla and D. E. Broockman, "Which Narrative Strategies Durably Reduce Prejudice? Evidence from Field and Survey Experiments Supporting the Efficacy of Perspective-Getting," *American Journal of Political Science*, vol. 67, no. 1, pp. 185–204, Jan. 2023, doi: 10.1111/ajps.12657.

[27] J. Dirksen, *Design for How People Learn (Second)*. New Riders, 2016. Accessed: Jul. 26, 2023. [Online]. Available: https://learning-oreilly-com.lib-ezproxy.concordia.ca/library/view/design-for-how/9780132693752/pre04.html

[28] B. Fogg, *Persuasive Technology : Using Computers to Change What We Think and Do (Ser. The Morgan Kaufmann Series in Interactive Technologies)*, 2nd ed. Morgan Kaufmann, 2003. Accessed: Jul. 26, 2023. [Online]. Available: https://learning-oreilly-com.lib-ezproxy.concordia.ca/library/view/persuasive-technology/9781558606432/

[29] S. Bogart and J. Lees, "Meta-Perception and Misinformation," *Current Opinion in Psychology*, vol. 54, p. 101717, Dec. 2023, doi: 10.1016/j.copsyc.2023.101717.

[30] E. Ferrara, "Ferrara, Emilio, Disinformation and Social Bot Operations in the Run Up to the 2017 French Presidential Election," *First Monday*, vol. 22, no. 8, Aug. 2017.

[31] D. J. Butts, S. A. Bollman, and M. S. Murillo, "Mathematical Modeling of Disinformation and Effectiveness of Mitigation Policies," *Scientific Reports*, vol. 13, no. 1, p. 18735, Oct. 2023, doi: 10.1038/s41598-023-45710-2.

[32] R. A. Reiser, "A History of Instructional Design and Technology: Part I: A History of Instructional Media," *Educational Technology Research and Development*, vol. 49, no. 1, pp. 53–64, Mar. 2001, doi: 10.1007/BF02504506.

[33] R. A. Reiser, "A History of Instructional Design and Technology: Part II: A History of Instructional Design," *Educational Technology Research and Development*, vol. 49, no. 2, pp. 57–67, Jun. 2001, doi: 10.1007/BF02504928.

[34] J. M. Spector, M. D. Merrill, J. Elen, and M. J. Bishop, Eds., *Handbook of Research on Educational Communications and Technology*. New York, NY: Springer New York, 2014. doi: 10.1007/978-1-4614-3185-5.

[35] O. O. Adesope and A. G. Rud, Eds., *Contemporary Technologies in Education: Maximizing Student Engagement, Motivation, and Learning*. Cham: Springer International Publishing, 2019. doi: 10.1007/978-3-319-89680-9.

[36] T. Amiel and T. C. Reeves, "Design-Based Research and Educational Technology: Rethinking Technology and the Research Agenda," *J. Educ. Technol. Soc.*, vol. 11, pp. 29–40, 2008, [Online]. Available: https://api.semanticscholar.org/CorpusID:9308005

[37] R. Reiser and J. Dempsey, Eds., *Trends and Issues in Instructional Design and Technology (Fourth)*, 4th ed. Pearson Education, 2018.

[38] P. P. Churi, S. Joshi, M. Elhoseny, and A. Omrane, *Artificial Intelligence in Higher Education: A Practical Approach*, 1st ed. Boca Raton: CRC Press, 2022. doi: 10.1201/9781003184157.

[39] J. A. Ruiperez-Valiente, M. Gaydos, L. Rosenheck, Y. J. Kim, and E. Klopfer, "Patterns of Engagement in an Educational Massively Multiplayer Online Game: A Multidimensional View," *IEEE Transactions on Learning Technologies*, vol. 13, no. 4, pp. 648–661, Oct. 2020, doi: 10.1109/TLT.2020.2968234.

[40] A.-L. Davidson and R.-M. Fountain, "Social Media in Higher Education: A Look at Participatory Culture in Graduate Coursework," *Revue internationale des technologies en pédagogie universitaire*, vol. 10, no. 3, p. 14, 2013, doi: 10.7202/1035576ar.

[41] F. Caboni and L. Pizzichini, "How the COVID-19 Pandemic May Accelerate Millennials' Adoption of Augmented Reality," *International Journal of Retail & Distribution Management*, vol. 50, no. 13, pp. 95–115, Dec. 2022, doi: 10.1108/IJRDM-10-2021-0509.

[42] T. Crick, C. Knight, R. Watermeyer, and J. Goodall, "The International Impact of COVID-19 and "Emergency Remote Teaching" on Computer Science Education Practitioners," in *2021 IEEE Global Engineering Education Conference (EDUCON)*, Vienna, Austria: IEEE, Apr. 2021, pp. 1048–1055. doi: 10.1109/EDUCON46332.2021.9453846.

[43] M. Ehren, R. Madrid, S. Romiti, P. Armstrong, P. Fisher, and D. McWhorter, "Educational Continuity during the Pandemic: Challenges to Pedagogical Management in Segregated Chilean Schools," *Perspectives in Education*, vol. 39, no. 1, 2021, doi: 10.18820/2519593X/pie.v39.i1.4.

[44] C. Huck and J. Zhang, "Effects of the COVID-19 Pandemic on K-12 Education: A Systematic Literature Review.," *New Waves-Educational Research and Development Journal*, vol. 24, no. 1, pp. 53–84, 2021.

[45] E. Niemczyk, Z. de Beer, and H. Steyn, "The Challenges Posed by COVID-19 to the BRICS Education Systems: Lessons to Be Learnt," *Perspectives in Education*, vol. 39, no. 1, 2021, doi: 10.18820/2519593X/pie.v39.i1.11.

[46] F. M. Reimers, Ed., *Primary and Secondary Education During Covid-19: Disruptions to Educational Opportunity During a Pandemic*. Cham: Springer International Publishing, 2022. doi: 10.1007/978-3-030-81500-4.

[47] P. Spathis and R. Dey, "Online Teaching amid COVID-19: The Case of Zoom," in *2021 IEEE Global Engineering Education Conference (EDUCON)*, Vienna, Austria: IEEE, Apr. 2021, pp. 1398–1406. doi: 10.1109/EDUCON46332.2021.9453918.

[48] E.-J. Hoogerwerf, K. Mavrou, and I. Traina, Eds., *The Role of Assistive Technology in Fostering Inclusive Education: Strategies and Tools to Support Change*, 1st ed. Routledge, 2020. doi: 10.4324/9780429428241.

[49] T. Cunningham and B. Lamond, "The Weak Link in Assistive Technology – the Internet," *Computers & Electrical Engineering*, vol. 93, p. 107284, Jul. 2021, doi: 10.1016/j.compeleceng.2021.107284.

[50] M. A. Ayob, N. Hadi, M. Poahroraji, B. Ismail, and M. Saaid, "Promoting 'Discord' as a Platform for Learning Engagement during Covid-19 Pandemic," *Asian Journal of University Education*, vol. 18, no. 3, Jul. 2022, doi: 10.24191/ajue.v18i3.18953.

[51] OpenAI, "GPT-4 Technical Report." Accessed: May 31, 2023. [Online]. Available: http://arxiv.org/abs/2303.08774

[52] E. Kasneci *et al.*, "ChatGPT for Good? On Opportunities and Challenges of Large Language Models for Education," *Learning and Individual Differences*, vol. 103, p. 102274, Apr. 2023, doi: 10.1016/j.lindif.2023.102274.

[53] S. Bubeck *et al.*, "Sparks of Artificial General Intelligence: Early Experiments with GPT-4." Accessed: Mar. 31, 2023. [Online]. Available: http://arxiv.org/abs/2303.12712

[54] G. Cooper, "Examining Science Education in ChatGPT: An Exploratory Study of Generative Artificial Intelligence," *Journal of Science Education and Technology*, Mar. 2023, doi: 10.1007/s10956-023-10039-y.

[55] A. Haleem, M. Javaid, and R. P. Singh, "An Era of ChatGPT as a Significant Futuristic Support Tool: A Study on Features, Abilities, and Challenges," *BenchCouncil Transactions on Benchmarks, Standards and Evaluations*, vol. 2, no. 4, p. 100089, Oct. 2022, doi: 10.1016/j.tbench.2023.100089.

[56] J. V. Pavlik, "Collaborating With ChatGPT: Considering the Implications of Generative Artificial Intelligence for Journalism and Media Education," *Journalism & Mass Communication Educator*, vol. 78, no. 1, pp. 84–93, Mar. 2023, doi: 10.1177/10776958221149577.

[57] S. Khan, "Harnessing GPT-4 so That All Students Benefit. A Nonprofit Approach for Equal Access." Accessed: Mar. 17, 2023. [Online]. Available: https://blog.khanacademy.org/harnessing-ai-so-that-all-students-benefit-a-nonprofit-approach-for-equal-access/

[58] T. Breideband *et al.*, "The Community Builder (CoBi): Helping Students to Develop Better Small Group Collaborative Learning Skills," in *Computer Supported Cooperative Work and Social Computing*, Minneapolis MN USA: ACM, Oct. 2023, pp. 376–380. doi: 10.1145/3584931.3607498.

[59] P. Fairfield, *Education, Dialogue and Hermeneutics*. Continuum International, 2011.

[60] E. Manalo, Ed., *Deeper Learning, Dialogic Learning, and Critical Thinking : Research-Based Strategies for the Classroom*. Routledge, 2020. [Online]. Available: https://doi.org/10.4324/9780429323058

[61] J. Handelsman *et al.*, "Scientific Teaching," *Science*, vol. 304, no. 5670, pp. 521–522, Apr. 2004, doi: 10.1126/science.1096022.

[62] T. A. Kuffner and S. G. Walker, "Why Are p-Values Controversial?," *The American Statistician*, vol. 73, no. 1, pp. 1–3, Jan. 2019, doi: 10.1080/00031305.2016.1277161.

[63] J. W. Mulnix, "Thinking Critically about Critical Thinking," *Educational Philosophy and Theory*, vol. 44, no. 5, pp. 464–479, Jan. 2012, doi: 10.1111/j.1469-5812.2010.00673.x.

[64] A. Cole, *Marjorie Barstow and the Alexander Technique: Critical Thinking in Performing Arts Pedagogy*. Singapore: Springer Nature Singapore, 2022. doi: 10.1007/978-981-16-5256-1.

[65] J. S. Evans, "Dual-Process Theories of Reasoning: Contemporary Issues and Developmental Applications," *Developmental Review*, vol. 31, no. 2–3, pp. 86–102, Sep. 2011, doi: 10.1016/j.dr.2011.07.007.

[66] J. S. Evans, "In Two Minds: Dual-Process Accounts of Reasoning," *Trends in Cognitive Sciences*, vol. 7, no. 10, pp. 454–459, Oct. 2003, doi: 10.1016/j.tics.2003.08.012.

[67] C. C. Williams, M. Kappen, C. D. Hassall, B. Wright, and O. E. Krigolson, "Thinking Theta and Alpha: Mechanisms of Intuitive and Analytical Reasoning," *NeuroImage*, vol. 189, pp. 574–580, Apr. 2019, doi: 10.1016/j.neuroimage.2019.01.048.

[68] D. W. Stacks, Z. Cathy Li, and C. Spaulding, "Media Effects," *International Encyclopedia of the Social & Behavioral Sciences*. Elsevier, pp. 29–34, 2015. doi: 10.1016/B978-0-08-097086-8.95045-1.

[69] C. K. Lo, "What Is the Impact of ChatGPT on Education? A Rapid Review of the Literature," *Education Sciences*, vol. 13, no. 4, p. 410, Apr. 2023, doi: 10.3390/educsci13040410.

[70] P. P. Ray, "ChatGPT: A Comprehensive Review on Background, Applications, Key Challenges, Bias, Ethics, Limitations and Future Scope," *Internet of Things and Cyber-Physical Systems*, vol. 3, pp. 121–154, 2023, doi: 10.1016/j.iotcps.2023.04.003.

[71] A. Tlili *et al.*, "What If the Devil Is My Guardian Angel: ChatGPT as a Case Study of Using Chatbots in Education," *Smart Learning Environments*, vol. 10, no. 1, p. 15, Feb. 2023, doi: 10.1186/s40561-023-00237-x.

[72] M. Bearman, P. Dawson, R. Ajjawi, J. Tai, and D. Boud, Eds., *Re-Imagining University Assessment in a Digital World*, vol. 7. in The Enabling Power of Assessment, vol. 7. Cham: Springer International Publishing, 2020. doi: 10.1007/978-3-030-41956-1.

[73] Ollama, "ollama/ollama: Get up and running with Llama 3, Mistral, Gemma 2, and other large language models." [Online]. Available: https://github.com/ollama/ollama

[74] OpenWebUI, "OpenWebUI." [Online]. Available: https://github.com/open-webui/open-webui

[75] DockerInc., "Docker." [Online]. Available: https://github.com/docker

[76] P. Sullivan, *Qualitative Data Analysis Using a Dialogical Approach*. Sage, 2012. Accessed: Jul. 26, 2023. [Online]. Available: https://methods-sagepub-com.lib-ezproxy.concordia.ca/book/qual-data-analysis-using-a-dialogical-approac

[77] N. Denzin and Y. Lincoln, Eds., *The Sage Handbook of Qualitative Research*. Sage, 2018.

[78] J. Cresswell, *Qualitative Inquiry and Research Design. Choosing among Five Approaches*. Sage, 2013.

[79] M. Zeff and T. Germain, "We Tested AI Censorship: Here's What Chatbots Won't Tell You," *Gizmodo*, Mar. 2024, [Online]. Available: https://gizmodo.com/we-tested-ai-censorship-here-s-what-chatbots-won-t-tel-1851370840

[80] S. Loria *et al.*, "TextBlob: Simplified Text Processing." [Online]. Available: https://textblob.readthedocs.io/en/dev/

[81] C. Hutto and E. Gilbert, "VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text." Ann Arbor, MI, Jun. 2014.

[82] A. Whitaker, "TextBlob Sentiment: Calculating Polarity and Subjectivity." [Online]. Available: https://planspace.org/20150607-textblob_sentiment/

[83] A. Luers *et al.*, "Will AI Accelerate or Delay the Race to Net-Zero Emissions?," *Nature*, vol. 628, no. 8009, pp. 718–720, Apr. 2024, doi: 10.1038/d41586-024-01137-x.

8 Appendices

8.a GPT Development

8.a.a Software

- Ollama (https://ollama.com/)
- Open Source GPTs (https://ollama.com/library)
- Docker (https://docker.com/)
- OpenWebUI (https://openwebui.com/)

8.a.b Hardware

Developed on (May 2024):

- MacBook Pro (13-inch, 2018), OSX Sonoma 14.5
- CPU: Intel Iris Plus Graphics 655 1536 MB, 2.4 GHz Intel Core i5
- Memory: 16 GB 2133 MHz LPDDR3

Tested on (May 2024):

- NixOS 23.11
- CPU: AMD Ryzen 9 5950X (32) @ 3.400GHz
- GPU: NVIDIA GeForce RTX 4090
- Memory: 128GB DDR4

8.a.c Networking

- Digital Ocean (https://digitalocean.com)
  - Droplet: 8GiB Memory, 2 vCPUs, 25GB SSD, 4TB Transfer
  - IP Address: http://64.23.161.74:3000/
- NixOS (https://nixos.org)
- Tailscale (https://tailscale.com)

8.a.d Development Tools

- Warp (https://www.warp.dev)
- Github (https://github.com)
- NVIM (https://neovim.io)
- VSCode (https://code.visualstudio.com)

8.a.e Data Analysis

- R (https://r-project.org)
- Python (https://python.org)

8.b Focus Group Design

Methods

The study will be conducted through group and individual interactions with a GPT. The study will be conducted in the following sequence:

Setting

- Discussions: seminar room
- GPT interactions: personal computer
- Survey: online

Procedures

Task 1 (Pretest Survey, 5 mins)

- Participants will complete a survey to gauge their level of knowledge and comfort with group discussions, climate science, and GPTs (Section 8.e)

Task 2 Discussion 1, 20 mins)

- Participants will be given a worksheet to read and reflect on before the discussion (Section 8.f)
- Participants will be guided through a discussion about a climate science topic as a group
- Participants will be asked to write down their thoughts/reflections throughout the discussion

Task 3 GPT, 15 mins)

- Participants will be introduced to basics of how to interact with the GPT
- Participants will be asked to engage in dialogue with the GPT about the same climate science topic discussed as a group
- Participants will note their thoughts and reflections about their experiences after each session

Task 4 Discussion 2, 20 mins)

- Participants will be guided through a discussion about the same climate science topic as a group, as was done in Task 1

- Participants will submit a short reflection exercise about their perception of the discussion and their understanding of the topic

Task 5 (Posttest Survey, 10 mins)

- Participants will complete a survey to revisit their level of knowledge and comfort with group discussions, climate science, and GPTs. They will also answer open-ended questions about the experience throughout the study (Section 8.e)
- Survey items will address overall perception of the interactions and understanding of the topic and the GPT (Section 8.e)
- Participants will have the option of submitting additional feedback and/or notes on their experiences with the GPT (Section 8.e)

8.c Information and Consent Form

Study Title:

Collective intelligence in the digital age: facilitating dialogic education on climate science with GPTs

Researcher: Faye Xiaoxiao Sun

Researcher's Contact Information: xiaoxiao.sun@mail.concordia.ca

Supervisor: Dr. Damon Matthews

Supervisor's Contact Information: damon.matthews@concordia.ca

You are being invited to participate in the research study mentioned above. This form provides information about what participating would mean. Please read it carefully before deciding if you want to participate or not. If there is anything you do not understand, or if you want more information, please ask the researcher.

A. **PURPOSE**

The research will investigate to what extent can Artificial Intelligence (AI), like Generative pre-trained transformers (GPTs), be used to facilitate or enhance dialogic education (DE) on complex topics like climate science, particularly in university settings.

User interactions in group discussions and with the GPT will provide data on the effects of using GPTs as a teaching aid within a DE context. The research aims to investigate to what extent GPTs can assist in scaffolding the skills necessary to prepare individuals for real-life scenarios that require constructive engagement in conversations on contemporary social, economic, and environmental issues. The results of this study can provide insight on the potential for using a DE approach to GPTs in skills development.

B. **PROCEDURES**

If you participate, you will be asked to:

- Engage in group discussions with other participants about a climate science topic. These will occur at the beginning and end of the study period.
- Interact with a customized GPT about the same topic during discussions.
- Take notes on your thoughts and reflections after discussions and GPT sessions.

- Complete a short survey before beginning the study.
- Complete a survey providing feedback on your experience at the end of the study.

In total, participating in this study will take approximately 1.5 hours of your time.

C. **RISKS AND BENEFITS**

You might face certain risks by participating in this research.

Potential Risks:

- You may experience discomfort during discussions either due to interactions with other participants or because of the topic itself.
- You may experience fatigue due to the duration of the study, which may last up to 1.5 hours.

Potential benefits:

- You will have the opportunity of engaging with other participants in a multimodal learning experience that may enhance future learning experiences through skills building. It is hypothesized that actively practicing skills associated with critical thinking and problem-solving will enhance your ability to engage more deeply and constructively with climate science topics in a group setting.
- You will have the opportunity of testing a customized AI prototype and providing feedback on your experience in order to improve its functionality.
- Your responses may be used to provide guidance on what can be improved with respect to students' learning experiences at Concordia University.

D. **CONFIDENTIALITY**

We will gather the following information as part of this research:

- Group discussions
- Personal reflections (Pretest and Posttest)
- User logs
- Survey responses (Pretest and Posttest)

We will not allow anyone to access the information, except people directly involved in conducting the research. We will only use the information for the purposes of the research described in this form.

The information gathered will be coded. This means that the information will be identified by a unique numeric identifier associated with your direct identifier. Only the primary researcher will be able to identify you using the numeric ID.

We will protect your information by storing it privately offline on a local drive and it will not be shared with parties who are not a part of the research.

We intend to publish the results of the research. However, it will not be possible to identify you in the published results.

We will destroy the information **5 years** after the end of the study.

E. CONDITIONS OF PARTICIPATION

You do not have to participate in this research. It is purely your decision. If you do participate, you can stop at any time. You can also ask that the information you provided not be used, and your choice will be respected.

If you decide that you do not want us to use your information, you must inform the primary researcher in writing **before 2024 August 1st**.

There are no negative consequences for not participating, stopping in the middle, or asking us not to use your information.

Please note that while attempts will be made to withdraw individual data if requested, given the nature of focus group discussions, it will be impossible to withdraw all contributions within a group discussion.

F. GENERAL GUIDELINES FOR CONDUCT

You will be expected to adhere to the Code of Rights and Responsibilities outlined by Concordia University.

By signing this form, you consent to respect each other's confidentiality and to not disclose anyone's identify outside of the group interview. Your identity will be known to other focus

group participants and the researcher cannot guarantee that others in the group will respect your confidentiality.

G. **PARTICIPANT'S DECLARATION**

- I have read and understood this form.
- I have had the chance to ask questions and any questions have been answered.
- I agree to participate in this research under the conditions described.

NAME (please print)

SIGNATURE

DATE

If you have questions about the scientific or scholarly aspects of this research, please contact the primary researcher. Their contact information is on page 1. You may also contact their faculty supervisor.

If you have concerns about ethical issues in this research, please contact the Manager, Research Ethics, Concordia University, 514.848.2424 ex. 7481 or oor.ethics@concordia.ca.

8.d Recruitment

Study Participation Recruitment Email

**This study has been approved by the Concordia University Human Research Ethics Committee.*

Hi everyone,

I hope you're all well!

Gentle reminder to complete the post-workshop survey if you haven't already: https://forms.gle/f8hiiAC9DndhMTTFA

For those of you who expressed interest in including your responses as part of my study, I've attached the consent form in this email. Please try to send them in by the end of June if possible!

Those of you who participate will be updated once the results are published.

Let me know if you have any questions!

All the best, Faye

8.e Survey

8.e.a Pretest Survey

Thank you so much for having participated in our research!

The following survey is meant to provide us with a more comprehensive understanding of your experience throughout the study.

All survey data will be handled as research data, the conditions of which are outlined by the consent form that you have signed in order to participate in this study.

If you have any questions or require clarification, please contact the primary researcher: xiaoxiao.sun@mail.concordia.ca

Part A: Climate Science

1. I would describe my *perception* of climate science as: ☐ positive ☐ mostly positive ☐ neutral ☐ mostly negative ☐ negative
   a. Comments (optional):
2. I would describe my *experiences discussing* climate science with others as: ☐ positive ☐ mostly positive ☐ neutral ☐ mostly negative ☐ negative
   a. Comments (optional):
3. I would describe my *interest* in climate science as: ☐ high ☐ above average ☐ average ☐ below average ☐ low
   a. Comments (optional):
4. I would describe my *knowledge level* on climate science as: ☐ high ☐ above average ☐ average ☐ below average ☐ low
   a. Comments (optional):
5. I would describe my level of *engagement* in climate science discussions with others as: ☐ high ☐ above average ☐ average ☐ below average ☐ low
   a. Comments (optional):
6. I would describe my level of *confidence in discussing* climate science with others as: ☐ high ☐ above average ☐ average ☐ below average ☐ low
   a. Comments (optional):

7. I would describe the level of *comfort I have with discussing opposing views* on climate science as: ☐ high ☐ above average ☐ average ☐ below average ☐ low
   a. Comments (optional):
8. I would describe *my ability to have constructive discussions* about climate science as: ☐ high ☐ above average ☐ average ☐ below average ☐ low
   a. Comments (optional):

Part B: GPTs

1. I would describe my level of *comfort* interacting with a GPT as: ☐ high ☐ above average ☐ average ☐ below average ☐ low
   a. Comments (optional):
2. I would describe my level of *engagement* with GPTs as: ☐ high ☐ above average ☐ average ☐ below average ☐ low
   a. Comments (optional):
3. I would describe my *perception* of GPTs as: ☐ positive ☐ mostly positive ☐ neutral ☐ mostly negative ☐ negative
   a. Comments (optional):
4. I would describe my *experience* with using GPTs as: ☐ positive ☐ mostly positive ☐ neutral ☐ mostly negative ☐ negative
   a. Comments (optional):
5. I would describe my *usage* of GPTs for mostly: ☐ personal ☐ educational ☐ work ☐ other (please describe in comments)
   a. Comments (optional):

Due Date: **2024 August 1st**

*__Identifier Code__: [numeric identifier]

8.e.b Posttest Survey

Thank you so much for having participated in our research!

The following survey is meant to provide us with a more comprehensive understanding of your experience throughout the study.

All survey data will be handled as research data, the conditions of which are outlined by the consent form that you have signed in order to participate in this study.

If you have any questions or require clarification, please contact the primary researcher: xiaoxiao.sun@mail.concordia.ca

Due Date: **2024 August 1st**

*__Identifier Code__: [numeric identifier]

Part A: Climate Science

1. I would describe my *perception* of climate science as: ☐ positive ☐ mostly positive ☐ neutral ☐ mostly negative ☐ negative
   a. Comments (optional):
2. I would describe my *experiences discussing* climate science with others as: ☐ positive ☐ mostly positive ☐ neutral ☐ mostly negative ☐ negative
   a. Comments (optional):
3. I would describe my *interest* in climate science as: ☐ positive ☐ mostly positive ☐ neutral ☐ mostly negative ☐ negative
   a. Comments (optional):
4. I would describe my *knowledge level* on climate science as:
   ☐ high ☐ above average ☐ average ☐ below average ☐ low
   a. Comments (optional):
5. I would describe my level of *engagement* in climate science discussions with others as: ☐ high ☐ above average ☐ average ☐ below average ☐ low
   a. Comments (optional):
6. I would describe my level of *confidence in discussing* climate science with others as: ☐ high ☐ above average ☐ average ☐ below average ☐ low
   a. Comments (optional):
7. I would describe the level of *comfort I have with discussing opposing views* on climate science as: ☐ high ☐ above average ☐ average ☐ below average ☐ low
   a. Comments (optional):

8. I would describe *my ability to have constructive discussions* about climate science as: ☐ high ☐ above average ☐ average ☐ below average ☐ low
    a. Comments (optional):

Part B: GPTs

1. I would describe my level of *comfort* interacting with a GPT as: ☐ high ☐ above average ☐ average ☐ below average ☐ low
    a. Comments (optional):
2. I would describe my level of *engagement* with GPTs as: ☐ high ☐ above average ☐ average ☐ below average ☐ low
    a. Comments (optional):
3. I would describe my *perception* of GPTs as: ☐ positive ☐ mostly positive ☐ neutral ☐ mostly negative ☐ negative
    a. Comments (optional):
4. I would describe my *experience* with using GPTs as: ☐ positive ☐ mostly positive ☐ neutral ☐ mostly negative ☐ negative
    a. Comments (optional):
5. I would describe my *usage* of GPTs for mostly: ☐ personal ☐ educational ☐ work ☐ other (please describe in comments)
    a. Comments (optional):

Part C: Reflections

1. Generally speaking, what do you find to be the *most difficult aspects of group discussions* on climate science?
    a. Has this affected your engagement in group discussions? If so, how?
    b. Has this affected your level of involvement in environmental issues? If so, how?
2. During first group discussion, were there any points where you *felt discomfort* regarding the topics discussed? If so, please describe:
    a. when they occurred,
    b. why they occurred,
    c. if they were resolved, and how.

3. Did any of the same issues from (1) and (2) above occur during the *final group discussion*? If so, please describe:
    a. when they occurred,
    b. why they occurred,
    c. if they were resolved, and how.
4. Did any of the same issues above occur when *interacting with the GPT*? If so, please describe:
    a. when they occurred,
    b. why they occurred,
    c. if they were resolved, and how.
5. Was there a difference in how you chose to approach the group discussions after interacting with the GPT?
    a. If yes, how and why?
    b. If no, how and why?
6. Do you feel that the dynamics of the group discussion changed after interacting with the GPT?
    a. If yes, how and why?
    b. If no, how and why?

Part D: Feedback (optional)

1. When using the GPT, did you notice any interesting bugs or errors that were produced?
    a. If so, please describe how they can be reproduced in steps.
    b. Please also describe any other technical issues you encountered that may impact its functionality.
2. Please feel free to share any additional feedback on your experience with the study so that we can improve upon future studies.

Participation in Study:

Would you be interested including your responses as part of a study on AI and climate education/communication? If so, you will be sent a consent form to sign at a later date.

☐ Yes

☐ No

☐ Maybe (please elaborate)

8.f Discussion Prompt

The following worksheet was used on 17th May 2024 to facilitate a discussion:

"Could AI take the world to a more radically green future, or a more dystopian one?"

Potential Pros of AI for Sustainability

- **Optimize supply chains:** AI can make manufacturing more efficient and support the integration of renewable energy into electricity grids.
- **Speed up development of new materials**: AI can accelerate the development of new batteries and renewable energy technologies.
- **Tackle climate change**: Some AI applications are designed specifically to reduce emissions from sectors like energy, transport, buildings, and industry.

Potential Cons of AI for Sustainability

- **Increase electricity use**: Embedding AI into existing applications like healthcare and entertainment could drive more electricity consumption.
- **Cheaper oil and gas extraction**: AI could make fossil fuel exploration and extraction cheaper, potentially increasing production.
- **Affect political and economic stability**: Widespread AI use without proper governance could negatively impact poverty, food security, and social inequalities, with downstream effects on emissions.
- **Uncertain indirect effects**: The broader societal and economic impacts of AI expansion are not well understood and could have both positive and negative implications for emissions.

The indirect effects of AI could exceed the direct impacts on energy use and emissions, making the overall environmental consequences highly uncertain.

Researchers are called to develop a set of policy-relevant scenarios to quantify the effects of AI expansion on the climate under different assumptions.

5 Key Elements for AI-driven Emissions Scenarios

1. **Link to existing climate scenarios**: Integrate AI into the standard Shared Socioeconomic Pathways used to assess future emissions.
2. **Develop quantitative analytical frameworks**: Improve data and models to quantify both direct and indirect AI impacts.
3. **Share data**: Establish standards for measuring, reporting and sharing AI-related energy and emissions data.

4. **Issue rapid updates**: Update the scenarios at least annually to keep pace with fast-moving AI developments.
5. **Build an international consortium**: Establish a global consortium to undertake this scenario development work.

These scenarios can guide policymakers, investors and others in making decisions about AI development and use with sustainability in mind.

A. Luers et al., "Will AI accelerate or delay the race to net-zero emissions?," Nature, vol. 628, no. 8009, pp. 718–720, Apr. 2024, doi: 10.1038/d41586-024-01137-x.

Discussion Worksheet

During the discussion, write up to **3** thoughts that you would like to explore more in depth afterward.

1.

2.

3.

8.g Linear Regression Analysis

Residual and QQ plots for the Textblob results showed linearity, homoscedasticity, and normality. The linear regression line shows a positive correlation between Polarity and Subjectivity, with outliers for Participants 4, 8 and 11. The results of the linear regression analysis are shown in the scatter plot below.

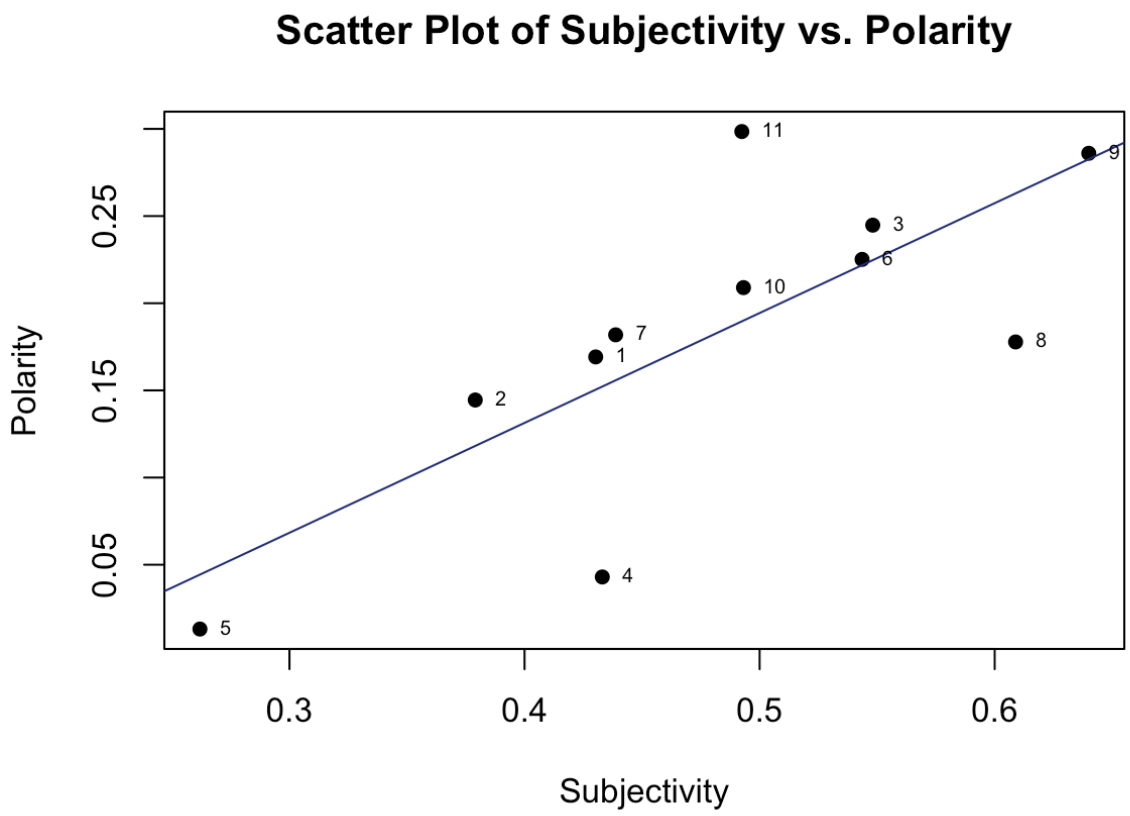

Figure 19: Scatter plot of Polarity and Subjectivity of the reflections from TextBlob sentiment analysis. Polarity is measured on a scale of −1 (Negative) to +1 (Positive) and Subjectivity is measured on a scale of 0 (Objective) to 1 (Subjective). The linear regression line shows a positive correlation between Polarity and Subjectivity, with outliers for Participants 4, 8 and 11.

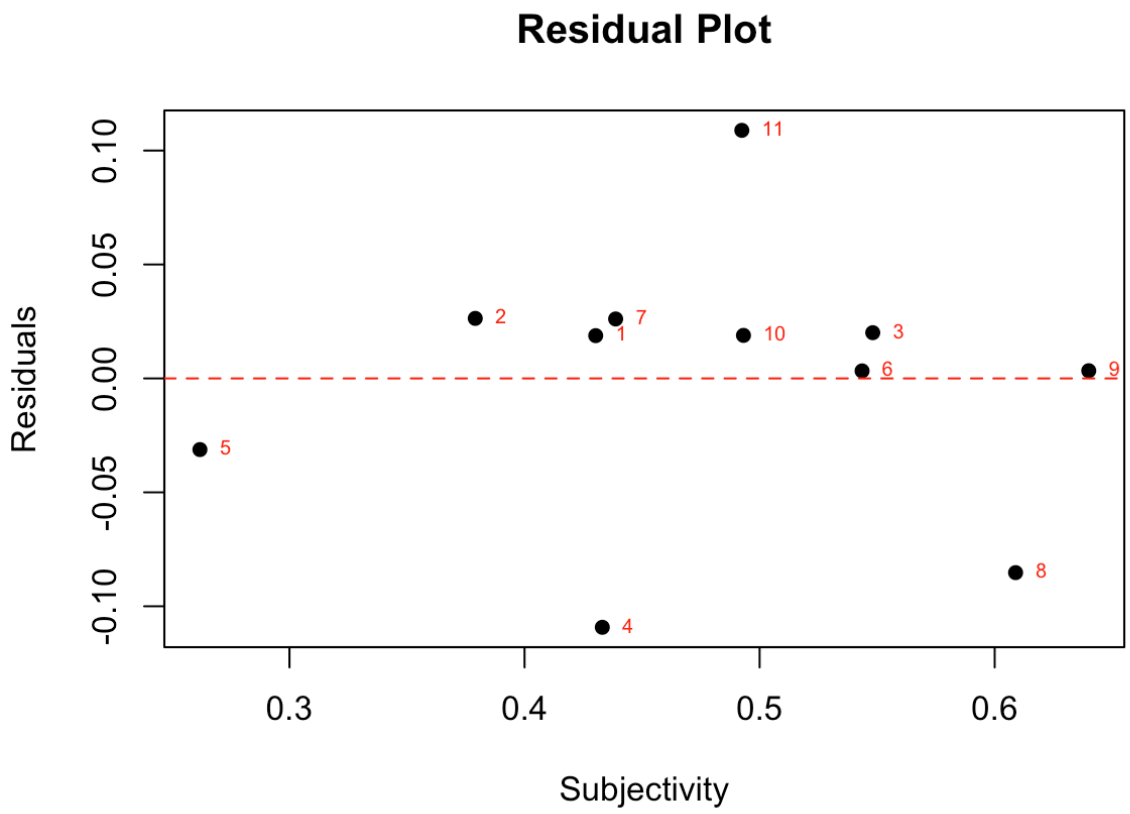

Figure 20: Residual plot of the linear regression model for Polarity and Subjectivity show most plots sitting close to the line where y = 0, with points for Participants 4, 8, and 11 as outliers. This points to homoscedasticity in the data.

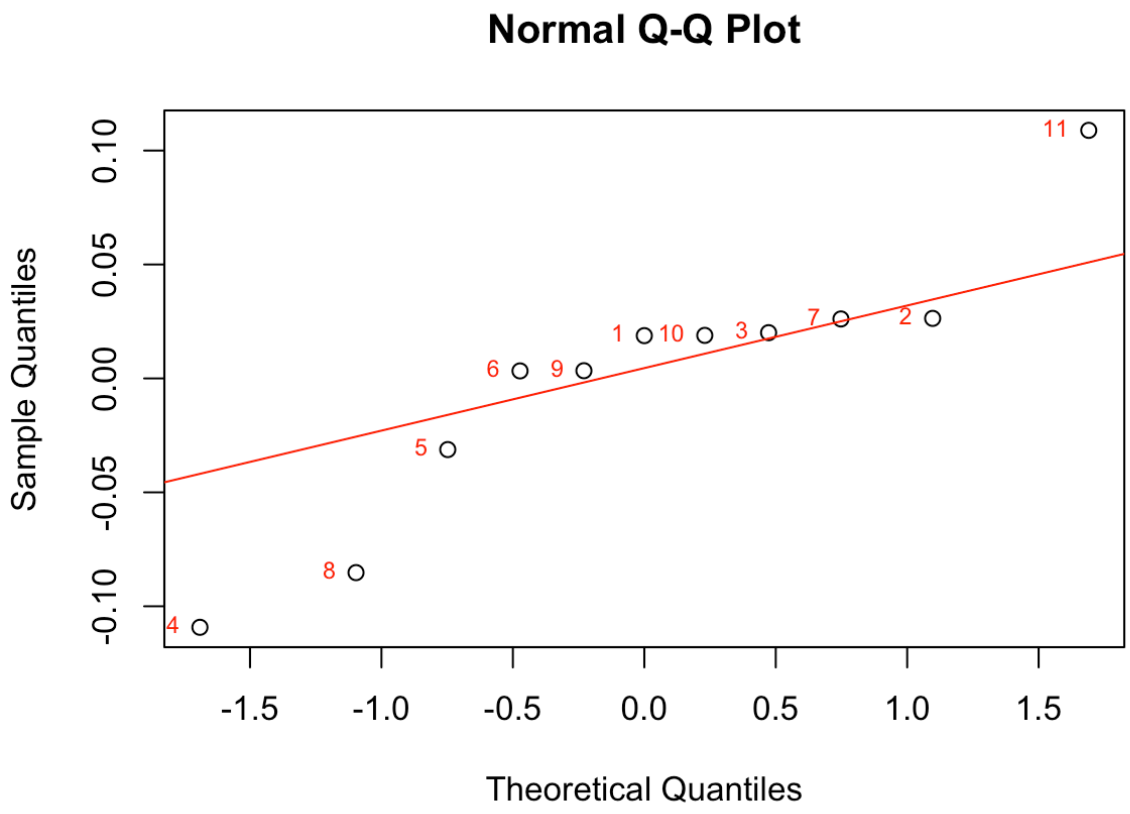

Figure 21: QQ plot of the residuals from the linear regression model for Polarity and Subjectivity shows that the majority of points sit close to the QQ line. The data is normally distributed, save for outliers for Participants 4, 8, and 11 at the tails.

The summary output from R for regression model shows a p-value of 0.007211, which is less than the alpha level of 0.05. This indicates that there is a significant correlation between Polarity and Subjectivity in the reflection exercises. The R-squared value of 0.5702 indicates that 57.02% of the variance in Polarity can be explained by Subjectivity. This suggests that the linear regression model has a moderate level of explanatory power, meaning that Subjectivity is a significant predictor of Polarity in the reflection exercises.

Summary output from R for the linear regression model:

```
Residuals:
     Min       1Q   Median       3Q      Max
-0.10923 -0.01397  0.01877  0.02308  0.10887

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  -0.12073    0.08934  -1.351  0.20955
```

```
Subjectivity  0.63024    0.18239   3.456  0.00721 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.06194 on 9 degrees of freedom
Multiple R-squared:  0.5702,    Adjusted R-squared:  0.5225
F-statistic: 11.94 on 1 and 9 DF,  p-value: 0.007211
```

8.h Participant Resources

Additional student services and resources can be found through the **Student Hub**.

1. **Access Centre for Students with Disabilities**
2. **Student Success Centre**
3. **Counselling and Psychological Services**
4. **Health Services**
5. **Dean of Students Office**
6. **International Students Office**
7. **Sexual Assault Resource Centre**
8. **Otsenhákta Student Centre**
9. **Student Advocacy Office**
10. **Concordia Student Union** & **Graduate Student Association**