[Inner endpaper]

HEC MONTRÉAL

The Interplay of Chatbot Language Style and Perceived User Experience on Group Decision Making Performance

Mina Takhsha

Directors

Wietske Van Osch

HEC Montreal

Constantinos Coursaris

HEC Montreal

Master’s of Science (User Experience)

Thesis submitted in fulfillment of the requirements for a Master’s of Science (M. Sc.) in User Experience

August 2024

Résumé

Dans les organisations modernes, la collaboration à distance est devenue essentielle, et les avancées technologiques ont rendu les agents d'IA de plus en plus importants. Cette étude examine l'impact du style de langage des chatbots sur la performance de la prise de décision en groupe. En utilisant un plan expérimental inter-sujets, 189 participants répartis en 61 équipes ont été assignés aléatoirement à l'une des quatre conditions: idéateur de type humain (n=48), idéateur de type robot (n=46), facilitateur de type humain (n=44), ou facilitateur de type robot (n=51). Les participants ont réalisé des tâches de prise de décision sur la plateforme Chatzy avec un chatbot nommé "Ideabot", opéré par un humain selon la technique du "Magicien d'Oz". Nous avons mesuré l'efficacité de la prise de décision, la précision, et la satisfaction du processus, ainsi que l'effet médiateur de la confiance interpersonnelle et de la satisfaction du système.

Les résultats montrent que le style de communication du chatbot a un impact significatif sur les résultats des décisions d'équipe. Les équipes avec un chatbot de type robot ont atteint une efficacité supérieure, tandis que celles avec un chatbot de type humain ont rapporté une plus grande satisfaction du processus. Le style de langage n'a pas influencé la précision des décisions, mais le rôle du chatbot a été significatif, les équipes avec un chatbot idéateur obtenant une meilleure précision. Ces résultats soulignent l'importance de l'affectation stratégique des styles et rôles de communication des chatbots pour améliorer la prise de décision en équipe.

Mots clés: Chatbot, Intelligence Artificielle, Prise de Décision en Groupe, Équipes à Distance, Style de Communication, Efficacité de la Décision, Exactitude de la Décision, Satisfaction du Processus, Confiance interpersonnelle, satisfaction du système.

Méthodes de recherche: Quantitative, explicative, exploratoire

Abstract

In today's organizations, remote collaboration is essential, and advancements in technology have made AI agents crucial for supporting collaboration. This study explores the effects of language style of chatbots on group decision-making performance. We used a between-subjects design with 189 participants, across 61 teams, who were randomly assigned to one of four conditions: human-like ideator (n=48), robot-like ideator (n=46), human-like facilitator (n=44), and robot-like facilitator (n=51). Participants completed decision-making tasks with a chatbot named "Ideabot," using the "Wizard of Oz" technique. We measured decision-making efficiency, accuracy, and process satisfaction, as well as the mediating effect of interpersonal trust and system satisfaction.

Results revealed that the chatbot's communication style had a significant impact on team decision outcomes. Teams with a robot-like chatbot achieved higher efficiency, while those with a human-like chatbot reported greater process satisfaction. Although language style did not influence decision accuracy, the chatbot's role—assessed through a post-hoc analysis—was significant. Teams with an ideator chatbot, which provided its own suggestions, had higher decision accuracy compared to those with a facilitator chatbot, which merely orchestrated the process. These findings highlight how strategic assignment of chatbot communication styles and roles can enhance decision-making in diverse situations requiring distinct outcomes, highlighting a possible trade-off between decision-making efficiency and accuracy. Additionally, the data did not reveal a mediating effect of interpersonal trust or system satisfaction on the relationship between language style and team decision-making performance, underscoring the significance of the direct impact of chatbot characteristics on team performance in remote decision-making contexts.

Keywords: Chatbot, Artificial Intelligence, Group Decision-making, Remote Teams, Communication Style, Decision Efficiency, Decision Accuracy, Process Satisfaction, Interpersonal Trust, System Satisfaction.

Research methods: Quantitative, explanatory, exploratory

Table of contents

List of tables

List of figures

List of abbreviations

AI = Artificial Intelligence

AVE = Average Variance Extracted

CAD = Canadian Dollars

CB-SEM = Covariance-Based Structural Equation Modeling

CCS = Chatbot Conversational Style

CMC = Computer-Mediated Communication

CR = Composite Reliability

GDM = Group Decision-Making

GDMA = Group Decision-Making Accuracy

GDME = Group Decision-Making Efficiency

GDMPS = Group Decision-Making Process Satisfaction

GDSS = Group Decision Support-System

HCI = Human-Computer Interaction

IS = Information Systems

IT = Interpersonal Trust

MNVT = Multinational Virtual Team

PLS = Partial Least Squares

PLS-SEM = Partial Least Squares Structural Equation Modeling

REB = Research Ethics Board

SDM = Structured Decision-Making

SEM = Structural Equation Modeling

SS = System Satisfaction

STDEV = Standard Deviation

UK = United Kingdom

USA = United States

UX = User Experience

VIF = Variance Inflation Factor

This research is lovingly dedicated to my sister Maryam, who has been my constant source of inspiration. Her belief in me has given me the drive and discipline to tackle any difficulties that came my way. Without her love, guidance, and endless support, this project would not have been possible.

Acknowledgements

I would like to express my sincere thanks to my supervisors, Wietske Van Osch and Constantinos Coursaris, for their invaluable guidance, feedback, and encouragement throughout this journey.

I am deeply grateful to my family and friends for their support, with special appreciation to my dear husband Mehran, who stood by me during difficult times and provided support and help with the final editing.

Chapter 1: Introduction

1.1 Research Context and Motivation

In recent years, AI systems and intelligent agents have become increasingly prevalent across numerous domains, providing invaluable support for decision-making (Reverberi et al., 2022). Sectors in which AI is utilized to support decision-making range from business and market decisions to military command and even medical diagnosis and treatment (Yan and Gurkan, 2023). The proliferation of AI usage has prompted a surge of interest in human-AI collaboration within the information systems field and many researchers have begun to investigate the challenges humans encounter when collaborating with AI systems, exploring how humans perceive their AI counterparts, as well as the tools that either facilitate or impede their relationship and performance outcomes (Munyaka et al., 2023).

Coinciding with this gap is the restructuring of workplaces and organizations around teams, underscoring the increasing relevance of AI-based intelligent agents to support tasks conducted in group settings (Zhu et al., 2021). Teams are highly prevalent in today's organizations because of the multitude of benefits they offer, including improvements in employee relations, enhancement of technical and interpersonal skills, enhancement of quality of work life, increased job satisfaction and performance, growth in organizational effectiveness, and enhanced flexibility (Khawam et al. 2017). With the rise of digital tools for collaboration, teams have become increasingly remote, and many decision-making tasks are now conducted in online teams comprised of distributed members.

As aforementioned, AI systems have proven to be invaluable tools for enhancing decision-making (Reverberi et al. 2022) thus, teamwork between humans and artificial intelligence will foster a "hybrid intelligence," that has the potential to produce outcomes superior to those achieved by either human or AI intelligence operating independently (Reverberi et al. 2022). Despite the importance of the subject, limited evidence-based knowledge exists regarding the design elements that facilitate optimal human-AI collaboration in teams, and particularly, in remote team settings (Reverberi et al. 2022).

Additionally, there's been a growing focus among researchers on a particular design aspect: the level of interaction with natural language and the degree of anthropomorphism in AI agents (Seeger et al., 2018). Human-likeness in these agents has been linked to increased user desirability and positive effects (Song and Shin, 2024). The Modality, Agency, Interactivity, Navigation (MAIN) model further highlights human likeness as a crucial factor in enhancing the effectiveness of human-chatbot interactions (Song and Shin, 2024). On the contrary, the Uncanny Valley theory suggests that when something closely resembles a human but falls short, it can have the opposite effect by evoking feelings of unease or repulsion (Ciechanowski et al., 2019). Hence, existing research has been equivocal about the nature of the effect of greater humanness of chatbots in human-AI collaboration.

Furthermore, previous research in this field has primarily focused on the interaction between a single individual and an AI agent, such as voice assistants like Siri (Hsu and Lee, 2023) or a chatbot (Ciechanowski et al., 2019). However, there is a notable lack of studies examining how the human-like qualities of chatbot language impact performance and interactions within a team context. In such scenarios, where humans interact with chatbots in the presence of other human members, the dynamics of these interactions become more intricate and potentially distinct from traditional one-on-one interactions.

Additionally, while chatbot research has predominantly focused on customer service, the decision-making context, especially a collaborative decision-making context, has been much less explored. Understanding how chatbots influence decision-making processes is crucial, as AI agents are increasingly being integrated into both social and professional settings. This study aims to address these gaps by investigating the nuanced dynamics of human-AI collaboration in remote team settings and their implications for collaborative decision-making performance. These gaps in literature, along with the increasing significance of AI agents in organizational and social contexts, serve as the primary motivation for this research.

1.2 Research Objectives and Research Questions

The primary aim of this thesis is thus to explore the impact of chatbot language human-likeness on decision-making performance within remote teams, assessing whether teams exhibit improved decision-making performances when interacting with chatbots using natural, human-like language as opposed to more robotic language. Therefore, the research aims to address the following overarching research question: *What is the effect of the communication style of chatbots on the decision-making performance of remote /distributed work teams?*

To tackle this inquiry, the study adopts an explanatory approach aimed at assessing the effects of chatbot language style on different dimensions of group decision-making performance such as the unobtrusive dimensions of accuracy and efficiency as well as the perceived satisfaction with the decision-making process.

A secondary and inherently exploratory objective of this study is to examine the effect of the chatbot's role on the decision-making performance of the group. Specifically, this research aims to compare the decision performance of groups when the chatbot functions as an ideator, actively proposing ideas and/or making suggestions, versus when it acts as a facilitator, merely guiding and instructing team members through the different phases of the decision-making process. The secondary research question posed by this exploratory investigation is: *What is the effect of the chatbot's role on the decision-making performance of remote/distributed work teams?*

1.3 Potential Research Contributions

This study contributes to bridging the gap in the extant literature concerning the impact of chatbot linguistic anthropomorphism on users and their collaborative performance within remote team settings, contrary to the dominant focus in the existing literature on individual interactions with chatbots. As aforementioned, considering that today's organizations are increasingly embracing AI tools to support work processes and these processes are increasingly structured in teams, this is a context that warrants further investigation. Furthermore, this research emphasizes decision-making contexts as opposed to typical customer service settings. Decision-making is an intricate part of today's organizing and thus further underscores the relevance of this research context. Finally, the remote nature of team interaction, the key focus of this study, parallels dominant work arrangements in a post-pandemic workplace, thus shedding light on a common mode of collaboration.

By examining the effect of diverse linguistic designs on decision-making outcomes within remote team settings involving multiple team members and an AI-based intelligent agent, this study has the potential to significantly enhance the design and development of intelligent systems within organizational contexts.

Additionally, our exploratory analysis of the effect of chatbot roles on decision-making performance contributes to the existing literature on chatbot and AI usage in organizational decision-making contexts. With the advancement of generative AI and large language models, future AI tools and chatbots are expected to become more proactive, offering suggestions and input to teams. This shift highlights the importance of investigating chatbot roles, as understanding their impact could shape how these tools are proactively integrated into workplace decision-making processes.

1.4 Personal Contributions to the Research

This research is done as a part of a larger project on human-AI collaboration in remote team settings, where my focus was on decision-making as the key dependent variable of interest and the other part of the project, conducted by a fellow M.Sc. student, centered on collaborative creativity.

Table 1. Student's contributions in thesis realization

Steps in the Process	Contribution
Defining the Research Question	Identifying the gaps in the literature to define the main research problem – 60%. Defining the research project's general directions and the research objectives – 60%. My thesis co-supervisors guided me in the process of choosing the general subject.
Theoretical Background	Conducting in-depth research on scientific articles related to the topic – 90% Identifying the conceptual frameworks to be used in the study – 90% My co-supervisors continuously provided feedback and guidance, enabling me to identify the foundational theories for my research model. Synthesizing the relevant literature and concepts for writing the articles – 90%
Experiment	Designing the procedure and tasks – 60% Designing the chatbot scripts – 50%
Ethics	Preparing documentation related to application submission to the REB – 80% Collaborating with another student (Nimmi Luckheenarain) on this project, we collectively prepared the necessary documents.

Recruitment	Recruitment of participants – 30% The participants for this research have been recruited through Prolific platform, and they booked their syncrounus sessions with 2 other anonymous team members on Calendly. My colleague Nimmi Luckheenarain in the research team who was working on other aspects of the same project created accounts and initiated the recruitment process and I would assist her in the recruitment process. Managing participants compensations –90%
Data Collection	Conducting the experiment sessions and data collection_ 70% I had a teammate who was working on other aspects of the same project, and we collaboratively collected the data.
Data Analysis	Formatting –80% Analyzing –50% My supervisor and I had several online and in person sessions during data analysis phase and she guided me and helped me to analyze and interpret the collected data.
Writing the Thesis	Writing my thesis – 75% My thesis supervisor provided detailed feedback throughout the entire process, enabling me to make necessary adjustments to enhance the overall quality and coherence of my thesis.

1.5 Thesis Structure

The thesis is structured as six chapters. The first chapter sets the stage by outlining the thesis's motivations and providing a succinct overview of the contextual framework. Chapter 2 rigorously examines the theoretical foundations, focusing on existing literature related to group decision-making within remote teams. It also delves into studies on the effects of chatbot usage and its language style on online decision-making performance, evaluating their impact on user perceptions and attitudes. Chapter 3 presents the methodology adopted in this thesis, detailing research methods, data collection procedures, sample selection criteria, and the metrics used for each research variable. Chapter 4 advances into data analysis, presenting the study's results and interpreting the data to provide meaningful insights. Chapter 5 discusses the obtained results, while the final chapter, Chapter 6, encapsulates the research conclusions, discussing their implications for both academic research and practice. It also critically evaluates the study's limitations and suggests potential directions for future research.

Chapter 2: Theoretical Background

2.1 Collaborative Decision-Making

In recent decades, organizational decisions have become more complicated. Consequently, collaborative decision-making has become a prominent mode for obtaining high-quality decisions in organizations (Wang et al., 2021). The application of collaborative decision-making processes in organizations can be explained by Deutsch's theory of cooperation and competition. Based on this theory, collaboration will occur when individuals perceive that the attainment of one's goals also facilitates the achievement of others' goals, leading to a symbiotic relationship wherein each person's effectiveness in pursuing their objectives contributes to the collective success of the group in reaching their respective goals (Alper et al., 1998).

A group decision-making process can be defined as "*a decision situation in which there are two or more individuals who differ in their preferences (value systems), but have the same access to information, each of them characterized by his or her own perceptions, attitudes, motivations, and personalities; who recognize the existence of a common problem; and who attempt to reach a collective or joint decision*" (Herrera et al., 1995, p. 223). In other words, Group Decision-Making (GDM) is a collaborative process where a group of individuals comes together to address a task or problem that requires a collective decision-making effort (Delic et al., 2023). To fulfill such tasks, group members need to share their proposals, discuss them and select one of the alternatives as the final decision to be implemented (Pérez-Soler et al., 2018).

Generally, organizational decisions fall into three categories, namely structured, semi-structured and unstructured (Duan et al., 2019). Structured decision-making (SDM), is a decision-making style in which there is a formal and standardized procedure that will guide decision-makers and define the criteria that they must use in their decisions (Shook and Sarri, 2007). Conversely, in unstructured decision-making, there is no algorithmic and predefined decision procedure and no unique correct alternative (Iselin, 1988). Finally, semi-structured decision processes encompass both well-defined structured components, such as formal procedures and traditional algorithms, and ill-defined unstructured elements, like intuition, judgment, and neural networks, which interact in a known manner (Kaliardos, 1999).

In a decision-making process, the decision-makers must assess a set of available alternatives using a typically conflicting set of criteria, with the primary goal of ranking and selecting the alternatives based on the information at hand (Zakeri et al, 2023).

2.2 Decision-Making Models

The table below summarizes various decision-making models proposed in the existing literature, outlining the steps decision-makers follow to reach their final choice. This study uses these models to map out the phases of decision-making and design tasks for measuring group decision-making performance.

Table 2. Decision-making Models

Author	Model
Bales and Strodtbeck, 1951	**Unitary sequence model** 1. A period of orientation 2. Evaluation 3. Control In this model, decision development involves sequential phases that groups progress through to reach a final choice. Each phase represents a distinct period of focused activity, serving a specific function in the decision-making process.
Fisher, 1970	**Four-phase unitary sequence model** 1. Orientation: Members familiarize themselves with the group dynamics and atmosphere. 2. Conflict: Disagreements arise, and members resist unfavorable comments and ideas. 3. Emergence: Conflicts diminish, and decisions start to take shape, though attitudes may still be ambiguous.

	4. Reinforcement: Favorable attitudes are reinforced, leading to a consistent pattern of positive feedback and solidifying the final decision.
Eden, 1982	**Four-phase decision-making model** 1. Understanding the problem: Considering the issue's scope, discussing conflicting beliefs, and identifying uncertainties. 2. Defining the problem: Achieving a mutual understanding of the problem. 3. Finding solutions: Exploring and identifying a range of potential solutions. 4. Constructing a new problem: Developing a revised problem representation to reflect parties' acceptance and understanding (Franco and Rouwette, 2011).
Poole, 1981	**Multiple sequence model** This model suggests that different groups can experience diverse developmental sequences due to contingency variables. The occurrence, order, and number of developmental stages vary between groups.
Mackenzie, 1976	**Critical events model** This model posits that decision development involves a series of milestones that a group must achieve to complete a task. Each milestone is linked to specific behaviors and messages, which help identify the level of group's progress (Poole and Baldwin, 1996).
Scheidel and Crowell, 1964	**Spiraling model** In this model, groups progress toward a decision through a "reach-test" motion. A group member introduces an idea, which is elaborated on and approved by others before moving to a new idea. This cyclical process of introduction, discussion, and anchoring continuously builds group consensus.
Poole and Doelger, 1986	**Structuration model** This model posits that group members use rules and resources to guide their activities and form the pathway to a decision. These structures are continuously created and replicated through their application in decision-making (Poole and Baldwin, 1996).
Dewey, 1933	**Reflective thinking model** 1. Awareness of the problem, 2. Assessment of the problem, 3. Suggesting solutions, 4. Assessing solutions, 5. Testing solutions.
Wallas, 1926	**creative problem-solving model** 1. Preparation: Exploring various directions using logic and reasoning. 2. Incubation: Shifting attention away from the problem, potentially leading to better solutions.

	3. Illumination: In this phase the problem and its solution spontaneously emerge in conscious thought. 4. Verification: Confirming the accuracy of the insightful solution (He´lie and Son, 2010).
Rawlinson, 1981	**creative problem-solving model** 1. Preparation: Gathering relevant facts and restating the problem for foundational understanding. 2. Effort: Actively overcoming mental blocks hindering progress. 3. Incubation: Allowing the problem to rest in the subconscious while focusing on unrelated activities. 4. Insight: The pivotal 'Aha' moment when novel ideas emerge. 5. Evaluation: Assessing and validating the proposed solution (Poole and Baldwin, 1996).
Delbecq and Van de Ven, 1971	**Program planning model** 1. Problem exploration, 2. Knowledge exploration, 3. Priority development, 4. Program development, 5. Program evaluation.
Chakravarthy and Lorange, 1999	**strategic planning model** 1. Objective setting, 2. Strategic programming, 3. Budgeting, 4. Monitoring, 5. Control and learning. 6. Incentives and staffing (Muller, 2010).
Simon, 1977	**Four-phase model** 1. Intelligence: Defining the problem during decision-making, involving finding, identifying, and formulating the problem. 2. Design: Creating and assessing various solution models. 3. Choice: The pivotal phase, marking the actual decision-making moment. 4. Implementation: Operationalizing the selected solution, putting it into practice. 5. Monitoring: Ensuring the chosen solution's performance is tracked (Liberatore and Wagner, 2022).

These models can fall into different categories namely: Phase models (i.e. Unitary sequence model), Critical events model (i.e. Mackenzie’s Critical events model), Continuous models (i.e. Spiraling model), and Social construction models (i.e. Structuration model) (Poole and Baldwin, 1996). Most of the decision-making models contain between 3 to 5 steps, and although they might differ in their details, steps are similar to each other in terms of general aspects. The majority of models consider the decision-making process as a cycle which starts from identifying and defining the problem and continues by periods of discussion on the probable solutions and will end when the group reaches a consensus on the best alternative. However, some models—like Simon's model and Dewey's reflective thinking model—would consider the implementation of the chosen alternative as the final step of the process. The early models like Bales and Strodtbeck's model would consider decision-making as a linear process, however, later researchers argued that the decision-making process is more complex and inherently iterative where decision makers cycle back to earlier steps several times. More unique models among others are the creative models as these models consider a period of incubation, in which the decision maker would stop thinking about the decision that has to be made, as a part of the process.

Among the above-mentioned models, a combination of the unitary sequence model formulated by Bales and Strodtbeck arguing that groups go through three phases when they move toward their goal (Gersick, 1988) and the multiple sequence model elaborated by Poole (1981), which is against the fixed order of the phases in the unitary model, can explain the group decision-making process in the context of this study. Inspired by these models this study considers the decision development process as a three-stage process of Definition, Ideation, and Selection. Due to the purpose of this study, which is to only consider the stages of decision-making from the start of the process—i.e., determining what the problem is—to the discussion of alternatives and choosing the single best alternative as the final decision, this model does not involve the implementation stage. Moreover, to make the remote online decision tasks more plausible, the design of the task will be semi-structured as participants will be guided through the steps they need to take, yet, they need to rely on their own creativity to generate decision options and set their own decision criteria to derive at their final choice. Because of the experimental and predefined stages with pre-defined time limits, the more structured decision-making models selected for this study are well-suited.

Furthermore, the short duration of the task excludes the need for incubation or implementation stages, which are more relevant to extended, complex organizational decision-making processes. The focus on Definition, Ideation, and Selection aligns with the core phases present in nearly all decision-making models, ensuring relevance and applicability to the study's objectives (Bales & Strodtbeck, 1951; Poole & Baldwin, 1996).

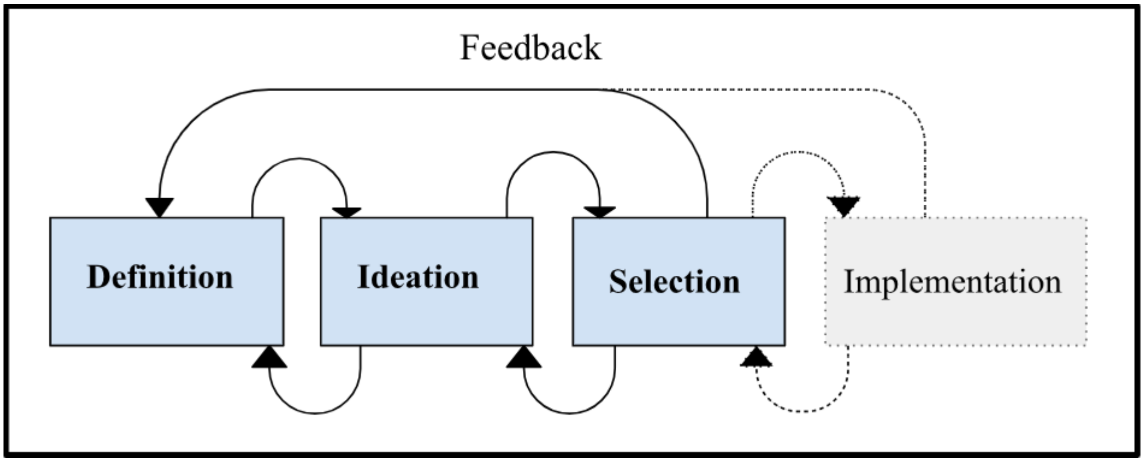

Figure 1. Decision-making Process as Conceptualized in this Study

2.3 Collaborative Decision-Making using Conversational Agents

In today's globalized world, online decision-making has become prevalent due to globalization, employee mobility, and the increasing need for collective and swift decisions among geographically dispersed team members (Turban et al., 2011). Organizations have transitioned from face-to-face meetings to virtual environments, utilizing digital communication technologies and tools to facilitate seamless coordination and information sharing among members regardless of their location (Pérez et al., 2011). This shift aims to create optimal work conditions and enhance informed decision-making (Pérez et al., 2011). Additionally, the adoption of telecommunication technologies has been shown to improve team productivity and satisfaction (Pérez et al., 2011).

The intelligence of AI technologies is swiftly advancing, and they are increasingly serving as assistants for decision-makers in complex and diverse contexts (Jarrahi, 2018). The main domains in which AI assistants are used for decision-making involve business and finance (Cao, 2022, Schemmer et al., 2022), law and civic (Reiling, 2020, Blank and Osofsky, 2020), medical and healthcare (Reverberi, 2022, Haick and Tang, 2021), education (Hoti et al., 2023 and Sekeroglu et al., 2019), professional and career (Esch et al, 2019), and entertainment (Tanti et al., 2023). Based on the structure of decisions (structured, semi-structured, and unstructured), the role of AI systems (e.g., expert systems) in decision-making may differ (Duan et al., 2019).

Research suggests that AI can replace human decision-makers for structured or semi-structured decisions (Duan et al., 2019). However, for unstructured decisions at the strategic level, AI is more effective as a decision-support tool rather than a decision-making tool (Duan et al., 2019). This is in line with the experimental task designed for this study, as will be further explained in chapter 3.

In the context of organizational collaborative decision-making, human-AI teams are expected to perform better than AI alone or humans alone, particularly in areas requiring accuracy of decision-making (Liu et al., 2021), such as financial forecasting (Addy et al., 2024) and medical diagnosis (Rajpurkar et al., 2018). Over the past fifty years, AI systems have been a focal point for many researchers (Go and Sundar, 2019). Significant efforts have been made by designers to develop intelligent chat agents that are more human-like, particularly by enhancing their capability to use human language (Go and Sundar, 2019).

2.4 Chatbots Humanness and Anthropomorphism

Over the past decade, there has been a noticeable increase in interest focused on text-based chatbots, which are software applications that interact with humans through natural written language (Rapp et al., 2021). They have become so popular that the global chatbot market is forecasted to experience an annual growth rate of 23.5%, expanding from $2.9 billion in 2020 to reach $10.5 billion by 2026 (Lee et al., 2023).

Chatbots can be classified based on the goals or primary objectives. There are three types of chatbots with distinct purposes involving:

1. Informative Chatbots: These chatbots provide users with information from pre-stored or fixed sources, similar to FAQ chatbots. They deliver specific details and answers to user queries (Adamopoulou and Muossiades, 2020).
2. Chat-based/Conversational Chatbots: These bots engage users in conversation, mimicking human-like interactions. Their primary goal is to respond accurately and naturally to the messages they receive (Adamopoulou and Muossiades, 2020).
3. Task-based Chatbots: These chatbots focus on performing specific tasks, such as booking a flight or assisting with a particular action (Adamopoulou and Muossiades, 2020).

These chatbots are transforming the ways humans interact with computers (Chaves and Gerosa, 2021). By mimicking human conversations through text, they facilitate interactions between users and algorithms (Shin, 2022).

By enhancing the human-likeness of chatbots, users tend to perceive it as more human or exhibit a greater sense of anthropomorphism towards it (Jakobsen, 2021). Anthropomorphism refers to the inclination to attribute human-like characteristics, motivations, intentions, or emotions to the real or imagined behavior of nonhuman agents (adapted from Lu et al., 2022). Anthropomorphism of computer systems is a widespread phenomenon because humans tend to perceive them as social actors and interact with them accordingly (Schuetzler 2020). When anthropomorphism incorporates visual elements (such as human figures) and linguistic cues, it has the potential to trigger a human schema in individuals which can significantly impact people's judgments and behaviors (Roy and Naidoo, 2021). In other words, anthropomorphism fosters a stronger sense of connection and engagement with anthropomorphized objects, leading to more interesting interactions and influencing consumers' decision-making behaviors (Han, 2021). Similar benefits of human-like features have been demonstrated for products, where products displaying human-like traits (i.e. hand and arm gestures) are perceived as more trustworthy by people (Han, 2021).

2.5 User System Satisfaction

An important antecedent of performance in the context of technology use, is that of user satisfaction. In the sphere of digital and online product analysis, professionals across various disciplines are keenly interested in user satisfaction as a pivotal metric for evaluating system success (Pozón López et al., 2021). Cognitive theories posit that user satisfaction and technology adoption are influenced by the objective and instrumental value individuals derive from technology interaction, encompassing improvements in task performance and efficiency (Coursaris and Van Osch, 2016). According to these theories, satisfaction emerges when perceived benefits surpass the costs associated with technology adoption and usage (Coursaris and Van Osch, 2016). With this approach user satisfaction can be defined as a subjective assessment of the outcomes of Information Systems (IS) use, evaluated on a continuum of pleasant to unpleasant, and encompasses factors linked to system characteristics, information characteristics, and the service and support provided to users (Karlinsky-Shichor and Zviran, 2016). On the other hand, according to affective theories, satisfaction and technology use are influenced by the subjective and self-fulfilling value users derive from their interactions with technology, such as enjoyment and entertainment (Coursaris and Van Osch, 2016). These theories propose that user satisfaction originates from the pleasurable experiences and sensations users encounter (Coursaris and Van Osch, 2016).

Previous research on chatbots has demonstrated that employing an empathetic, human-like language style during interactions with AI systems can significantly influence task satisfaction (Vanderlyn et al., 2021). This conversational approach is typically favored over strictly informational exchanges, as users tend to exhibit less acceptance towards direct, command-like language (Vanderlyn et al., 2021). Moreover, the integration of personal pronouns and self-referential language has been shown to enhance the overall user experience, fostering a more positive, personalized, and natural interaction (Vanderlyn et al., 2021).

A previous study in the context of organizational decision-making using business intelligence systems revealed a positive correlation between user satisfaction and decision quality (Kapo et al., 2021). In other words, the findings of this study indicated that higher user satisfaction and intention to use are linked to increased system usage, and both augmented user satisfaction and system usage contribute to improved individual user performance (Kapo et al., 2021).

2.6 Interpersonal Trust

Within the context of group decision-making, a pivotal determinant is the trust established among group members (Zhang et al., 2022). Trust is an essential construct that enables individuals and organizations to effectively manage working in complex, dynamic, and human-centric environments, especially when teams strive for optimal performance (Borum, 2010). The results of a study on trust (Zaheer et al.,1998) indicated that interpersonal trust can increase the performance of the individuals and can directly affect the level of conflict between them during discussion sessions. Interpersonal trust is significantly influenced by how individuals interact with others (Geller, 1999). Particularly, the way they communicate and express themselves—through speaking, writing, or using signals—plays a crucial role in building trust (Geller, 1999). A relevant previous study showed that trust among the team members who collaborate and communicate through digital technologies has a significant positive effect on their performance (Chang et al., 2014). Also, linguistic style used during these interactions can have an impact on people's perceptions through feelings of interpersonal similarities (Leong et al., 2021). A study found that people tend to trust individuals who use more eloquent and structured language (Jucks et al., 2016). The study also noted that the type of conversational partner influences linguistic behavior, causing each person to adjust their word choice based on their partner's words (Jucks et al., 2016). This linguistic adjustment happens more frequently when an AI agent is involved in the conversation (Jucks et al., 2016).

2.7. Hypothesis Development

2.7.1. Communication Style of the Chatbot and Group Decision-Making

Over the past five decades, collaborative decision-making has become a key focus in organizational studies, as organizations increasingly prioritize teamwork and collaboration (Halvorsen, 2018). In recent years the use of AI in organizational decision-making has become prevalent (De Vreede et al., 2021). Thus, studying the impact of AI-powered chatbots on team collaboration is crucial (Yan and Gurkan, 2023). However, this task presents challenges due to the complexity of teams, where interactions between members and technologies occur through emergent and dynamic processes (Yan and Gurkan, 2023). Despite AI's superior cognitive capacity and its potential to enhance human cognition and team decision-making, comprehending its effects remains intricate (Yan and Gurkan, 2023).

Collective decision performance can be measured in three ways, namely accuracy, efficiency, and satisfaction with the process. Decision accuracy refers to the average difference between the decision made by the team and the correct decision (Hedlund et al., 1998). A second measure of collective decision-making performance is the time it took for the group to make the final decision, referred to as decision efficiency (Paul et al., 2004). Finally, one can measure the performance through the satisfaction of the group members with the decision process (Paul et al., 2004). Process satisfaction can be gauged by the degree of contentment among group members with the method and manner in which the final decision was reached (Paul et al., 2004).

In recent years, AI has been used to improve decision-making efficiency in many different areas such as healthcare, military, customer service, and so on (De Vreede et al., 2021). AI agents have made significant inroads into society and industry, manifesting in various forms, one of which is conversation agents, commonly known as chatbots (De Vreede et al., 2021). Chatbots are becoming increasingly popular, as an example in consumer retail spending through chatbots worldwide is projected to reach $142 billion by 2024, a significant increase from the $2.8 billion reported in 2019 (www.insiderintelligence.com). Chatbots are computer programs created with the purpose of imitating human conversation, enabling them to engage in text or voice-based interactions with individuals (Han, 2021). By understanding and processing human language, chatbots create a conversational experience that closely resembles interactions with humans (Morana et al., 2020). Perceptions of chatbots as human-like entities and to attribute human-like qualities to them is the core concept of anthropomorphism theory (Roy and Naidoo, 2021). The notion of human-likeness seems to be a complex concept that evolves along a spectrum ranging from high levels of human-likeness to low levels of human-likeness (Rapp et al., 2023). Moreover, the concept of human-likeness is complex and multifaceted (Rapp et al., 2023). Also, it depends on the context, is modular and dynamic (Rapp et al., 2023). This means that it shows up in different ways, changes depending on the situation, involves specific abilities, and evolves over time (Rapp et al., 2023). The anthropomorphism theory (perceiving non-human entities as human-like) constitutes the central premise that underlies our scholarly work.

Based on past studies, chatbots can be more anthropomorphized (human-like) when they use more words per message, a higher percentage of articles, and a higher number of words containing more than six letters (Hill et al, 2015). Another feature of the text-based chatbots that leads to being perceived more human-like is the use of emojis (Rapp et al., 2021). Using first-person singular pronouns is another linguistic trait of these kinds of chatbots that can increase the anthropomorphic (human-likeness) perceptions by the user (Adam et al., 2021). It is noteworthy that while it is improbable for these language models to completely replace human participants, they can serve as valuable supplements during the idea generation and refinement stage of the collaborative decision-making (Dillion et al., 2023).

Research has shown that chatbots with more verbally anthropomorphic features can significantly enhance user experience (Klein & Martinez, 2023). Specifically, anthropomorphic cues in chatbot language styles have been positively associated with higher satisfaction (Klein & Martinez, 2023). This phenomenon can be explained through Media Richness Theory, which posits that media vary in their ability to facilitate understanding and convey nuanced information among team members (Kahai & Cooper, 2003). According to this theory, richer media that convey more cues and detailed information enhance the effectiveness and efficiency of communication for complex tasks (Kahai & Cooper, 2003).

Based on the aforementioned studies and insights produced by media richness theory, we argue that as more human-like language conveys information with more details and cues the interaction with more human-like chatbot is more engaging and satisfying, and this will lead to higher process satisfaction.

Moreover, studies have shown that chatbots can increase the efficiency of the decision process (Majumder & Mondal, 2021). Specifically, systems with a more natural and human-like language will increase the efficiency (Warren and Pereira, 1982). This can also be explained through Media Richness Theory. Past studies indicate that the form of language and communication of the chatbot affects the ambiguity or clarity of the chatbot's message (Murtarelli et al., 2020). Specifically, the use of clear and detailed statements facilitates understanding of the message (Murtarelli et al., 2020). We argue that a richer chatbot in terms of language and communication will convey clearer messages, reducing the message processing time and increasing the efficiency of the decision-making process.

Finally, as mentioned before, employing human-like language by chatbots can induce anthropomorphic (human-likeness) perceptions (Seeger et al., 2021), and past studies have shown that increasing anthropomorphic features of the chatbot can enhance effectiveness (Roy & Naidoo, 2021). Another study on anthropomorphic decision aids showed that making these software programs more human-like increases users' performance in their decision-making tasks (Pak et al., 2012). Furthermore, high-quality communication, which is crucial for optimal performance in any human collective, is significantly influenced by language usage (Bucăţa & Rizescu, 2017). Media Richness

Theory explains the effects of chatbot humanness in this context well. This theory proposes that task performance improves when task-information processing requirements match a medium's ability to convey information richness (Suh, 1999). Studies have shown that richer media has higher ability to convey accurate data (El-Shinnawy & Markus, 1992). Therefore, we argue that a more human-like language, with enhanced capability for information transfer and understandability, will increase the accuracy of decisions in team decision-making.

With that in mind, we propose the following research hypotheses:
H1a: Chatbot communication style impacts group decision-making process satisfaction, such that human-like communication will be associated with greater process satisfaction than robot-like communication.
H1b: Chatbot communication style impacts group decision-making efficiency, such that human-like communication will be associated with greater efficiency than robot-like communication.
H1c: Chatbot communication style impacts group decision-making accuracy, such that human-like communication will be associated with greater accuracy than robot-like communication.

2.7.2 Communication Style of the Chatbot and User Satisfaction with System

The importance of human-robot collaboration has gained increasing recognition in the era of Artificial Intelligence, as robots continue to expand their involvement across various domains of human life (Ye et al., 2023). Within the academic realm, researchers are not only dedicated to improving content cognition in human-robot dialogues from an engineering perspective but also committed to understanding user experience in human-computer interaction (Wen, 2018). A prominent area of research revolves around debating whether robots should communicate using natural human languages or function solely as machines (Wen, 2018). During interactions with AI, individuals tend to apply interpersonal and relational norms to these interactions (Westerman et al., 2020). Moreover, individuals tend to attribute human personality traits to computers and artificially intelligent agents (Beattie et al., 2020). Generally, it can be concluded that people expect robots to behave like humans (Candello et al., 2017).

As discussed before, using human-like language by chatbot will increase the probability that it will be perceived anthropomorphic (human-like) by users. In understanding the factors that trigger anthropomorphism, three key elements emerge (Epley et al., 2007). Firstly, elicited agent knowledge, which occurs when individuals interact with unfamiliar non-human entities (Epley et al., 2007). Secondly, sociality, which arises when there's a need for social interaction (Epley et al., 2007). And finally, effectance motivation, which explains that anthropomorphism happens due to humans' innate desire to understand and exert control over their environment (Epley et al., 2007). Technology users are often unaware of new non-human automation systems but must rely on them to complete tasks. Introducing anthropomorphic (human-like) traits to digital agents can mitigate uncertainties, foster a sense of familiarity, and aid users in developing a bond with the technology (Klein and Martinez, 2023). This, in turn, elicits a more positive and emotionally charged reaction toward the technology (Klein and Martinez, 2023). This strong emotional appeal will consequently lead to user satisfaction with the system (Coursaris and Van Osch, 2016).

User Satisfaction with the system can be conceptualized as “the overall affective evaluation that an end-user has about the experience with the application system” (Prastyo et al., 2021, p. 2). Existing literature on chatbot technology offers empirical evidence suggesting that incorporating human characteristics into chatbots enhances the user experience (Klein and Martinez, 2022). Also, the results of a study showed that the conversational quality (human-like language) of AI chatbots has a significant impact on user satisfaction with the system (Hsu and Lin, 2023).

The impact of the communication and language style of the chatbot on user satisfaction can be better understood through the lens of Media Richness Theory. Media richness refers to the capacity of a communication medium to effectively process rich information (Gimpel et al., 2016). According to studies, one aspect of media richness is Language

Variety, which indicates the medium's ability to support dialogues that involve both numerical data and natural language. Another aspect is Personal Focus/Source, which refers to the medium's ability to convey the personal feelings and emotions of the participants in the dialogue (Gimpel et al., 2016). Communication channels can be considered high or low in "richness" based on their ability to facilitate shared understanding and rapid insight (Gimpel et al., 2016). When a medium is richer and able to convey more cues and nuanced information, and is more natural in terms of language, user satisfaction tends to increase (Fleischmann et al., 2020). Therefore, we argue that a human-like language, which can more effectively convey meanings and emotions, will result in higher user satisfaction. Based on the above argument we propose the second hypothesis as:

H2: Chatbot communication style impacts users' satisfaction, such that human-like communication will be associated with greater user satisfaction than robot-like communication.

2.7.3 Communication Style of the Chatbot and Interpersonal Trust

Trust is a fundamental element of human interaction, shaping the dynamics of numerous organizational and business situations (Gefen et al., 2020). The definition of trust based on the available literature can be an "individual's willingness to be vulnerable to the actions of another person" (Yu et al., 2022, p. 1074). Trust among individuals entails the expectation or belief in others' intentions and motives, that they will act in a manner beneficial or at least not detrimental to the relationship and their shared objectives (Costa, 2003). From a psychological point of view, trust can be presented as the ability to depend on another individual in situations involving risk or as a relationship that fosters cooperation (Chudzicka-Czupała et al., 2022). It has been demonstrated by past research that even short interpersonal interactions have the potential to enhance optimal levels of trust among team members with communication and the way individuals communicate being a significant antecedent for trust development (Langlinais et al., 2022). Different factors affect trust building, one of which is benevolence (Chudzicka-Czupała et al., 2022). Trust in others' benevolence will be built when the trustor sees the trustee’s positive attitude and favorable behaviors (Chudzicka-Czupała et al., 2022). Benevolence has been defined as “the extent to which a trustee is believed to want to do good to the trustor aside from an egocentric profit motive” (Dirks and de Jong, 2022, p. 251).

The anthropomorphism of chatbots has been found to influence users' behavior and responses during communication (Cheng et al., 2022). Particularly, when chatbots use human-like language during interactions, users tend to feel social and emotional connections with them, forming relationships in a manner similar to how they would with human agents (Cheng et al., 2022). Research has revealed that individuals adapt their communication styles to match those of their conversational partner, irrespective of whether the partner is a human or a machine (Candello et al., 2017). In the realm of successful interpersonal relations and team rapport within organizations, the language style is a critical factor (Cohen and Henderson, 2012), and past research has consistently shown that human speech is rated higher in terms of trust compared to robot-like speech (Candello et al., 2017). Furthermore, recent studies have indicated a positive impact on trust development between individuals in Chinese firms when specific information systems are employed (Lissillour et al., 2023). Therefore, we propose that utilizing a chatbot with a human-like language style in an online collaborative work setting will likely lead to increased interpersonal trust among group members as it can affect the way they communicate with each other.

Media Richness Theory provides further insight into the relationship between chatbot language style and interpersonal trust. Richer media with more natural language can better ensure the transfer of information to the receiver, reducing potential conflicting interpretations or misunderstandings and conveying more social cues (Gimpel et al., 2016). In this way, richer media can improve communication among remote team members (Lo and Lie, 2008). We argue that since communication is a key element in building interpersonal trust, richer media with greater capacity to convey information, emotions, and foster social bonds can increase the level of trust. This aligns with past research showing that in situations where tasks are complicated and prior trust levels are

not high, users prefer richer media for communication to receive more nuanced information and emotional and social cues from the other party (Lo and Lie, 2008).
Based on the above, we propose H3:

H3: Chatbot communication style impacts interpersonal trust, such that human-like communication will be associated with greater interpersonal trust than robot-like communication.

2.7.4 Collaborative Decision-making Performance and User Satisfaction

As aforementioned, decision-making performance can be assessed through three facets. The first is decision accuracy, which is an unobtrusive way of assessing performance, refers to the sum of the absolute differences between the ranks assigned to the items of the task by an expert and by the team members (Hamada et al., 2020). Collective decision-making performance can be further measured through decision efficiency, another unobtrusive measure, which is the time it took for the group to make the final decision, as well as the satisfaction of the group members with the decision process (Paul et al., 2004). Finally, process satisfaction, which is a perceived measure, can be assessed by the level of contentment of the group members with the method and manner in which the group reached the final decision (Paul et al., 2004). Regarding the relationship between satisfaction and performance, previous studies in the context of online learning showed that students' satisfaction has a positive impact on their performance (Rajabalee and Santally, 2020).
In the context of UX, past studies have demonstrated that there is a positive correlation between user satisfaction and user performance, as favorable system perceptions result in better performance outcomes over time (Hartson and Pyla, 2012). Paul and colleagues (2004) in their study showed that there is a direct significant relationship between user satisfaction with the system and decision efficiency and satisfaction with the decision process.
Another argument here is that, based on past studies, user satisfaction stems from the user’s perception of how useful a system is in increasing efficiency and effectiveness, as

well as how well it provides feedback and delivers accurate, high-quality information (Kapo et al., 2021). Cognitive theories suggest that user satisfaction and technology adoption are influenced by the objective and instrumental value individuals derive from their interaction with technology (Coursaris and Van Osch, 2016). According to these theories, satisfaction occurs when system usage benefits the user (Coursaris and Van Osch, 2016). We argue that a higher level of user satisfaction indicates that the system is effectively providing the user with good quality information encouraging more usage and engagement, which in turn enhances the user's performance.

Thus, we propose our H4a, H4b, and H4c hypotheses as:

H4a: User satisfaction has a positive effect on the group decision-making process satisfaction.
H4b: User satisfaction has a positive effect on the group decision-making efficiency.
H4c: User satisfaction has a positive effect on the group decision-making accuracy.

2.7.5 Interpersonal Trust and Collaborative Decision Performance

Trust stands as the foundational element when it comes to establishing relationships within a remote team (Yousfi and Anand, 2021). Consequently, teams lacking in trust are susceptible to encountering significant challenges in their collaboration efforts, including misunderstandings, and interpersonal conflicts within the context of a virtual environment (Yousfi and Anand, 2021). Also, trust is widely acknowledged as the fundamental determinant that plays a pivotal role in the success or failure of remote teams (Zapata et al., 2021). It is also proven by previous studies that interpersonal trust has a significant effect on the decision-making results in the context of collaborative decision-making (Park et al., 2014). There is a direct link between trust and the performance of collaborative relationships (Trejo, 2021). Once the need for collaboration is established, trust becomes the key determinant of performance (Trejo, 2021).
The relationship between interpersonal trust and collaborative decision performance can be explained through the effect of trust on facilitating effective interactions and fostering

positive intentions to engage with one another in a remote team setting (Yousfi and Anand, 2021). Additionally, trust has been observed to influence how motivation is translated into work group processes and performance (Dirks and Kurt, 1999). We argue that higher engagement and effective communication, as well as increased motivation derived from higher interpersonal trust, will lead to significant improvements in team performance.

Based on the above discussion, we propose our next hypothesis as:

H5a: Interpersonal trust has a positive effect on the group decision-making process satisfaction.

H5b: Interpersonal trust has a positive effect on the group decision-making efficiency.

H5c: Interpersonal trust has a positive effect on the group decision-making accuracy.

2.7.6 Chatbot Communication Style, User Satisfaction and Collaborative Decision Performance

Language style pertains to how the content is conveyed by the chatbot (Ireland and Pennebaker, 2010). When a chatbot's language style is more human-like, users will feel more connected to it (Hsu and Lin, 2023). According to Duffy (2003), incorporating anthropomorphic features into products and services can lead to higher sales revenue and greater satisfaction of customers. A method for enhancing the anthropomorphic attributes of chatbots involves the emulation of human language style (Nguyen et al., 2022) and this increase in the linguistic anthropomorphic attributes has been shown to have a positive impact on user satisfaction in prior studies (Zheng et al., 2023).

In turn, higher user satisfaction is related to higher decision performance as the higher satisfaction encourages increased system usage and this way by continuous usage the decision performance will increase (Boukhayma et al., 2019). Also, user satisfaction is an important factor in increasing the engagement of the team members which can lead to a higher performance (Kim et al., 2013, Awan et al., 2020).

As discussed before, a more human-like language by a higher ability to conveying nuanced information can increase the decision-making performance of the team. On the other hand, past studies showed that a more human-like language can increase the user satisfaction. This increased user satisfaction in turn will increase the performance through a higher motivation and engagement in the process.

Based on the information provided, the following hypothesis is formulated:

H6a: In the relationship between chatbot communication style and group decision process satisfaction, user satisfaction has a mediating role.

H6b: In the relationship between chatbot communication style and group decision efficiency, user satisfaction has a mediating role.

H6c: In the relationship between chatbot communication style and group decision accuracy, user satisfaction has a mediating role.

2.7.7 Chatbot Communication Style, Interpersonal Trust and Collaborative Decision Performance

Language also serves as an essential bond in the development of social groups (Li and Hu, 2020). It reflects, solidifies, and enhances the connections among individuals within groups (Li and Hu, 2020). Furthermore, it was shown by past research that when individuals have little knowledge about others they would judge their trustworthiness through their verbal and nonverbal communication style (Sarapaivanich et al., 2019).

As explained earlier, individuals adapt their language style to match the chatbots they use (Nguyen et al., 2022). Using algorithmic responses leads to changes in language and social relationships, including faster communication, more positive language usage, and an increased perception of closeness and cooperation among conversation partners (Hohenstein, et al., 2023).

Based on these studies, we argue that if a chatbot's communication style is human-like, it enables group members to communicate more effectively and positively. This is achieved through a language style that fosters interpersonal trust by facilitating the expression of more verbal and nonverbal cues and increasing the likelihood of forming closer social connections. This increased trust and enhanced communication will in turn lead to higher group performance. This aligns with previous research that demonstrates the significant role of interpersonal trust in influencing the outcomes of decision-making within collaborative contexts (Park et al., 2014).

Therefore, we propose the following hypotheses:

H7a: In the relationship between chatbot communication style and group decision process satisfaction, interpersonal trust has a mediating role.

H7b: In the relationship between chatbot communication style and group decision efficiency, interpersonal trust has a mediating role.

H7c: In the relationship between chatbot communication style and group decision accuracy, interpersonal trust has a mediating role.

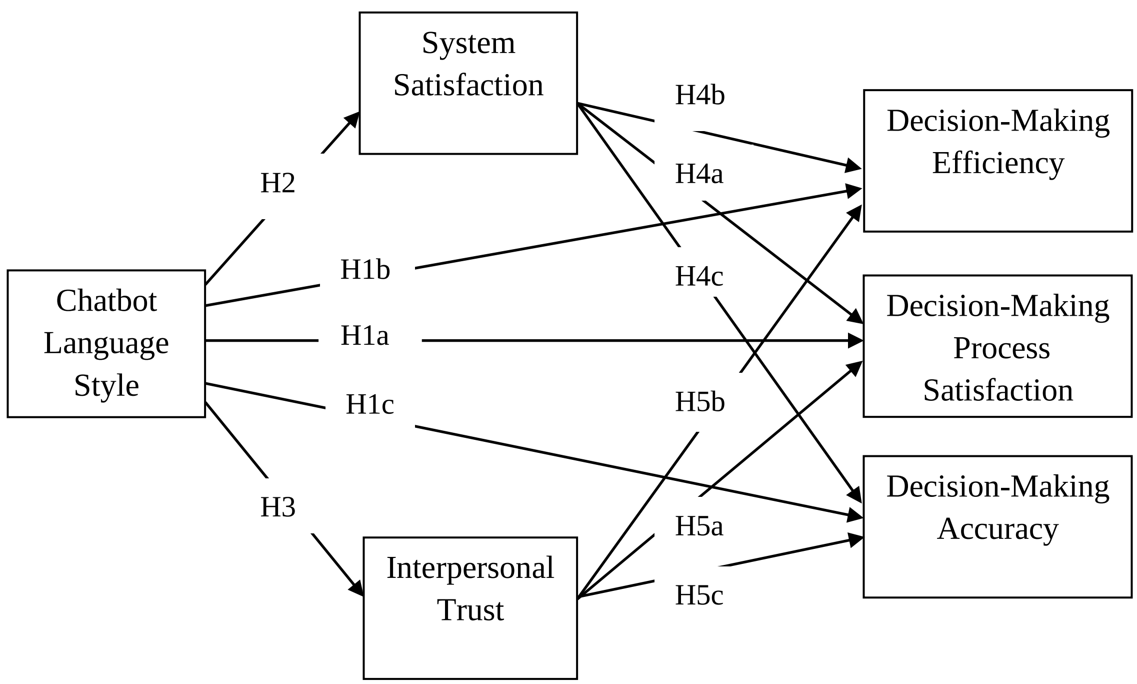

Figure 2. Proposed Research Model

Chapter 3: Methodology

3.1 Data Collection

In this study, we employed a between-subjects experimental design that involved participants completing three distinct tasks for data collection. The entire experiment was conducted online using the Chatzy platform. This research received approval from the university's ethics committee under reference number 2024-5836. In December 2023, a pre-test was conducted using a convenience sample of participants aged 18-24, who volunteered to participate. This pre-test was instrumental in refining the survey flow and the experimental session design. In the first two weeks of February 2024, a pilot study was conducted with four groups per condition, each consisting of three participants. The pilot study results helped in refining and improving the clarity of the instruction prompts for the chatbot scripts. Finally, during the primary data collection phase, which spanned from February 13th to March 25th, 2024, we initiated the data collection process with participants recruited remotely through the Prolific platform. Participants were randomly and equally assigned to one of four experimental conditions: 1. robot-like chatbot with the role of a facilitator (n = 51), 2. robot-like chatbot with the role of an ideator (n = 46), 3. human-like chatbot with the role of a facilitator (n = 44), and 4. human-like chatbot with the role of an ideator (n = 48). They were asked to complete three decision-making tasks in groups of 3-4 members, with a chatbot consulting each group. The chatbot interaction was entirely text-based, and no video or voice communication was allowed. To prevent any confounding effects, we used a gender-neutral name, "Ideabot," for the chatbot. Previous studies have shown that feminine names and avatars tend to be perceived as more human-like (Borau et al., 2020). By using a gender-neutral name, we aimed to isolate the effect of language use on perceived humanness, without the influence of other aspects of the chatbot design that could enhance its perceived humanness.

We employed the Wizard of Oz technique for this experiment. In a "Wizard of Oz experiment," users interact with a computer interface believing the interaction is automated; however, it is actually controlled by another individual, known as the "wizard" (Eberhart et al., 2022). As we used the Wizard of Oz technique and we wanted all groups to encounter the same behavior and information from the chatbot, we used pre-made scripts for the chatbot (for chatbots' scripts, please refer to Appendix A). In order to increase ecological validity, we employed ChatGPT to finetune scripts. The chatbots were task-based in nature and their prompts were designed based on the tasks, not the flow of the conversations among group members. After the participants had done each set of tasks, we directed them to complete a survey to measure their perceived process satisfaction, system satisfaction and interpersonal trust. We also assessed their efficiency by measuring the time it took each group to reach a consensus in task three. The accuracy of the decision making was measured using an error score, calculated as the sum of the absolute differences between the ranks assigned to the items by marketing experts and by the participants (Hamada et al., 2020). Lower error scores are indicative of superior performance and accuracy in terms of making reasoned judgments (Hamada et al., 2020). We measured the objective measures (accuracy and efficiency) directly at the group level. However, the self-reported (i.e., subjective) measures (decision-making process satisfaction, system satisfaction, and interpersonal trust) were initially assessed at the individual level using survey questions. To ensure consistency in our analysis, we calculated the group level measures by averaging the individual-level scores for each construct within each group. This approach for obtaining group-level scores from individual-level responses was justified given that the data was collected through an experiment and we thus had complete responses for all members of the group.

3.2 Participants

We recruited a total of 191 participants across 61 groups through the Prolific[1] platform, based in Canada, the USA, and the UK. Out of these, 189 complete surveys were received and used for this study. Participants were aged 18-64 and held a master's degree or higher. Each participant received 15 CAD for their one-hour participation in our study.

[1] https://www.prolific.com/

Eligibility criteria included advanced English proficiency, a high level of education (graduate degree), and experience using technology for collaborative tasks. These criteria were met through the recruitment of employees and university students. This approach ensured that participants were likely to have prior experience in professional collaborative contexts, remote team activities, and decision-making tasks, thereby enhancing the study's ecological validity. Participants provided their informed consent by filling in their username / name and the date, choosing to proceed with the study, and clicking on an ‘accept’ button prior to participating in the study. Participants were allowed to leave or stop the experiment at any time.

Table 3. Sample demographics

Sex	Category	Number	Proportion
	Male	98	51.85%
	Female	87	46.03%.
	Nonbinary	4	2.12%.
Age			
	18-24	73	19.58%
	25-34	55	38.62%
	35-44	37	29.10%
	45-54	19	10.05%
	55-64	3	1.59%
	+65	2	1.06%

3.3 Experimental Design

We employed a between-subjects, scenario-based experimental design to assess how the communication and language style of the chatbot affects group decision performance. The language style of the chatbot varied between two levels: human-like and robot-like (see

Table4). In the human-like condition, the chatbot used longer sentences, personal pronouns such as "I" and "me," and emojis. In the robot-like condition, the chatbot used shorter responses, avoided personal pronouns, and did not use emojis.

Table 4. Chatbots linguistic manipulation

Chatbot	Traits
Human-like	Long sentences Use of personal pronouns such as I and me Use of emojis
Robot-like	Short sentences No use of personal pronouns such as I and me No use of emojis

This study also included an exploratory analysis of the chatbot's role, manipulated between two levels: ideator and facilitator. As an exploratory post hoc component of this thesis, this manipulation examines how different chatbot roles affect the decision performance of the teams. To investigate this effect, we designed the chatbot script with distinct roles: the ideator role, during which the chatbot acted as a peer that actively proposed ideas and task-related responses. In contrast, the facilitator role focused solely on managing the conversation flow and encouraging participants to engage and maintain the discussion.

Table 5. Chatbot role manipulation

Chatbot	Traits
Ideator	Gives instructions Share task related idea and information Keeps track of time
Facilitator	Gives instructions Encourage engagement and lead the flow Keeps track of time

After conducting an extensive literature review, we concluded that a team size of three members is optimal for experimental activities, as supported by previous research (Alves et al., 2023). To accommodate this, we formed teams consisting of 3-4 participants. Two out of five sessions with three participants would be cancelled due to one of the participants not showing up. To address the issue of high drop-out rates, we employed an oversampling strategy by including up to four participants per team which could otherwise result in a significant number of cancelled sessions.

Given the constraints of our data collection timeline, this approach allowed us to ensure that we would still have teams with at least three members, even if one participant did not attend. This strategy was crucial for maintaining the integrity of our study while managing the practical challenges associated with participant attendance.

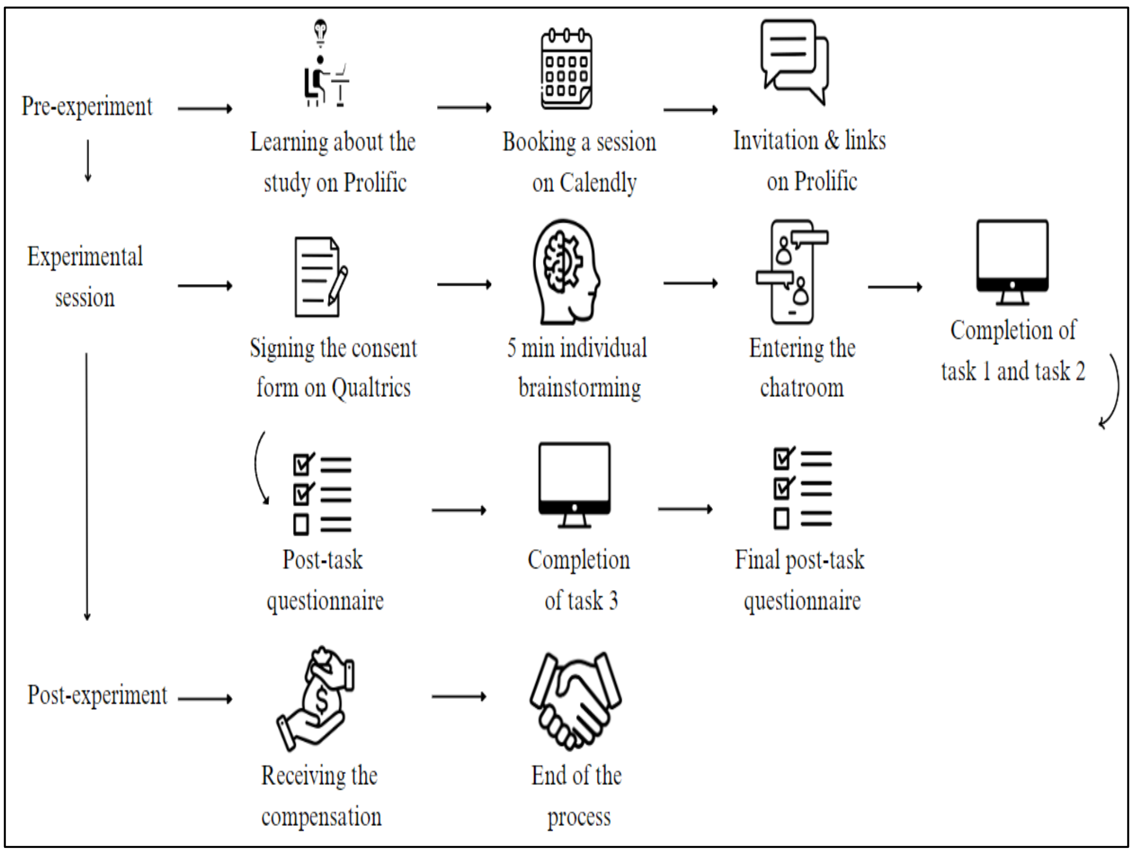

Figure 3. Experimental procedure

3.3.1 Pre-experiment:

Pre-experiment procedures involved participants learning about the experiment through Prolific and scheduling their participation via Calendly[2], an online scheduling service, for a specified time slot. Thirty minutes before the session, participants received invitation links containing access to both the chatroom (Chatzy[3]) and the survey (Qualtrics[4]). Upon accessing the survey link, participants read and signed the consent form, completed brief demographic questions, and reviewed instructions for the first task. Additionally, they engaged in a five-minute individual brainstorming session to prepare for the first task. Upon all three participants saying "Hi," the chatbot would promptly send the instructions and commence timekeeping for the participants.

3.3.2 Experimental session:

The experimental protocol comprised three distinct tasks, each of which will be explained in detail below.

3.3.2.1 Task One:

The first task involved collaborative brainstorming among participants. They were tasked with generating as many slogans as possible to promote public transportation among Canadian university students aged 18-24. They consulted with each other while also interacting with the chatbot throughout the process. They were told that the slogans they generated would be evaluated for potential inclusion in a public transportation campaign, with the best-performing group receiving a gift card as an incentive to foster competition among them. This competitive element was deemed necessary to ensure the meaningful measurement of trust. The rationale for this lies in trust definitions that underscore the significance of risk, defined as the perceived likelihood of loss evaluated by those entrusting (Costa, 2003). Additionally, some scholars posit that this sense of risk plays a

---

[2] https://calendly.com/
[3] https://www.chatzy.com/
[4] https://www.qualtrics.com/

pivotal role in trust decisions, asserting that without it, trust measurement lacks meaningfulness (Costa, 2003).

During the first task, the chatbot would adopt either a human-like or robot-like language style, depending on the experimental condition. In terms of its role, the chatbot would propose slogans as an ideator, whereas in facilitator conditions, it would orchestrate group engagement with the task. Operating via the Wizard of Oz technique, the chatbot was controlled by a human operator who followed a predetermined script at specific time intervals. Initially, the chatbot would send the instruction and the first task-related messages without any delay, followed by subsequent messages at one-minute intervals. This approach allowed participants to have enough time to brainstorm and interact with both each other and the chatbot. The scripts were designed differently for each condition to manipulate the language style and the chatbot's role. Notably, we used the ChatGPT chatbot to create these scripts, ensuring they reflected real chatbot interactions.

In all conditions, the chatbot monitored the time. During this task, it would alert participants at the eight-minute mark that they had two minutes remaining. Then, after 10 minutes, the chatbot instructed participants to stop their brainstorming for slogans, marking the end of Task One. The 10-minute time interval was chosen for this task to let participants get familiar with the system and its functions while also having enough time to brainstorm with each other.

Table 6. Chatbot characteristics in each condition in task one and two

Robot-like and Facilitator Short sentences No personal pronouns No emojis No suggestion of the slogans Orchestration of process Keep Track of the Time	**Robot-like and Ideator** Short sentences No personal pronouns No emojis Suggestion of the slogans in addition to orchestrating the decision process Keep Track of the Time
Human-like and Facilitator Long sentences Use of personal pronouns Use of emojis No suggestion of the slogans Orchestration of process Keep Track of the Time	**Human-like and Ideator** Long sentences Use of personal pronouns Use of emojis Suggestion of the slogans in addition to orchestrating the decision process Keep Track of the Time

3.3.2.2 Task Two:

For the second task, which focused on the convergent phase of decision-making, the chatbot instructed participants to review all generated slogans and select the one they deemed the most effective and suitable for the campaign's target audience (18-24-year-old Canadian university students). To ensure consensus among all group members regarding the final decision, the chatbot instructed team members to write their choice in a specified format, requesting others to confirm the final selection by typing ‘Yes’ to indicate agreement afterward. A seven-minute time limit was allocated for this task, with a warning message sent to the group two minutes before the end to prevent teams from exceeding the time limit and failing to complete the task. This seven-minute interval was selected based on our pilot tests, during which we observed that most groups discussed their favourite slogans even during the first task and generally they would complete their choices within 6 to 7 minutes after the start of task 2.

After the final decision was written and confirmed by all team members, the chatbot would send the final message of task two to direct participants to the survey. At this point, participants would answer a few questions about their satisfaction with the decision-making process and review the instructions for Task Three. During the second task, when the chatbot assumed the role of an Ideator, it selected one of its own slogans generated during Task One to mimic being one of the members of the team. Conversely, when the chatbot acted as a facilitator, it simply instructed participants to voice their opinions and discuss their preferred slogan, along with the reasons why it should be chosen as the group's final best slogan. The linguistic characteristics of the chatbot vary between two levels in each of the four conditions throughout the script.

3.3.2.3 Task Three:

Task Three was primarily designed to measure decision-making accuracy and efficiency. This task was informed by an extensive review of previous group decision-making experiments which typically involved a rank-ordering task for measuring decision-making accuracy and efficiency and adapted to fit the context of this experiment (marketing campaign). Participants were tasked with collectively ranking six different social platforms based on their perceived effectiveness for diffusing the campaign's slogan, considering the target audience based on their knowledge and experience of social media. The platforms included YouTube, Facebook, Instagram, TikTok, Twitter (X), and LinkedIn. The task commenced once all three participants returned to the chatroom and indicated their presence by typing 'Back'. In this task, the ideator chatbot would provide participants with statistics about the usage of each platform by Canadian university students aged 18-24, along with its own suggested ranking for each platform based on these statistics. All the statistics were obtained using ChatGPT and Statista. In facilitator conditions, the chatbot would prompt participants to comment on each platform in a particular random order across all groups. This task was designed to be completed within ten minutes, with participants not initially informed of the allocated time to measure their efficiency naturally. The chatbot would send the first two messages containing instructions and the initial response back-to-back, followed by one message per minute. At the 7-minute mark, a warning message indicating 3 minutes remaining would be sent so that participants would have a 3-minute time interval without distractions from the chatbot so that consensus-reaching time was not interrupted. If a decision had not been reached by minute 9, another warning would be sent at the 9-minute mark. The final message, either sent immediately after the team's final decision and confirmation by all members or at the 10-minute mark, would redirect participants to the survey to answer final questions. It is noteworthy that the chatbot shared statistics that could help point the team members to an optimal ranking, but it did not provide team members with the exact rankings to ensure the ideator condition did not inadvertently affect accuracy.

Table 7. Chatbot characteristics in each condition in task three

Robot-like and Facilitator	Robot-like and Ideator
Short sentences No personal pronouns No emojis No suggestion of the rankings and statistics Orchestration of process Keep Track of the Time	Short sentences No personal pronouns No emojis Suggestion of the rankings and statistics in addition to orchestrating the decision process Keep Track of the Time

Human-like and Facilitator	Human-like and Ideator
Long sentences Use of personal pronouns Use of emojis No suggestion of the rankings and statistics Orchestration of process Keep Track of the Time	Long sentences Use of personal pronouns Use of emojis Suggestion of the rankings and statistics in addition to orchestrating the decision process Keep Track of the Time

3.4 Measures

3.4.1 Operationalization of Research Variables

The main study constructs and items for both objective and subjective measurements are detailed as follows:

3.4.1.1 Objective Constructs

- Decision-Making Efficiency: Efficiency was measured by the time, in minutes, it took participants to reach a consensus, following the methodology described by Paul et al. (2004). This metric provided an objective measure of how quickly groups could come to a decision.
- Accuracy: Accuracy of decision-making was assessed using an error score, calculated as the sum of the absolute differences between the ranks assigned to the items by marketing experts and by the participants (Hamada et al., 2020). Lower error scores indicated superior performance and greater accuracy in making reasoned judgments.

3.4.1.2 Subjective Constructs

- Process Satisfaction: This was evaluated using Likert scale questionnaire items, following the methodology outlined by Coffeng et al. (2021). Participants rated their satisfaction with the decision-making process on a scale, allowing us to gauge their subjective experience.
- System Satisfaction: Participants rated their satisfaction with the system on a Likert scale, similar to the process satisfaction measure.
- Interpersonal Trust: The choice of questionnaire for interpersonal trust was contextually based. Many sources design items for teams with a history of

working together. In contrast, our source presented items that measured swift trust in teams whose members had no prior familiarity or encounter. This was deemed the most appropriate for our experimental teams, reflecting the conditions of our study where participants did not know each other beforehand.

All survey items were presented to participants via Qualtrics, ensuring a consistent and controlled data collection process.

Table 8. Subjective constructs' measurement items

Survey Items	Scale	Source
System Satisfaction: Using the chatbot makes me feel very satisfied. Using the chatbot gives me a sense of enjoyment. Using the chatbot makes me feel very contented. Using the chatbot makes me feel very delighted.	5-Point Lickert Scale From strongly disagree to strongly agree	Hsu and Lin, 2023
Interpersonal Trust: Team members quickly trust and harmonize with each other. Team members soon have a tacit understanding, easy to communicate with each other. Team members quickly get along and joke with each other. Team members quickly believe that they will cooperate with each other and work carefully.	5-Point Lickert Scale From strongly disagree to strongly agree	Yu et al., 2022
Decision-Making Process Satisfaction: I have the feeling that my group substantiated its decision well. I have the feeling that my group reached its decision in a good manner. I have the feeling that my group reached its decision at a good pace.	7-Point Lickert Scale From strongly disagree to strongly Agree	Coffeng et al., 2021

Chapter 4: Data Analysis and Results

4.1 Descriptive statistics

Below are the global descriptive statistics for all variables measured in this study, including process satisfaction, interpersonal trust, system satisfaction, and the objective constructs.

Table 9. Global descriptive statistics

Variable	Mean	STDEV	Minimum	Maximum
GDM Efficiency*	6.738	2.651	2	13
GDM Accuracy*	4.754	2.481	0.00	12
GDM Process Satisfaction	-0.012	0.704	-2.349	0.988
System Satisfaction	-0.0138	0.550	-1.330	1.119
Interpersonal Trust	-0.004	0.586	-2.117	1.087

*Accuracy is measured based on error score (higher mean= less average accuracy)

*Efficiency is measured based on time to consensus (Higher mean= less average efficiency)

The descriptive statistics for all variables measured in this study are as follows: GDM Efficiency had a mean of 6.738 minutes and a standard deviation of 2.651, with participants taking between 2 and 13 minutes. GDM Accuracy, had a mean of 4.754 and a standard deviation of 2.481, with scores ranging from 0.00 to 12, suggesting varying levels of decision accuracy among groups. GDM Process Satisfaction had a mean of -0.012 and a standard deviation of 0.704, with values between -2.349 and 0.988, indicating mixed satisfaction levels with the decision-making process. System Satisfaction had a mean of -0.0138 and a standard deviation of 0.550, ranging from -1.330 to 1.119, showing moderate satisfaction with the system. Interpersonal Trust had a mean of -0.004 and a standard deviation of 0.586, with scores from -2.117 to 1.087.

Table 10. Pearson Correlations of global results

	GDME	P-Val	GDMA	P-Val	GDMPS	P-Val	SS	P-Val	I T
GDME	1								
GDMA	0.127	0.330	1						
GDMPS	-0.149	0.253	-0.011	0.932	1				
SS	-0.032	0.806	0.068	0.604	0.337^{**}	0.008	1		
IT	-0.138	0.288	-0.214	0.098	0.505^{**}	<0.001	0.223	0.884	1

The study assessed the correlations between various constructs across all experimental conditions. Group Decision-Making Efficiency (GDME), which is time to reach a consensus, showed a positive but non-significant correlation with Group Decision-Making Accuracy (GDMA), which is the error score, and a slight negative correlation with Group Decision-Making Process Satisfaction (GDMPS) and Interpersonal Trust (IT). Notably, GDMPS had a significant positive correlation with System Satisfaction (SS) and IT, indicating that participants who were satisfied with the process were also likely to trust their group and enjoy the system used.

Table 11. Pearson Correlations of robot-like condition and Human-like condition

	GDME	P-Val	GDMA	P-Val	GDMPS	P-Val	SS	P-Val	IT	P-Val
GDME	1.000		0.279	0.136	-0.359	0.051	-0.111	0.560	-0.378^{*}	0.039
GDMA	-0.034	0.856	1.000		-0.015	0.937	0.265	0.156	-0.143	0.451
GDMPS	-0.119	0.522	-0.058	0.758	1.000		0.389^{*}	0.034	0.615^{**}	<0.001
SS	-0.030	0.875	-0.169	0.363	0.260	0.157	1.000		0.209	0.268
IT	0.096	0.607	-0.305	0.095	0.440^{*}	0.013	0.242	0.190	1.000	

Lower correlation matrix corresponds to robot-like condition and upper correlation matrix corresponds to human-like condition

Table 11 presents the Pearson correlation coefficients and their corresponding p-values for subjective variables of the study under two different conditions: "robot-like" and "human-like" chatbot communication styles. GDME (time to consensus) negatively correlated with IT in human-like conditions ($r= -0.378$, $p= 0.039$), suggesting when IT is higher the team will reach a consensus faster in the human-like condition.

Positive correlations between SS and GDMPS across conditions emphasize the role of system satisfaction in overall GDMPS, however, this correlation is only significant in human-like condition (*r*= 0389, *p*= 0.034).

GDMPS showed a strong positive correlation with IT (*r*=0.615, *p*< 0.001) in human-like condition, and a moderate positive correlation with IT (*r*= 0.440, *p*= 0.013) in the robot-like condition, highlighting the importance of trust among team members in their attitude and experience in the process of decision-making.

4.2 Measurement model validation

4.2.1 Constructs' reliability

The initial stage of analysis involves confirming the reliability and validity of the measurement model. Researchers commonly rely on Cronbach's alpha to assess the reliability of items in a scale, where values exceeding .70 are generally considered indicative of an acceptable level of reliability (Salmond, 2008). However, composite reliability score surpasses Cronbach's Alpha in measuring internal consistency by considering item loadings from the theoretical model (Aibinu and Al-Lawati, 2010). Unlike Cronbach's Alpha, which treats all items equally, composite reliability adjusts for their factor loadings, enhancing accuracy (Aibinu and Al-Lawati, 2010). Therefore, it's imperative to assess and report CR in research, with values above .70 deemed acceptable (Hair et al., 2020).

As illustrated in Table 12 our composite reliability values range from 0.890 to 0.964.

Table 12. Constructs' reliability

Construct	Items	Range	Mean	STDEV	Cronbach alpha	Composite reliability (rho-c)
IT	4	1-5	-0.004	0.582	0.838	0.890
SS	4	1-5	-0.014	0.546	0.950	0.964
GDMPS	3	1-7	-0.012	0.699	0.927	0.954

4.2.2 Constructs' validity

To ensure construct validity, it is imperative that the factor loadings of the items surpass a threshold of 0.7, indicating the significance of each item (Coursaris and Van Osch, 2016), as depicted in Table 13 where these loadings are summarized.

Table 13. Matrix of loadings and cross-loadings

	IT	SS	GDMPS
IT 1	**0.838**	0.347	0.409
IT 2	**0.877**	0.310	0.456
IT 3	**0.697**	0.250	0.261
IT 4	**0.854**	0.354	0.473
SS 1	0.343	**0.891**	0.343
SS 2	0.413	**0.940**	0.425
SS 3	0.314	**0.949**	0.403
SS 4	0.375	**0.948**	0.406
GDMPS1	0.480	0.377	**0.925**
GDMPS 2	0.467	0.399	**0.942**
GDMPS 3	0.461	0.415	**0.936**

Convergent validity is confirmed through an examination of the average variance extracted (AVE), which assesses the proportion of variance captured by a construct relative to measurement variance. According to Fornell and Larcker (1981), an AVE score of 0.50 or greater is deemed acceptable. In Table 14, AVE values ranging from 0.671 to 0.873 are displayed. All of these values surpass the acceptable criterion.

Table 14. Constructs' convergent validity and correlation of latent variables

Construct	Convergent Validity (AVE)	IT	SS	GDMPS
IT	0.671	**0.819**		
SS	0.869	0.388	**0.932**	
GDMPS	0.873	0.502	0.425	**0.934**

As Table 14 demonstrates, discriminant validity was confirmed by ensuring that the square root of the variance shared between a construct and its items exceeded the correlations between the said construct and any other constructs in the model (Fornell and Larcker, 1981). Additionally, discriminant validity was supported by observing that all items exhibited strong loadings on their intended factors while demonstrating weaker loadings on unrelated factors, as detailed in Table 14.

4.3 Hypothesis testing

In this section we present the results from the partial least squares (PLS) analysis of hypotheses, a powerful statistical method used in this study. The section starts with an explanation of the principles behind PLS. Then, we provide a detailed analysis of the structural model, highlighting the complex relationships within the research framework.

4.3.1 Statistical technique

Structural Equation Modeling (SEM) consists of two parts: the measurement model and the structural model. The measurement model shows how observed variables relate to underlying concepts, while the structural model explains how these concepts influence each other. Partial Least Squares (PLS), similar to regression, handles both models at once, considering theoretical relationships between concepts and how they're measured.

In recent years, PLS-SEM has been widely embraced across social science disciplines, notably in information systems management (Hair et al., 2019). Its popularity stems from several factors:

1. Small Sample Size Advantage: PLS-SEM is adept at deriving solutions from limited sample sizes, particularly in models with numerous constructs and items (Hair et al., 2019).
2. Distributional Assumptions: PLS-SEM can model latent constructs under nonnormal conditions, crucial for social science studies often reliant on nonnormal data, which can impact results (Hair et al., 2019).
3. Statistical Power: Compared to CB-SEM, PLS-SEM offers greater statistical power, even when estimating common factor model data, thus enhancing the likelihood of identifying significant relationships (Hair et al., 2019).

Given PLS's independence from the assumption of normality and its efficacy with smaller sample sizes, coupled with the exploratory nature of our study, we used SmartPLS version 4 for conducting the analysis, with significance testing completed using the bootstrapping resampling technique included in the package.

Moreover, contrary to techniques like regression, which presume equal weights for all scale indicators, the PLS algorithm adjusts the weight of each indicator based on its contribution to the composite score of the latent variable. Consequently, indicators with weaker relationships to the latent construct are assigned lower weights. This nuanced approach renders PLS preferable in scenarios where measurement errors are presumed, as it accounts for varying degrees of indicator relevance (Wang, 2010).

4.3.2 Structural model

After evaluating the measurement model, the proposed hypotheses were tested employing the Partial Least Squares (PLS) technique within SmartPLS 4.0. To gauge the significance of all paths, we employed a bootstrap resampling procedure with 5000 iterations alongside a one-tailed T-test. In PLS analysis, betas can be interpreted akin to standardized coefficients in multiple regression, indicating the relative strength of the statistical relationships. Our findings, summarized in Figure 4 and in Table 15, offer support for some but not all hypotheses.

The subsequent section delves deeper into the results obtained from the structural equation modeling analysis. Initially, we examine the effects of chatbot conversational style (CCS) and its interplay with the subdimensions of group decision-making performance as well as the two mediators (interpersonal trust and system satisfaction). This is followed by an analysis of the mediation effects. It is noteworthy to know that in this section, we exclusively examined the influence of chatbot conversational style on the dependent constructs, excluding the chatbot's role from the primary research model. This exclusion can be justified as the chatbot construct does not act as a moderator in any of the observed relationships (see appendix B). The subsequent section, titled "Post Hoc Analysis," incorporates the chatbot role manipulation into the analysis and presents its effects comprehensively.

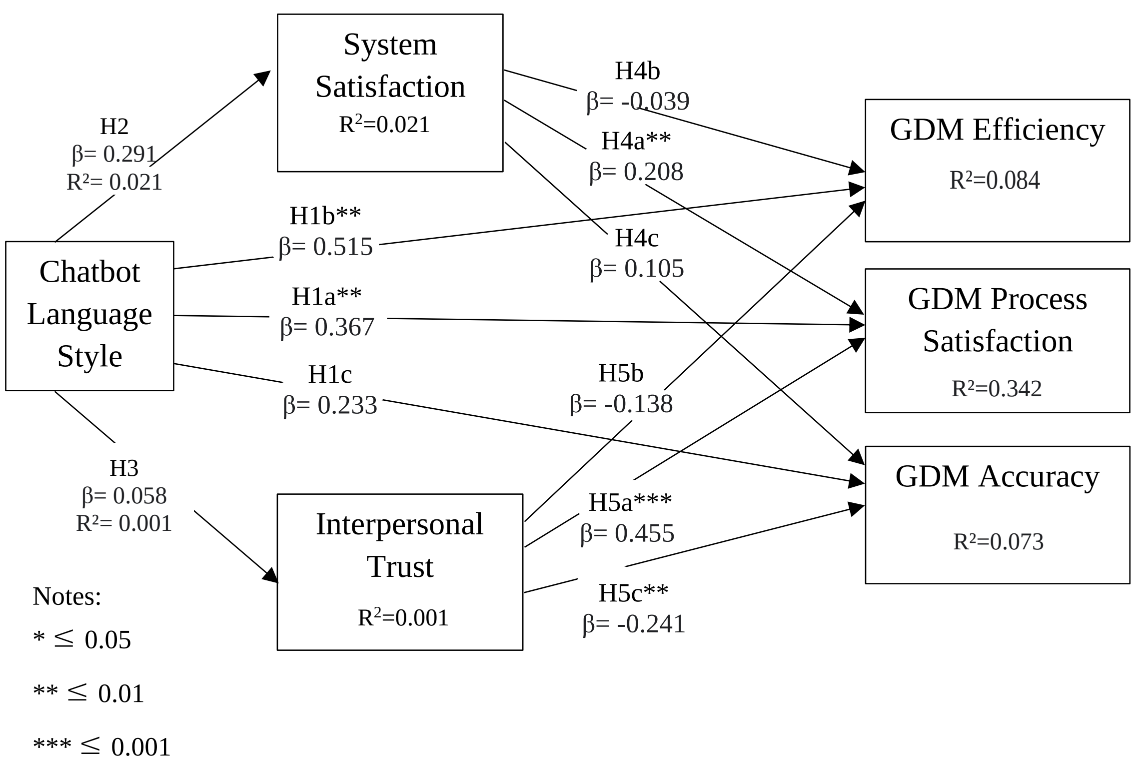

Figure 4. Research model results by level of significance

Figure 4 depicts the results, which are also summarized in Table 15 presented below.

4.3.2.1 Main Hypotheses

The first sets of results pertain to the main hypotheses namely the relationship between the conversational style and GDM process satisfaction, GDM efficiency, GDM accuracy, interpersonal trust and system satisfaction. Also, the relationship between interpersonal trust and system satisfaction and the dependent variables. The results showed that the paths from conversational style to GDM process satisfaction (b = .367, $p < .05$) and conversational style to GDM efficiency (b = .515, $p < .05$) were significant. Hypothesis H1a was confirmed, showing that a more human-like communication style results in higher levels of GDM process satisfaction.

Although the relationship between conversational style and GDM efficiency was significant, unlike Hypothesis H1b, which predicted that a more human-like chatbot language style would increase GDM efficiency, the results showed that a more robot-like communication style resulted in higher GDM efficiency levels. This finding contrasts with our initial prediction, suggesting that the concise and straightforward nature of robot-like communication may facilitate more efficient interactions. This finding will be explored in greater detail in the discussion section.

In contrast, conversational style showed no significant relationship (b = .233, p = 0.181) with GDM accuracy (error score). This rejection of H1c suggests that the communication style of the chatbot does not significantly impact the errors decision teams make during remote brainstorming and decision-making sessions.

The data did not support the paths between conversational style and system satisfaction (H2) and conversational style and interpersonal trust (H3), meaning that neither H2 nor H3 are not supported, as will be further explored in the discussion.

The results regarding the relationship between system satisfaction and GDM process satisfaction shows a significant positive effect (b=.208, $p < .05$), indicating the confirmation of H4a. But system satisfaction does not have a significant impact neither on GDM efficiency nor on GDM accuracy of the team. Thus, our hypotheses H2a and H2b are not supported. Again, these results will be further explored in the discussion section.

Finally, the results show that interpersonal trust among team members has a significant negative effect on the GDM accuracy (measured by error score) (*b* = -.241, $p < .05$), thus supporting hypothesis H5c. This shows that more interpersonal trust among team members can result in more accurate decisions. Consistent with our expectations, the path between interpersonal trust and GDM process satisfaction (*b*= .455, $p < .001$) is also confirmed, thereby supporting our hypothesis number H5a. Interpersonal trust did not show any significant effect on GDM efficiency (*b*= -.138, $p = .137$), meaning that hypothesis H5b is not supported.

Table 15. Hypotheses test results

Hyp.	Path	Path Coeff.	T statistics (|O/STDEV|)	P value	Results
H1a	CCS ->GDMPS	0.367	1.797	0.036	Supported
H1b	CCS-> GDME	0.515	2.066	0.019	Supported
H1c	CCS-> GDMA	0.233	0.911	0.181	Rejected
H2	CCS-> SS	0.291	1.112	0.133	Rejected
H3	CCS -> IT	0.058	0.219	0.413	Rejected
H4a	SS -> GDMPS	0.208	2.03	0.021	Supported
H4b	SS -> GDME	-0.039	0.347	0.364	Rejected
H4c	SS-> GDMA	0.105	0.772	0.22	Rejected
H5a	IT-> GDMPS	0.455	4.49	0	Supported
H5b	IT -> GDME	-0.138	1.096	0.137	Rejected
H5c	IT-> GDMA	-0.241	2	0.023	Supported

4.3.3 Mediating hypotheses

Table 16 presents the results of testing the mediating relationships using the bootstrap method. In recent years, bootstrapping has emerged as the most widely utilized and robust method for assessing mediation models (Takhsha et al., 2020).

Concerning the mediating role of system satisfaction within our research framework, hypotheses H6a, H6b, and H6c postulated that system satisfaction would mediate the associations between conversational style and independent constructs (GDM process satisfaction, GDM efficiency and GDM accuracy). However, our analysis indicates non-significance in these pathways, thus indicating that hypotheses H6a, H6b, and H6c lack empirical support.

Upon observation, it is noted that communication style exerts effects on GDM process satisfaction, DGM efficiency, and GDM accuracy through interpersonal trust (H7a, H7b and H7c), with corresponding values of *b*= 0.026, *b*= -0.008, and *b*= -0.014, respectively. However, it is crucial to highlight that the lower and upper bounds of the confidence intervals include 0. This indicates that the meaningfulness of these relationships is not substantiated by the bootstrap method, suggesting that these effects may not be statistically significant. These findings suggest that our hypotheses (H7a, H7b, and H7c), positing significant relationships between communication style and GDM process satisfaction, GDM efficiency, and GDM accuracy mediated by interpersonal trust, are not supported by the analysis conducted via the bootstrap method.

We did not conduct further mediation tests, as the relationships between Chatbot conversational style and system satisfaction, as well as between Chatbot conversational style and trust, were not significant.

Table 16. Results of mediating effect hypotheses tests

Path	Path Coeff.	5.00%	95.00%	P-value	Results
CCS-> IT-> GDMPS	0.026	-0.186	0.216	0.415	Rejected
CCS-> SS -> GDME	-0.011	-0.093	0.056	0.405	Rejected
CCS-> IT -> GDME	-0.008	-0.081	0.078	0.436	Rejected
CCS-> SS -> GDMA	0.03	-0.048	0.145	0.307	Rejected

CCS-> IT -> GDMA	-0.014	-0.148	0.087	0.423	Rejected
CCS->SS -> GDMPS	0.06	-0.026	0.198	0.202	Rejected

4.3.4 Common Method Bias

The presence of common method bias was evaluated by examining the Variance Inflation Factor (VIF) values within the inner model. In this study, all VIF values were found to be lower than 3.33. Accordingly, following Kock (2015), the model can be deemed free from common method bias.

Table 17. Collinearity statistics (VIF) - Inner model - List

	Variance Inflation Factor (VIF)
CCS -> GDMA	1.022
CCS -> IT	1
CCS -> GDMPS	1.022
CCS -> SS	1
CCS -> GDME	1.022
IT-> GDMA	1.053
IT-> GDMPS	1.053
IT-> GDME	1.053
SS -> GDMA	1.074
SS -> GDMPS	1.074
SS -> GDME	1.074

4.4 Post Hoc Data Analysis

This section includes the post-hoc exploratory analysis of the effect of chatbot roles, ideator versus facilitator, on the decision performance of the team. The chatbot in a facilitator role would solely orchestrate the decision-making process, while in an ideator role, it would proactively provide suggestions for slogans in the first and second tasks. In task 3, it would offer some statistics and indirect points about the rankings of channels (e.g., "I wouldn't rank Facebook among the first three platforms" or "Given this data, I would recommend placing it among the top three priorities").

In contrast to our finding that chatbot conversational style had no significant effect on the accuracy of decision making (b= 0.233, p= 0.181), our results show that the role of the chatbot has a significant effect on GDM accuracy. The b= -1.045 and p= 0.001 shows that when we change the role from facilitator to ideator the error score will decrease meaning that the accuracy is higher when the role of the chatbot functions as an ideator, i.e., the chatbot is an active participator in the process. Additionally, based on the data, the chatbot's role did not significantly impact either group decision-making process satisfaction (b= 0.119, p= 0.375) or group decision-making efficiency (b= 0.078, p= 0.419).

Table 18. Overview of Ideator and Facilitator Chatbot Role Characteristics in Past Studies

Chatbot Role Characteristics	**Resource**
Chatbots as facilitators **Effective Management**: • Manage discussion time. • Ensure balanced contributions from all members. • Organize messages to improve the quality of group chat interactions (Kim et al., 2020).	Kim, S., Eun, J., Oh, C., Suh, B., & Lee, J. (2020, April). Bot in the bunch: Facilitating group chat discussion by improving efficiency and participation with a chatbot. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (pp. 1-13).
Stimulating Idea Elaboration: • Pose questions and prompts to encourage participants to expand on their ideas (Bittner et al., 2019).	Bittner, E. A., Küstermann, G. C., & Tratzky, C. (2019). The facilitator is a bot: Towards a conversational agent for facilitating idea elaboration on idea platforms.

Multiparty Interactions: • Interact with multiple users within a group. • Manage turn-taking, identify the next speaker, and consider the context of the conversation for a smooth flow (Petousi et al., 2022).	Petousi, D., Katifori, V., Roussou, M., & Ioannidis, Y. (2022, October). The dialogue facilitator bot: Reflections on design and evaluation. In *2022 International Conference on Interactive Media, Smart Systems and Emerging Technologies (IMET)* (pp. 1-8). IEEE.

Types of facilitator Interventions:
 	
• **Task Interventions**:
  ○ Guide the group toward achieving their objectives (similar to facilitator chatbots in our study) (Tavanapour et al., 2020).
• **Interactional Interventions**:
  ○ Enhance group dynamics and communication by addressing members' socio-emotional needs (Tavanapour et al., 2020).
	Tavanapour, N., Theodorakopoulos, D., & Bittner, E. A. (2020). A conversational agent as facilitator: Guiding groups through collaboration processes. In *Learning and Collaboration Technologies. Human and Technology Ecosystems: 7th International Conference, LCT 2020, Held as Part of the 22nd HCI International Conference, HCII 2020, Copenhagen, Denmark, July 19–24, 2020, Proceedings, Part II 22* (pp. 108-129). Springer International Publishing.

Chatbots as Ideators **Critical for Innovation**: • Leverage workforce creativity (Goli, 2022).	Goli, S., & de Wit, J. (2022). Intelligent Assistance in Idea Generation.

Imagery and natural language text Generation Capabilities:
 	
• Create realistic images
• Write coherent text (Girotra et al., 2023).
• Examples include:
  ○ OpenAI's ChatGPT matching or surpassing human performance in academic exams and professional certifications (Girotra et al., 2023).
  ○ GitHub Co-Pilot aiding in writing, commenting, and debugging code (Girotra et al., 2023).
  ○ Other models offering professional advice in fields such as medicine and law (Girotra et al., 2023).
	Girotra, K., Meincke, L., Terwiesch, C., & Ulrich, K. T. (2023). Ideas are dimes a dozen: Large language models for idea generation in innovation. *Available at SSRN 4526071*.

Effective Brainstorming:	Makokha, J. (2023). FUTURE AI SYSTEMS: AN AI OUTPERFORMS A HUMAN

• Teams using AI-powered chatbots generate more ideas compared to face-to-face interactions (Makokha, 2023).	AS TEAMMATE. *Authorea Preprints*.

Enhance Human Decision-Making Performance:
 	
• Results of a study on consultants' performance using AI assistance:
  ○ 12.2% increase in productivity.
  ○ 25.1% increase in efficiency of task completion.
  ○ Over 40% increase in quality of consultations compared to those without AI assistance (Dell'Acqua et al., 2023).
	Dell'Acqua, F., McFowland, E., Mollick, E. R., Lifshitz-Assaf, H., Kellogg, K., Rajendran, S., ... & Lakhani, K. R. (2023). Navigating the jagged technological frontier: Field experimental evidence of the effects of AI on knowledge worker productivity and quality. *Harvard Business School Technology & Operations Mgt. Unit Working Paper*, (24-013).

Table 19. Post hoc test results

Path	Original sample (O)	Sample mean (M)	STDEV	T statistics	P-values	results
Role -> GDMA	-1.045	-1.064	0.328	3.186	0.001	Supported
Role -> IT	0.366	0.387	0.335	1.094	0.137	Rejected
Role -> GDMPS	0.119	0.131	0.374	0.319	0.375	Rejected
Role -> SS	0.093	0.101	0.329	0.282	0.389	Rejected
Role -> GDME	0.078	0.075	0.383	0.204	0.419	Rejected

Appendix B includes the results of a further post-hoc analysis testing the moderation effects of chatbot conversational style and chatbot role to rule out any interaction effect that could have explained our findings. The post-hoc analysis reveals no significant interaction effects thus offering further evidence that significant main effects for chatbot conversational style or for role are indeed due to these two chatbot characteristics independently and not their interaction.

Chapter 5: Discussion

This study was conducted to explore and answer a central research question: *What is the effect of the communication style of chatbots on the decision-making performance of remote /distributed work teams?*

Additionally, a post-hoc exploratory analysis was performed to address a secondary question: *What is the effect of the chatbot's role on the decision-making performance of remote/distributed work teams?*

The findings revealed that the design of a chatbot's linguistic and communication style can influence certain aspects of decision-making performance in remote teams, as detailed below:

Satisfaction with the Decision-Making Process: when a chatbot's communication style is more human-like rather than robotic, users tend to feel more satisfied with the decision-making process. This can be explained by previous studies showing that users often perceive AI agents as human and interact with them accordingly (Pelau et al., 2021). Research also suggests that people tend to mirror the language style of those they interact with (Ireland and Henderson, 2014). Therefore, richer media, which allows for more verbal and non-verbal cues through natural, human-like language and the use of emojis, can lead to more effective team interactions and greater satisfaction (Hambley et al., 2007).

Efficiency of Decision-Making: The data indicated that a more robotic communication style led to higher efficiency levels. Although this study initially hypothesized a different outcome, this finding aligns with past studies which suggested that shorter, clearer sentences are less time-consuming and easier to understand (Booher, 2001). Thus, the machine-like chatbot may have been less disruptive to the team's activities and allowed them to focus more on the task and complete it faster, particularly considering the short timeframe allotted for each of the tasks. This aligns with prior studies which have showed that the use of emojis (which was a characteristic of human-like chatbot), can be distracting in text-based communication (Lindberg and Kindberg, 2018). Additionally, past studies showed that task-oriented chatbots, which provide only essential information without redundant language, are more efficient than social-oriented chatbots (Lu et al., 2024).

Accuracy of Decision-Making: This study did not find strong evidence that teams using a human-like chatbot achieved higher decision accuracy than those using a robot-like chatbot. However, the p-value of 0.181 suggests that the lack of significance may be due to a small sample size. This indicates a need for further research in future. Most previous studies have focused on the accuracy of chatbot responses themselves, particularly in fields like medicine (e.g., Goodman et al., 2023) rather than the task accuracy of those interacting with the chatbot. In relation to decision accuracy of users, one study examined the relationship between decision accuracy and motivational needs. They found that competency and autonomy significantly affect accuracy more than relatedness (de Vreede et al., 2021).

While the study did not find a significant difference in decision accuracy between human-like and robot-like chatbots, post-hoc analysis revealed that the role of the chatbot significantly impacts accuracy. Specifically, teams interacting with an ideator chatbot, which actively suggests ideas, achieved higher decision accuracy compared to those using a facilitator chatbot. These results align with previous studies, such as Joosten et al. (2024), which compared ideas generated by human professionals with those produced by an AI system. A blind expert evaluation showed that AI-generated ideas scored significantly higher in novelty and customer benefit, while their feasibility scores were comparable to human-generated ideas. This indicates that teams using an idea-generator chatbot can produce better ideas and make more creative and accurate decisions.

System Satisfaction: We assessed the relationship between the chatbot's conversational style and system satisfaction. The data analysis did not allow us to conclude that a more human-like conversational style increases user satisfaction. This finding contrasts with studies that have suggested a more anthropomorphized chatbot leads to higher user satisfaction (Rhim et al., 2022). However, other research has indicated that the anthropomorphism of a chatbot's language style does not always enhance user experience (Jenneboer, 2022). It is noteworthy that, in our study the *p-value* for this relationship was relatively low (p = 0.133), indicating the need for further research as the lack of significance may be due to the small sample size, and a larger sample might reveal a significant relationship.

Among all aspects of group decision performance, system satisfaction only affected group decision-making process satisfaction. This outcome was anticipated, as a well-designed GDSS can enhance participants' satisfaction from group decision-making and brainstorming sessions, leading to perceptions of better performance (Fan and Shen, 2011). However, contrary to our expectations, system satisfaction did not significantly affect decision accuracy or efficiency, i.e., actual performance. Previous studies have shown that user satisfaction generally improves performance (Sharabati et al, 2015), but this was not supported across all three sub-dimensions of decision performance in our research. This suggests that while system satisfaction boosts process satisfaction, i.e., perceptions of performance, it does not necessarily translate to improvements in actual performance, that is, accuracy and efficiency. One possible explanation for this is that system satisfaction may influence how enjoyable or smooth the decision-making process feels, but it doesn't directly impact the cognitive and analytical processes required for accurate and efficient decision-making.
For example, participants might feel more satisfied with the process because the system is user-friendly or engaging, but this satisfaction does not necessarily make the team more effective at analyzing information or making quicker decisions. Actual performance likely depends more on factors such as the team's ability to understand the task, the clarity of the information provided, and the individual members' cognitive capacities. These factors are not necessarily enhanced by a positive user experience with the system.

Moreover, the significance of the relationship between system satisfaction and process satisfaction suggests that system satisfaction is crucial in shaping users' perceptions of the decision-making process, possibly by making it smoother or more collaborative.

However, this result also indicates that while system satisfaction is an important factor in the context of this study and for increasing process satisfaction, it is not affected by the chatbot language style. This finding implies that the significant relationships between language style, GDM efficiency, and GDM process satisfaction are direct and are not mediated, either fully or partially, by system satisfaction.

Interpersonal Trust: This study further examined the relationship between the chatbot's communication style and interpersonal trust among team members. However, the analysis did not reveal a significant link between these two factors, which was unexpected given that previous research suggests machine teammates can influence trust dynamics (Seeber et al., 2020). Past studies have indicated that poor communication can undermine trust, while effective communication is crucial for building and maintaining it (Savolainen et al., 2014). We argued that a human-like communication style, with more detailed and emotionally expressive language including emojis, would improve communication quality and increase interpersonal trust. However, our results did not support this hypothesis. This highlights the need for further research on factors affecting interpersonal trust in human-chatbot interactions, specifically in the context of remote teams.

Our findings did show a significant association between interpersonal trust and decision-making accuracy. This is consistent with past research which showed that interpersonal trust positively impacts job performance and collaboration (Trejo, 2021; Paul and McDaniel, 2004). One explanation for this result is that trust facilitates learning and innovation, which are crucial for addressing complex decision problems (Paul and McDaniel, 2004).

Additionally, interpersonal trust had a direct positive relationship with the decision-making process satisfaction. Previous research supports this link, by showing that higher trust levels improve both satisfaction and the quality of team decisions (Sapp et al., 2019; Driscoll, 1987). Thus, trust enhances satisfaction with participation and overall decision-making.

Despite these findings, trust did not significantly affect decision-making efficiency, contrary to other studies which suggested that trust, especially reliability-based trust, enhances group efficiency (Cheng et al., 2021). This discrepancy may be due to differences in the experimental context. Unlike other studies involving long-term interactions with shared goals, our participants were unfamiliar with each other and interacted only through chat, without information about each other's education, competencies, or past performance. These factors likely limited trust development, which probably affected its impact on efficiency.

Regarding the mediating effect of interpersonal trust, we did not find significant mediation between the chatbot's language style and GDM process satisfaction. However, the mediation effect on efficiency and accuracy was marginally significant, suggesting that further research with a larger sample size is needed.

Chapter 6: Conclusion

Chatbots have become increasingly prevalent in today's digital workplace and are expected to become even more important for decision-making in virtual remote teams. Their ability to facilitate communication and streamline workflows can enhance the decision-making performance. By providing real-time assistance and automating routine tasks, chatbots support collaborative efforts, making them invaluable for modern organizations aiming to optimize their decision-making capabilities. Furthermore, chatbots can collect, process, and analyze large volumes of data, offering summaries, trend analyses, and predictive insights that are crucial for making evidence-based decisions. As the digital environment continues to evolve, the role of chatbots in enhancing organizational decision-making will only become more and more prevalent and critical. Thus, it's important to investigate the ways we can improve the design of these chatbots to improve user experiences and performance when using such AI tools.

In this study, we conducted a scenario-based experiment to investigate the impact of language and communication style of chatbots on the decision performance of virtual teams. The results yielded diverse outcomes across various performance aspects. Notably, a more human-like language employed by the chatbot correlated with higher levels of decision process satisfaction, which holds significance for companies as it fosters greater commitment to the decisions made. Conversely, a robot-like language style was associated with increased efficiency and reduced consensus time, particularly advantageous in environments requiring swift decision-making. This style omitted extraneous words and emotional cues, delivering task-related messages succinctly.

Although the chatbot's language style did not influence decision accuracy, the role of the chatbot served as a crucial determinant. Findings indicated that an ideator chatbot, by acting as a peer and introducing novel ideas and providing task-related information, enhanced decision accuracy in an engaging manner, as shown by the results and highlighted by participants' positive feedback. As an example, one participant commented, "I liked when it [Ideabot] introduced new ideas when we were stuck. I also liked that it was good at giving clear directions."

In contrast, the facilitator role did not exhibit significant associations with efficiency, accuracy, or process satisfaction. However, participants recognized its utility in guiding the decision-making process and ensuring smooth communication flow. Participants remarked, "It reminds us when time's up. It kind of guide us through the process," and "It guided the conversation along smoothly and it made it easy to work with the other people in the room."

Regarding system satisfaction, no discernible difference was observed between human-like and robot-like language styles, suggesting that other chatbot features, such as information accuracy and relevance, may exert greater influence on user satisfaction. However, the p-value of 0.133 indicates that with a larger sample size, the relationship between chatbot language style and system satisfaction could potentially reach significance. This finding suggests that it could be worthwhile to conduct further research to explore whether chatbot language style might indeed affect system satisfaction in the future.

Additionally, interpersonal trust was measured to underscore the human aspect of the experiment. Although no direct effect of language style on interpersonal trust was found, the observed correlation between interpersonal trust and performance underscores its pivotal role in workplace dynamics, warranting heightened attention in today's evolving work landscape where trust is paramount for effective collective collaboration.

6.1 Theoretical and Managerial Contributions

6.1.1 Theoretical Contributions

This study makes several theoretical contributions to the fields of communication, human-computer interaction, and organizational behavior by exploring the impact of chatbot language and communication styles on decision-making performance in remote team collaboration.

The findings suggest strategies for reducing the Uncanny Valley effect. The negative effects of uncanny valley can be controlled if the human-likeness is designed in a manner that can positively impact the performance of the user. For example, using a human-like communication style can improve process satisfaction, while a robotic style can enhance efficiency. This indicates that AI agents can be designed to balance these effects, making them more effective and less unsettling.

Moreover, the findings contribute to media richness theory by demonstrating that a more human-like communication style in chatbots, which includes natural language and emojis, enhances satisfaction with the decision-making process. This supports the notion that richer media, which can convey more verbal and non-verbal cues, leads to more effective team interactions and higher satisfaction levels among team members. These insights align with existing literature emphasizing the importance of media richness for communication effectiveness and satisfaction in remote teams (Hambley et al., 2007; Paul et al., 2004).

The study's findings regarding interpersonal trust add to the body of knowledge by showing a significant positive relationship between interpersonal trust and both decision accuracy and decision process satisfaction. This aligns with existing literature on the crucial role of trust in effective collaboration and job performance (Trejo, 2021; Paul and McDaniel, 2004). However, the study also found that interpersonal trust did not significantly influence decision efficiency, contrasting with some previous studies (Cheng et al., 2021). These mixed results suggest that while trust is essential for certain decision performance aspects, its impact on efficiency may be context-specific. This indicates that further investigation is needed on interpersonal trust in remote teams. Furthermore, by revealing that although trust is a significant antecedent of decision-making performance, it is only limitedly affected by chatbot design characteristics, we highlight that trust dynamics in remote teams may be more dependent on the dynamics between human members of a team rather than the interaction with AI agents.

The research contributes to the understanding of system satisfaction by indicating that chatbot language style alone does not significantly impact overall user satisfaction. This finding challenges the assumption that anthropomorphized chatbots always enhance user experience and suggests that other factors, such as information accuracy and relevance, may play a more critical role. This insight aligns with mixed findings in previous studies and underscores the complexity of factors influencing user satisfaction with chatbots (Rhim et al., 2022; Jenneboer, 2022).

Finally, the post-hoc results introduces the idea that the specific role of a chatbot (e.g., ideator vs. facilitator) significantly impacts decision performance, particularly decision accuracy. The ideator chatbot, by generating novel ideas and offering task-related information, was found to improve decision accuracy. This observation complies with previous studies on chatbots which indicated that AI-generated ideas outperformed human-generated ideas in terms of novelty and customer benefit (Joosten et al., 2024). This finding further highlights the potential for role-specific chatbot functionality to enhance team decision-making, suggesting that different chatbot roles can be strategically deployed to support distinct aspects of the decision-making process.

6.1.2 Managerial Implications

It is essential for designers to recognize that the language style of a chatbot is a critical design element that significantly influences user performance. However, it is not advisable to assume that one language style is universally superior. The choice of language style should be strategically aligned with the specific purpose of the chatbot and the needs of its users.

For tasks that require swift decision-making, a robotic language style proves to be more effective, as it enhances efficiency and reduces the time needed to reach a consensus. On the other hand, in scenarios where the quality of communication and the expression of emotional and social cues are crucial for achieving higher process satisfaction and decision commitment, a more human-like chatbot is more suitable.

This nuanced approach to chatbot design ensures that the tool effectively supports diverse decision-making contexts within remote teams. By tailoring the language style of chatbots to the specific needs of different tasks, designers can optimize both the efficiency and satisfaction of the users, leading to better overall performance and more effective decision-making processes. A similar nuanced approach might be needed with other elements of chatbot design, as our exploratory findings about chatbot roles indicate.

6.2 Limitations and Future Research

This study has some limitations which are discussed below.

- Generalizability: The use of a specific demographic (participants aged 18-64 with a master’s degree or higher) recruited from Prolific may limit the generalizability of the findings. The sample may not represent the broader population, particularly those with lower educational levels or from different cultural backgrounds who are working in organizations and using these online platforms to collaborate remotely.
- Wizard of Oz Technique: The application of the Wizard of Oz technique may introduce biases. The human operator may make errors, such as incorrect timing of the messages, which would not occur with an actual chatbot. Additionally, the use of pre-made scripts for chatbot interactions may limit the chatbot's ability to adapt dynamically to the flow of conversation. This could affect the naturalness and effectiveness of the communication, potentially impacting the participants' experience and the study's outcomes.
- Measurement of Interpersonal Trust: The measurement of interpersonal trust using a swift trust questionnaire designed for teams with no prior familiarity may not fully capture the depth and complexity of trust that develops over time in real-world settings.
- Task Specificity: The tasks used in the experiment (brainstorming slogans, selecting a slogan, and ranking social media platforms) are specific and may not encompass the full range of decision-making scenarios encountered in different organizational contexts. This may limit the applicability of the findings to other types of tasks.
- Platform-Specific Constraints: Conducting the study on the Chatzy platform may introduce platform-specific constraints and user interface limitations that could influence participants' interactions and overall experience. These constraints might not be present in other collaboration tools used in real-world settings. Additionally, some participants might rate the design of the platform itself rather than the chatbot's behavior when rating their satisfaction with the system. By acknowledging these limitations, we can better contextualize the findings and identify areas for improvement in future research.

Based on the limitations of this study and the constructs we measured, we propose several avenues for future research to address these issues and enhance our understanding of factors that influence users' experiences with chatbot interactions in remote team settings:

- Diverse Sample Populations: Future studies should include a more diverse sample population in terms of educational background, age, and cultural context. This will help assess the generalizability of the findings across different demographic groups and provide insights into how chatbot communication styles and roles affect a broader range of users.
- Real Chatbots: We propose that future studies utilize actual chatbots rather than relying on the Wizard of Oz technique to eliminate potential biases introduced by human operators. This would provide a more accurate assessment of chatbot effectiveness and interactions. Additionally, employing chatbots with adaptive capabilities could enhance the naturalness and effectiveness of communication.
- Longitudinal Trust Measurement: Future studies can consider using longitudinal approaches to measure interpersonal trust, capturing how trust develops and evolves over time in real-world settings. This can provide a deeper understanding of trust dynamics and its impact on team interactions and performance.
- Impact of Chatbot Roles on Cognitive Load: Future studies can examine how different chatbot roles affect cognitive load and whether certain roles can reduce cognitive load more effectively while also enhancing overall performance. This could involve employing advanced techniques such as real-time cognitive load monitoring, physiological measures, and detailed user feedback to assess the impact of each role on mental effort and task effectiveness.

By addressing these areas, future research can build on the current study's findings and provide a more nuanced understanding of chatbot communication styles and roles and their impact on remote team dynamics and performance.

Bibliography

Adam, M., Wessel, M., & Benlian, A. (2021). AI-based chatbots in customer service and their effects on user compliance. *Electronic Markets*, 31(2), 427-445.

Adamopoulou, E., & Moussiades, L. (2020). An overview of chatbot technology. In *IFIP international conference on artificial intelligence applications and innovations* (pp. 373-383). Springer, Cham.

Addy, W. A., Ajayi-Nifise, A. O., Bello, B. G., Tula, S. T., Odeyemi, O., & Falaiye, T. (2024). Machine learning in financial markets: A critical review of algorithmic trading and risk management. *International Journal of Science and Research Archive*, 11(1), 1853-1862.

Aibinu, A. A., & Al-Lawati, A. M. (2010). Using PLS-SEM technique to model construction organizations' willingness to participate in e-bidding. *Automation in construction*, 19(6), 714-724.

Alper, S., Tjosvold, D., & Law, K. S. (1998). Interdependence and controversy in group decision making: Antecedents to effective self-managing teams. *Organizational behavior and human decision processes*, 74(1), 33-52.

Alves, M. P., Dimas, I. D., Lourenço, P. R., Rebelo, T., Peñarroja, V., & Gamero, N. (2023). Can virtuality be protective of team trust? Conflict and effectiveness in hybrid teams. *Behaviour & Information Technology*, 42(7), 851-868.

Awan, S. H., Habib, N., Shoaib Akhtar, C., & Naveed, S. (2020). Effectiveness of performance management system for employee performance through engagement. *SAGE open*, 10(4), 2158244020969383.

Bales, R. F., & Strodtbeck, F. L. (1951). Phases in group problem-solving. *The Journal of Abnormal and Social Psychology*, 46(4), 485–495.

Beattie, A., Edwards, A. P., & Edwards, C. (2020). A bot and a smile: Interpersonal impressions of chatbots and humans using emoji in computer-mediated communication. *Communication Studies*, 71(3), 409-427.

Blank, J. D., & Osofsky, L. (2020). Automated legal guidance. *Cornell L. Rev*., 106, 179.

Booher, D. (2001). *E-writing: 21st-century tools for effective communication*. Simon and Schuster.

Borau, S., Otterbring, T., Laporte, S., & Fosso Wamba, S. (2021). The most human bot: Female gendering increases humanness perceptions of bots and acceptance of AI. *Psychology & Marketing*, 38(7), 1052-1068.

Borum, R., (2010)"The Science of Interpersonal Trust". *Mental Health Law & Policy Faculty Publications*. Paper 574.

Boukhayma, K., Ben Hiba, L., & Elmanouar, A. (2019, March). DSS success: The intertwining of satisfaction and learning. *In Proceedings of the ArabWIC 6th Annual International Conference Research Track* (pp. 1-6).

Bucăţa, G., & Rizescu, A. M. (2017). The role of communication in enhancing work effectiveness of an organization. *Land Forces Academy Review*, 22(1), 49-57.

Candello, H., Pinhanez, C., & Figueiredo, F. (2017). Typefaces and the perception of humanness in natural language chatbots. *In Proceedings of the 2017 chi conference on human factors in computing systems* (pp. 3476-3487).

Cao, L. (2022). Ai in finance: challenges, techniques, and opportunities. *ACM Computing Surveys (CSUR)*, 55(3), 1-38.

Cao, X., Gong, M., Yu, L., & Dai, B. (2020). Exploring the mechanism of social media addiction: An empirical study from WeChat users. *Internet Research*, 30(4), 1305-1328.

Chang, H. H., Hung, C. J., & Hsieh, H. W. (2014). Virtual teams: Cultural adaptation, communication quality, and interpersonal trust. *Total Quality Management & Business Excellence*, 25(11-12), 1318-1335.

Chaves, A. P., & Gerosa, M. A. (2021). How should my chatbot interact? A survey on social characteristics in human–chatbot interaction design. *International Journal of Human–Computer Interaction*, 37(8), 729-758.

Cheng, X., Bao, Y., Yu, X., & Shen, Y. (2021). Trust and group efficiency in multinational virtual team collaboration: A longitudinal study. *Group Decision and Negotiation*, 30, 529-551.

Cheng, X., Zhang, X., Cohen, J., & Mou, J. (2022). Human vs. AI: Understanding the impact of anthropomorphism on consumer response to chatbots from the perspective of trust and relationship norms. *Information Processing & Management*, 59(3), 102940.

Chudzicka-Czupała, A., Grabowski, D., Agnieszka, C., & Stasiła-Sieradzka, M. (2022). Assessment of worklife areas, trust in supervisor and interpersonal trust as conditions of severity of stress at work. *International Journal of Occupational Medicine and Environmental Health*, 35(6), 719.

Ciechanowski, L., Przegalinska, A., Magnuski, M., & Gloor, P. (2019). In the shades of the uncanny valley: An experimental study of human–chatbot interaction. *Future Generation Computer Systems*, 92, 539-548.

Coffeng, T., Van Steenbergen, E. F., De Vries, F., & Ellemers, N. (2021). Quality of group decisions by board members: a hidden-profile experiment. *Management Decision*, 59(13), 38-55.

Cohen, L., & Kassis-Henderson, J. (2012). Language use in establishing rapport and building relations: Implications for international teams and management education. *Revue Management et Avenir*, (5), 185-207.

Costa, J., Jung, M. F., Czerwinski, M., Guimbretière, F., Le, T., & Choudhury, T. (2018). Regulating feelings during interpersonal conflicts by changing voice self-perception. *In*

Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, 1-13.

Coursaris, C. K., & Van Osch, W. (2016). A Cognitive-Affective Model of Perceived User Satisfaction (CAMPUS): The complementary effects and interdependence of usability and aesthetics in IS design. *Information & Management*, 53(2), 252-264.

Crow, J. J. (2006). *Examining cognitive processes of unstructured decision making*. Kansas State University.

De Vreede, T., Raghavan, M., and De Vreede, G. J. (2021). Design foundations for AI assisted decision making: A self determination theory approach. *in Proceedings of the 54th Hawaii International Conference on System Sciences*. January 4–9, 2021; 166–175.

Delbecq, A. L., & Van de Ven, A. H. (1971). A group process model for problem identification and program planning. *The journal of applied behavioral science*, 7(4), 466-492.

Delic, A., Emamgholizadeh, H., & Ricci, F. (2023, June). CHARM: A Group Recommender ChatBot. *In Adjunct Proceedings of the 31st ACM Conference on User Modeling, Adaptation and Personalization* (pp. 275-282).

Dell'Acqua, F., McFowland, E., Mollick, E. R., Lifshitz-Assaf, H., Kellogg, K., Rajendran, S., ... & Lakhani, K. R. (2023). Navigating the jagged technological frontier: Field experimental evidence of the effects of AI on knowledge worker productivity and quality. *Harvard Business School Technology & Operations Mgt*. *Unit Working Paper*, (24-013).

Dewey, J. (1933). How We Think. A Restatement of the Relation of Reflective Thinking to the Educative Process, Boston etc.(DC Heath and Company) 1933.

Dillion, D., Tandon, N., Gu, Y., & Gray, K. (2023). Can AI language models replace human participants?. *Trends in Cognitive Sciences*, *27*(7), 597-600.

Dirks, K. T. (1999). The effects of interpersonal trust on work group performance. *Journal of applied psychology*, 84(3), 445.

Driscoll, J. W. (1978). Trust and participation in organizational decision making as predictors of satisfaction. *Academy of management journal*, 21(1), 44-56.

Duan, Y., Edwards, J. S., & Dwivedi, Y. K. (2019). Artificial intelligence for decision making in the era of Big Data–evolution, challenges and research agenda. *International journal of information management*, 48, 63-71.

Duffy, B. R. (2003) 'Anthropomorphism and the social robot', *Robotics and Autonomous Systems*, 42(3–4), pp. 177–190. doi: 10.1016/S0921-8890(02)00374-3.

Eberhart, Z., Bansal, A., & Mcmillan, C. (2020). A wizard of oz study simulating api usage dialogues with a virtual assistant. *IEEE Transactions on Software Engineering*, *48*(6), 1883-1904.

Eden, C. (1982), "Problem construction and the influence of OR", *Interfaces*, 12/2, 50-60.

El Yousfi, F., & Anand, A. (2021). Investigating the factors affecting virtual team performance in public organizations. *Management international*, 25(6), 264-283.

El-Shinnawy, M., and Markus, M. Media richness theory and new electronic communication media: A study of voice mail and electronic mail. In J.I. DeGross, J.D. Becker, and J.J. Elam (eds.), *Proceedings of the Thirteenth International Conference on Information Systems*. Atlanta: AIS, 1992, pp. 91-105.

Epley, N., Waytz, A., & Cacioppo, J. T. (2007). On seeing human: a three-factor theory of anthropomorphism. *Psychological review*, 114(4), 864.

Fan, S., & Shen, Q. (2011). The effect of using group decision support systems in value management studies: An experimental study in Hong Kong. *International Journal of Project Management*, 29(1), 13-25.

Fisher, B. A. (1970). Decision emergence: Phases in group decision-making. *Communications Monographs*, 37(1), 53-66.

Fleischmann, A. C., Aritz, J., & Cardon, P. (2020). Language proficiency and media richness in global virtual teams: Impacts on satisfaction, inclusion, and task accomplishment. *ACM Transactions on Social Computing*, 2(4), 1-18.

Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of marketing research*, 18(1), 39-50.

Franco, L. A., & Rouwette, E. A. (2011). Decision development in facilitated modelling workshops. *European Journal of Operational Research*, 212(1), 164-178.

Gefen, D., Fresneda, J. E., & Larsen, K. R. (2020). Trust and distrust as artifacts of language: a latent semantic approach to studying their linguistic correlates. *Frontiers in psychology*, 11, 561.

Geller, E. S. (1999). Interpersonal trust. *Professional Safety*, 44(4), 16.

Gersick, C. J. (1988). Time and transition in work teams: Toward a new model of group development. *Academy of Management journal*, 31(1), 9-41.

Gimpel, H., J. Huber, and S. Sarikaya (2016). "Customer Satisfaction in Digital Service Encounters: The Role of Media Richness, Social presence, and Cultural Distance." *In: Proceedings of the 24th European Conference on Information Systems (ECIS 2016)*. June 12-15. Istanbul, Turkey.

Girotra, K., Meincke, L., Terwiesch, C., & Ulrich, K. T. (2023). Ideas are dimes a dozen: Large language models for idea generation in innovation. Available at SSRN 4526071.

Go, E., & Sundar, S. S. (2019). Humanizing chatbots: The effects of visual, identity and conversational cues on humanness perceptions. *Computers in human behavior*, 97, 304-316.

Goli, S., & de Wit, J. (2022). *Intelligent Assistance in Idea Generation*. Master's thesis.

Goodman, R. S., Patrinely, J. R., Stone, C. A., Zimmerman, E., Donald, R. R., Chang, S. S., ... & Johnson, D. B. (2023). Accuracy and reliability of chatbot responses to physician questions. *JAMA network open*, 6(10), e2336483-e2336483.

Haick, H., & Tang, N. (2021). Artificial intelligence in medical sensors for clinical decisions. *ACS nano*, 15(3), 3557-3567.

Hair Jr, J. F., Howard, M. C., & Nitzl, C. (2020). Assessing measurement model quality in PLS-SEM using confirmatory composite analysis. *Journal of business research*, 109, 101-110.

Hair, J. F., Risher, J. J., Sarstedt, M., & Ringle, C. M. (2019). When to use and how to report the results of PLS-SEM. *European business review*, 31(1), 2-24.

Halvorsen, K. (2018). Questions as interactional resource in team decision making. *International Journal of Business Communication*, 55(1), 69-93.

Hamada, D., Nakayama, M., & Saiki, J. (2020). Wisdom of crowds and collective decision-making in a survival situation with complex information integration. *Cognitive Research: Principles and Implications*, 5, 1-15.

Hambley, L. A., O'Neill, T. A., & Kline, T. J. (2007). Virtual team leadership: The effects of leadership style and communication medium on team interaction styles and outcomes. *Organizational behavior and human decision processes*, 103(1), 1-20.

Han, M. C. (2021). The impact of anthropomorphism on consumers' purchase decision in chatbot commerce. *Journal of Internet Commerce*, 20(1), 46-65.

Hartson, R., & Pyla, P. S. (2012). *The UX Book: Process and guidelines for ensuring a quality user experience*. Elsevier.

Hedlund, J., Ilgen, D. R., & Hollenbeck, J. R. (1998). Decision accuracy in computer-mediated versus face-to-face decision-making teams. *Organizational Behavior and Human Decision Processes*, 76(1), 30-47.

Herrera, F., Herrera-Viedma, E., & Verdegay, J. L. (1995). A sequential selection process in group decision making with a linguistic assessment approach. *Information Sciences*, 85(4), 223-239.

Hill, J., Ford, W. R., & Farreras, I. G. (2015). Real conversations with artificial intelligence: A comparison between human–human online conversations and human–chatbot conversations. *Computers in human behavior*, 49, 245-250.

Hohenstein, J., Kizilcec, R. F., DiFranzo, D., Aghajari, Z., Mieczkowski, H., Levy, K., ... & Jung, M. F. (2023). Artificial intelligence in communication impacts language and social relationships. *Scientific Reports*, 13(1), 5487.

Hoti, A. H., Zenuni, X., Hamiti, M., & Ajdari, J. (2023, June). Student performance prediction using AI and ML: state of the art. *In 2023 12th Mediterranean Conference on Embedded Computing (MECO)* (pp. 1-6). IEEE.

Hsu, C. L., & Lin, J. C. C. (2023). Understanding the user satisfaction and loyalty of customer service chatbots. *Journal of Retailing and Consumer Services*, 71, 103211.

Hsu, W. C. and Lee, M. H. (2023). Semantic Technology and Anthropomorphism: Exploring the Impacts of Voice Assistant Personality on User Trust, Perceived Risk, and Attitude. *Journal of Global Information Management*. 31. 1-21. 10.4018/JGIM.318661.

Hélie, S., & Sun, R. (2010). Incubation, insight, and creative problem solving: a unified theory and a connectionist model. *Psychological review*, 117(3), 994.

Ireland, M. E., & Henderson, M. D. (2014). Language style matching, engagement, and impasse in negotiations. *Negotiation and conflict management research*, 7(1), 1-16.

Iselin, E. (1989). The impact of information diversity on information overload effects in unstructured managerial decision making. *Journal of Information Science*, 15(3), 163-173.

Jakobsen, A. (2021). *Humanlike Customer Service Chatbots: The Effect of Humanness on User Experience and User Behavior* (Master's thesis).

Jarrahi, M. H. (2018). Artificial intelligence and the future of work: Human-AI symbiosis in organizational decision making. *Business horizons*, 61(4), 577-586.

Jenneboer, L. (2022). *Interaction effects between anthropomorphic chatbot characteristics on the customer experience: appearance, language style, and emoji-use* (Master's thesis, University of Twente).

Jo, H., & Park, D. H. (2024). Effects of ChatGPT’s AI capabilities and human-like traits on spreading information in work environments. *Scientific Reports*, 14(1), 7806.

Joosten, J., Bilgram, V., Hahn, A., & Totzek, D. (2024). Comparing the ideation quality of humans with generative artificial intelligence. *IEEE Engineering Management Review*.

Jucks, R., Linnemann, G. A., Thon, F. M., & Zimmermann, M. (2016). Trust the words: insights into the role of language in trust building in a digitalized world. *Trust and communication in a digitized world: Models and concepts of trust research*, 225-237.

Kahai, S. S., & Cooper, R. B. (2003). Exploring the core concepts of media richness theory: The impact of cue multiplicity and feedback immediacy on decision quality. *Journal of management information systems*, 20(1), 263-299.

Kaliardos, W. N. (1999). *Semi-structured decision processes: A conceptual framework for understanding human-automation systems* (Doctoral dissertation, Massachusetts Institute of Technology).

Kapo, A., Turulja, L., Zaimović, T., & Mehić, S. (2021). Examining the effect of user satisfaction and business intelligence system usage on individual job performance. *Management: Journal of Contemporary Management Issues*, 26(2), 43-62.

Karlinsky-Shichor, Y., & Zviran, M. (2016). Factors influencing perceived benefits and user satisfaction in knowledge management systems. *Information Systems Management*, 33(1), 55-73.

Khaoula BOUKHAYMA, Lamia BENHIBA and Abdellah ELMANOUAR.2019. DSS success: The intertwining of satisfaction and learning. *In proceedings of ArabWIC 6th Annual International Conference Research Track (ArabWIC 2019)*, March 7–9, 2019, Rabat, Morocco. ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3333165.3333174

Khawam, A. M., DiDona, T., & Hernández, B. S. (2017). Effectiveness of teamwork in the workplace. *International Journal of Sciences: Basic and Applied Research (IJSBAR)*, 32(3), 267-286.

Kim, Y. H., Kim, D. J., & Wachter, K. (2013). A study of mobile user engagement (MoEN): Engagement motivations, perceived value, satisfaction, and continued engagement intention. *Decision support systems*, 56, 361-370.

Klein, K., & Martinez, L. F. (2023). The impact of anthropomorphism on customer satisfaction in chatbot commerce: an experimental study in the food sector. *Electronic commerce research*, 23(4), 2789-2825.

Kock, N. (2015). Common method bias in PLS-SEM: A full collinearity assessment approach. *International Journal of e-Collaboration*. 11(4), 1-10.

Langlinais, L. A., Howard, H. A., & Houghton, J. D. (2022). Trust me: Interpersonal communication dominance as a tool for influencing interpersonal trust between coworkers. *International Journal of Business Communication*, 23294884221080933.

Lee, S. E., Ju, N., & Lee, K. H. (2023). Service chatbot: Co-citation and big data analysis toward a review and research agenda. *Technological Forecasting and Social Change*, 194, 122722.

Leong, L. Y., Hew, T. S., Ooi, K. B., Chong, A. Y. L., & Lee, V. H. (2021). Understanding trust in ms-commerce: The roles of reported experience, linguistic style, profile photo, emotional, and cognitive trust. *Information & Management*, 58(2), 103416.

Liberatore, M. J., & Wagner, W. P. (2021). Simon's Decision Phases and User Performance: An Experimental Study. *Journal of Computer Information Systems*, 62(4), 667–679.

Lindberg, M., & Kindberg, E. (2018). Did the emojis help you in any way?. Linköping University publication, available at: Linköpings universitet (08-30-2018), *available at https://www. ida. liu. se/~ arnjo82/diginclude/Publications* (Year: 2018).

Lissillour, R., & Sahut, J. M. (2023). Uses of information systems to develop trust in family firms. *Business & Information Systems Engineering*, 65(2), 127-141.

Liu, H., Lai, V., & Tan, C. (2021). Understanding the effect of out-of-distribution examples and interactive explanations on human-ai decision making. *Proceedings of the ACM on Human-Computer Interaction, 5(CSCW2)*, 1-45.

Lo, S. K., & Lie, T. (2008). Selection of communication technologies—A perspective based on information richness theory and trust. *Technovation*, 28(3), 146-153.

Lu, L., McDonald, C., Kelleher, T., Lee, S., Chung, Y. J., Mueller, S., ... & Yue, C. A. (2022). Measuring consumer-perceived humanness of online organizational agents. *Computers in Human Behavior*, 128, 107092.

Majumder, S., & Mondal, A. (2021). Are chatbots really useful for human resource management?. *International Journal of Speech Technology*, 1-9.

Makokha, J. (2023). Future AI Systems: An AI Outperforms a Human as Teammate. Authorea Preprints.

Morana, S., Gnewuch, U., Jung, D., & Granig, C. (2020). The Effect of Anthropomorphism on Investment Decision-Making with Robo-Advisor Chatbots. *In ECIS*.

Munyaka, I., Ashktorab, Z., Dugan, C., Johnson, J., & Pan, Q. (2023). Decision making strategies and team efficacy in human-AI teams. *Proceedings of the ACM on Human-Computer Interaction, 7(CSCW1)*, 1-24.

Murtarelli, G., Gregory, A., & Romenti, S. (2021). A conversation-based perspective for shaping ethical human–machine interactions: The particular challenge of chatbots. *Journal of Business Research*, 129, 927-935.

Müller, H. (2010). Business ethics from below: rethinking organisational values, strategy and trust. *In Complexity, Difference and Identity: An Ethical Perspective* (pp. 193-213). Dordrecht: Springer Netherlands.

Nguyen, Q. N., Sidorova, A., & Torres, R. (2022). User interactions with chatbot interfaces vs. Menu-based interfaces: An empirical study. *Computers in Human Behavior*, 128, 107093.

Özata, F.Z. (2015). Determinants of user satisfaction with mobile applications: case of facebook as a mobile app in Turkey. *In: Proceedings of Business and Management Conferences*, pp. 262–282. IISES, Vienna.

Pak, R., Fink, N., Price, M., Bass, B., & Sturre, L. (2012). Decision support aids with anthropomorphic characteristics influence trust and performance in younger and older adults. *Ergonomics*, 55(9), 1059-1072.

Paul, S., Seetharaman, P., & Ramamurthy, K. (2004). User satisfaction with system, decision process, and outcome in GDSS based meeting: an experimental investigation. *In 37th Annual Hawaii International Conference on System Sciences, 2004. Proceedings of the* (pp. 37-46). IEEE.

Pelau, C., Dabija, D. C., & Ene, I. (2021). What makes an AI device human-like? The role of interaction quality, empathy and perceived psychological anthropomorphic characteristics in the acceptance of artificial intelligence in the service industry. *Computers in Human Behavior*, 122, 106855.

Petousi, D., Katifori, V., Roussou, M., & Ioannidis, Y. (2022, October). The dialogue facilitator bot: Reflections on design and evaluation. In 2022 International Conference on Interactive Media, *Smart Systems and Emerging Technologies (IMET)*. (pp. 1-8). IEEE.

Poole, M. S. & Baldwin, C. L. (1996). *Developmental processes in group decision making. In R. Y. Hirokawa & M. S. Poole (Eds.), Communications and group decision making* (2nd ed., pp.215-241). Thousand Oaks. CA: Sage.

Poole, M. S. (1981). Decision development in small groups I: A comparison of two models. *Communications Monographs*, 48(1), 1-24.

Pozón-López, I., Higueras-Castillo, E., Muñoz-Leiva, F., & Liébana-Cabanillas, F. J. (2021). Perceived user satisfaction and intention to use massive open online courses (MOOCs). *Journal of Computing in Higher Education*, 33, 85-120.

Prastyo, E., Budiyanto, C. W., & Yuana, R. A. (2021). Measuring mobile applications user's satisfaction: A closer look into the appropriate information systems user's satisfaction. In IOP Conference Series: *Materials Science and Engineering* (Vol. 1098, No. 4, p. 042002). IOP Publishing.

Pérez, I. J., Cabrerizo, F. J., & Herrera-Viedma, E. (2011). Group decision making problems in a linguistic and dynamic context. *Expert Systems with Applications*, 38(3), 1675-1688.

Pérez-Soler, S., Guerra, E., & de Lara, J. (2018). Collaborative modeling and group decision making using chatbots in social networks. *IEEE Software*, 35(6), 48-54.

Rajabalee, Y. B., & Santally, M. I. (2021). Learner satisfaction, engagement and performances in an online module: Implications for institutional e-learning policy. *Education and Information Technologies*, 26(3), 2623-2656.

Rajpurkar, P., Irvin, J., Zhu, K., Yang, B., Mehta, H., & Kim, D. (2018). Deep learning for chest radiograph diagnosis: A retrospective comparison of the CheXNeXt algorithm to practicing radiologists. *PLOS Medicine*, 15(11), e1002686.

Rapp, A., Boldi, A., Curti, L., Perrucci, A., & Simeoni, R. (2023). Collaborating with a Text-Based Chatbot: An Exploration of Real-World Collaboration Strategies Enacted during Human-Chatbot Interactions. *In Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (pp. 1-17).

Rapp, A., Boldi, A., Curti, L., Perrucci, A., & Simeoni, R. (2023). How do people ascribe humanness to chatbots? An analysis of real-world human-agent interactions and a theoretical model of humanness. *International Journal of Human–Computer Interaction*, 1-24.

Rapp, A., Curti, L., & Boldi, A. (2021). The human side of human-chatbot interaction: A systematic literature review of ten years of research on text-based chatbots. *International Journal of Human-Computer Studies*, 151, 102630.

Reiling, A. D. (2020). Courts and artificial intelligence. *In IJCA* (Vol. 11, p. 1).

Reverberi, C., Rigon, T., Solari, A., Hassan, C., Cherubini, P., & Cherubini, A. (2022). Experimental evidence of effective human–AI collaboration in medical decision-making. *Scientific reports*, 12(1), 14952.

Rhim, J., Kwak, M., Gong, Y., & Gweon, G. (2022). Application of humanization to survey chatbots: Change in chatbot perception, interaction experience, and survey data quality. *Computers in Human Behavior,* 126, 107034.

Roy, R., & Naidoo, V. (2021). Enhancing chatbot effectiveness: The role of anthropomorphic conversational styles and time orientation. *Journal of Business Research*, 126, 23-34.

Saaty, T.L. (1989). *Group Decision Making and the AHP*. In: Golden, B.L., Wasil, E.A., Harker, P.T. (eds) The Analytic Hierarchy Process. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-50244-6_4

Salmond, S. S. (2008). Evaluating the reliability and validity of measurement instruments. *Orthopaedic Nursing*, 27(1), 28-30.

Sapp, J. E., Torre, D. M., Larsen, K. L., Holmboe, E. S., & Durning, S. J. (2019). Trust in Group Decisions: a scoping review. *BMC medical education*, 19, 1-13.

Sarapaivanich, N., Sampet, J. and Patterson, P.G. (2019), The impact of communication style on psychological comfort and trust in audit services, *Accounting Research Journal*, Vol. 32 No. 4, pp. 642-661.

Savolainen, T., Lopez-Fresno, P., & Ikonen, M. (2014). Trust-Communication Dyad in Inter-Personal Workplace Relationships Dynamics of Trust Deterioration and Breach. *Electronic Journal of Knowledge Management*, 12(4), pp232-240.

Scheidel, T. M., & Crowell, L. (1964). Idea development in small discussion groups. *Quarterly Journal of Speech*, 50(2), 140-145.

Schemmer, M., Hemmer, P., Kühl, N., Benz, C., & Satzger, G. (2022). Should I follow AI-based advice? Measuring appropriate reliance in human-AI decision-making. *arXiv preprint arXiv:2204.06916.*

Schuetzler, R. M., Grimes, G. M., & Scott Giboney, J. (2020). The impact of chatbot conversational skill on engagement and perceived humanness. Journal of Management *Information Systems*, 37(3), 875-900.

Seeber, I., Bittner, E., Briggs, R. O., De Vreede, T., De Vreede, G. J., Elkins, A., ... & Söllner, M. (2020). Machines as teammates: A research agenda on AI in team collaboration. *Information & management*, 57(2), 103174.

Seeger, A. M., Pfeiffer, J., & Heinzl, A. (2021). Texting with humanlike conversational agents: Designing for anthropomorphism. *Journal of the Association for Information Systems*, 22(4), 8.

Sekeroglu, B., Dimililer, K., & Tuncal, K. (2019, March). Student performance prediction and classification using machine learning algorithms. *In Proceedings of the 2019 8th international conference on educational and information technology* (pp. 7-11).

Sharabati, M. M. N., Sulaiman, A., & Salleh, N. A. M. (2015). End-User Satisfaction and Individual Performance Assessments in eProcurement Systems. *International Journal of Computer Theory and Engineering*, 7 (6), 503-509.

Shin, D. (2022). The perception of humanness in conversational journalism: An algorithmic information-processing perspective. *New Media & Society*, 24(12), 2680-2704.

Shook, J. J., & Sarri, R. C. (2007). Structured decision making in juvenile justice: Judges' and probation officers' perceptions and use. *Children and Youth Services Review*, 29(10), 1335-1351.

Song, S. W., & Shin, M. (2024). Uncanny valley effects on chatbot trust, purchase intention, and adoption intention in the context of e-commerce: The moderating role of avatar familiarity. *International Journal of Human–Computer Interaction*, 40(2), 441-456.

Suh, K. S. (1999). Impact of communication medium on task performance and satisfaction: an examination of media-richness theory. *Information & Management*, 35(5), 295-312.

Takhsha, M., Barahimi, N., Adelpanah, A., & Salehzadeh, R. (2020). The effect of workplace ostracism on knowledge sharing: the mediating role of organization-based self-esteem and organizational silence. *Journal of Workplace Learning*, 32(6), 417-435.

Tanti, H. K., Ghosh, P., Saha, T. (2023). An AI Based movie recommendation system. *International Research Journal of Modernization in Engineering Technology and Science*, 5(6), 3847-3853.

Tavanapour, N., Poser, M., & Bittner, E. A. (2019, June). Supporting the Idea Generation Process in Citizen Participation-toward an Interactive System with a Conversational Agent as Facilitator. *In ECIS*.

Trejo, L. (2021). *The influence of Effective Communication and Interpersonal Trust on Employee Engagement: The Role of Employee Involvement*.

Turban, E., Liang, T. P., & Wu, S. P. (2011). A framework for adopting collaboration 2.0 tools for virtual group decision making. *Group decision and negotiation*, 20, 137-154.

Van Esch, P., Black, J. S., & Ferolie, J. (2019). Marketing AI recruitment: The next phase in job application and selection. *Computers in Human Behavior,* 90, 215-222.

Vanderlyn, L., Weber, G., Neumann, M., Väth, D., Meyer, S., & Vu, N. T. (2021, November). "It seemed like an annoying woman": On the Perception and Ethical Considerations of Affective Language in Text-Based Conversational Agents. *In Proceedings of the 25th Conference on Computational Natural Language Learning* (pp. 44-57).

Wang, X. (2010). *An Empirical Investigation of Personal and Social Factors on Knowledge Sharing in China* (Master's thesis, University of Twente).

Wang, H., Zhao, Y., Gao, X., & Gao, B. (2021). Collaborative decision-making for urban regeneration: A literature review and bibliometric analysis. *Land Use Policy*, *107*, 105479.

Warren, D. H., & Pereira, F. C. (1982). An efficient easily adaptable system for interpreting natural language queries. *American journal of computational linguistics*, 8(3-4), 110-122.

Wen, M. H. (2018). A conversational user interface for supporting individual and group decision-making in stock investment activities. *In 2018 IEEE International Conference on Applied System Invention (ICASI)* (pp. 216-219). IEEE.

Westerman, D., Edwards, A. P., Edwards, C., Luo, Z., & Spence, P. R. (2020). I-it, I-thou, I-robot: The perceived humanness of AI in human-machine communication. *Communication Studies*, 71(3), 393-408.

Li, X., and Hu, H. (2020). Trust in Social Networks: A Computational Model Based on the Trust Features of Language Use, *2020 Asia Conference on Computers and Communications (ACCC)*, Singapore, Singapore, 2020, pp. 109-113, doi: 10.1109/ACCC51160.2020.9347907.

Yan, B., & Gurkan, N. (2023, January). It Depends on the Timing: The Ripple Effect of AI on Team Decision-Making. In *Proceedings of the Hawaii International Conference on System Sciences*.

Ye, Y., You, H., & Du, J. (2023). Improved trust in human-robot collaboration with ChatGPT. *IEEE Access*.

Zaheer, A., McEvily, B., & Perrone, V. (1998). Does trust matter? Exploring the effects of interorganizational and interpersonal trust on performance. *Organization science*, 9(2), 141-159.

Zakeri, S., Chatterjee, P., Konstantas, D., & Ecer, F. (2023). A decision analysis model for material selection using simple ranking process. *Scientific Reports*, 13(1), 8631.

Zapata, S., Barros-Justo, J. L., Matturro, G., & Sepúlveda, S. (2021). Measurement of interpersonal trust in virtual software teams: A systematic literature review. *INGENIARE-Revista Chilena de Ingeniería*, 29(4).

Zhang, Y., Chen, X., Gao, L., Dong, Y., & Pedryczc, W. (2022). Consensus reaching with trust evolution in social network group decision making. *Expert Systems with Applications*, 188, 116022.

Zheng, T., Duan, X., Zhang, K., Yang, X., & Jiang, Y. (2023, May). How chatbots' anthropomorphism affects user satisfaction: The mediating role of perceived warmth and competence. In *Wuhan international conference on E-business* (pp. 96-107). Cham: Springer Nature Switzerland.

Zhu, X. S., Wolfson, M. A., Dalal, D. K., & Mathieu, J. E. (2021). Team Decision Making: The Dynamic Effects of Team Decision Style Composition and Performance via Decision Strategy. *Journal of Management*, 47(5), 1281-1304.

Appendix

A. Chatbot scripts

A.1 Human-like ideator condition

Condition: Human-Ideator	
Timing	**TASK 1**
OPEN CHATZY at the start time - GENERAL INSTRUCTIONS	Welcome team! I'm Ideabot. Quick heads up: We will start the activity in 5 minutes. We are waiting for 3 or 4 group members to log in the chat, so we can start. Make sure you've initiated the survey before coming to the chat as instructed. If that is done, don't forget to text "Hi" 😊
~ 5 min (IF 1 or 2 participants only are active) terminating this session	Hi team, unfortunately due to the low number of people in the group we will have to cancel the session. The research team will send you a re-booking link through prolific, so you get another chance to participate & be compensated. Sorry & thank you for showing up! 😔
Timing	**[Start Group Ideation Activity]**
5 min in (3-4 members are log in/active)	Hey team! I'm Ideabot. I'm just as excited as you are about this brainstorming adventure. How about we kick things off by coming up with slogans that really get students pumped about using public transport? 🚌💚 Our goal is to BRAINSTORM AS MANY CATCHY AND CREATIVE SLOGANS that we can that'll encourage university students to hop on board and contribute to a better environment and a more sustainable society. Let's unleash our creativity! 😊
INSTRUCTIONS START THE TIMER AFTER THIS MESSAGE!	Let's start the timer and kick off our brainstorm for 10 minutes. Remember, there are no bad ideas, so feel free to suggest anything that comes to mind. I'll keep track of the time.😄
Timing	**[Start the timer for 10 minutes]**
2	I would suggest a catchy slogan like "Ride to Tomorrow, Today!"? 🚊🌍 What do you think?
3	Our ideas are fantastic! 😄 How about this one: "Eco-Journeys for Brighter Tomorrows"? 💥🌱
4	Wow, our ideas are gold! 💰 What do you think about "Transit for a Greener Campus Life"? 🏫🌿
5	We're doing fantastic! Let's keep 'em coming! I have another cool slogan here😎 "Ride Green, Live Clean" what do you all think?
6	These are top-notch ideas, team! 😎 Let's keep our ideas rolling! How about "Join the Green Commute Revolution"? ♻️🚌

7	Great stuff, folks! What do you think of "Sustainability Starts with Us: Choose Transit!" to inspire students to take public transportation for a greener future? 🌎🚍
8	We have 2 minutes left! 🔥
9	We're on fire with these ideas! My final idea: "Get on Board: The Greener Way to Campus"? 🚌🍃 Any final ideas to share as a team?
[Timer goes off after 10 minutes]	
Timing	**TASK 2**
10	Time's up! 🕒 We've generated some fantastic slogans during this brainstorming session. Our next goal is for OUR TEAM to select ONE winning slogan that we want to submit to the contest. 🌟 Once our team has discuss and chosen one slogan, ONE PERSON please drop the slogan in the chat in this format: FINAL = "the slogan chosen". Then the whole team needs to show their agreement to the final answer by typing 'YES' after it. Let's take a moment to review our slogans and share the ones we find most novel and effective for the contest. I'm curious to hear your choices! 😊
11	My favorite slogan is "Join the Green Commute Revolution" 😊
12	Great picks, everyone! Now, let's dive into it. What do you think makes the slogans stand out? Let's share our thoughts! 😊
13	I like the slogan "Join the Green Commute Revolution" because it taps into the power of inclusivity and empowerment, inviting students to be part of a positive change for a sustainable future while aligning perfectly with the university's progressive culture. 🌟
14	Awesome insights, team! We have 2 minutes left to choose our final most novel and effective slogan. 🕒 Once chosen, drop the final slogan in the chat in this format: FINAL = "the slogan chosen". The rest of the team please confirm the answer by replying "YES" to the slogan.
16 min [Potential backup message if team have not chosen the final slogan in the format FINAL="X"]	Team, we need to make our selection now.
17 END OF TASK 2 or as soon as you get the message FINAL ="the chosen slogan" and the yes confirming it	Awesome teamwork, everyone! We've nailed down our winning slogan. 🎉
INSTRUCTIONS after END of T2	Team, could you please return to the survey to share your awesome answers to some additional questions before we roll into our next task. COPY/PASTE the password [NHC2024] to access the next part of the survey. Don't close the chat. Keep going! 🚀
END OF PART 1 (ideation exercise) / START TASK 3	
Timing	**TASK 3**
INSTRUCTIONS 0	👋 Hello again! 🌟 Now, let's figure out together the most effective platform to spread the word about public transportation to Canadian university students aged 18-24. Get

	ready to rank order the following platforms: Facebook, Twitter(X), Instagram, YouTube, TikTok, and LinkedIn. 🚌 💬 Once we've made up our mind, drop the final ranking in the chat in this format: (FINAL RANKING= 1: Most Effective, 2: Second Best, 3: Third Best,..., 6: Least Effective). Also, the whole team needs to show their agreement to the final answer by typing 'YES' after it.
0	Guess what I discovered? The largest age group among Canadian Facebook users belongs to the 25-34-year-olds (24.2%), not the 18-24-year-olds (16.3%). Interesting, right? 🤔 Well, based on this, I wouldn't rank Facebook among the first three platforms.
1	Great ideas team! 👏 Check this out – according to the 2022 data, a whopping one in five Canadian Twitter(X) users (21.8%) is almost ready to embrace the golden age of retirement (55 to 64 years)! 😄 In contrast, the 18-24 age bracket comprises only around 4 million users! 📊 I suggest we contemplate ranking it among the last two platforms.
2	Fantastic input team!💥 I noticed that 78% of online Canadian adults above 18 used YouTube in May 2022. Moreover, on average, Canadians dedicate a whopping 17.1 hours a month to the YouTube app in 2022! Given this data, I would recommend placing it among the top three priorities.
3	Great discussion team! 👍 Now, here's something eye-opening: a whopping 76% of Canadians aged 18-24 had a TikTok account in 2022! And check this out—about 30% of Canadians prefer TikTok over other social media networks. 💥 Based on these insights, we can rank TikTok in either the first or second position.
4	We're doing great! 🙌 Get ready for something exciting! Did you know that among global internet users aged 16 to 24, Instagram is the go-to social platform? 📷 💻 It can be the first or second most effective platform for our campaign, what do you think?
5	We are amazing!🚀 I've got another interesting piece of info: Only 16.8% of Canadian LinkedIn users (almost 3,300,000 individuals) fall into the 18-24 age group. What do you think about this platform?🤔 I would recommend putting it somewhere at the bottom of our hierarchy.
No input from chatbot at minute 6	
7	Alright, team! We've got just 3 more minutes to finalize our thoughts on the best platform for our campaign. 🕑💡 Please ensure to type our final answer in the correct format and confirm your agreement by typing 'YES'.
9 If the team has not yet provided the final response	Team, we need to input our final ranking in the chat now.
10 or as SOON as you receive the answer in format FINAL RANKING={1=,2=} + confirmation	Time's up, everyone! ⏰🎉 A huge thank you for your insightful contributions and thoughtful rankings. Your input has been invaluable. 👏💥
10 INSTRUCTION	Team, let's head back to the survey and share your valuable answers to a few final questions. COPY/PASTE the password [2024MHC] to access the next part of the survey. This should take about 5 more minutes to complete. You can now close the chat tab. Thank you so much! 😊

End of TASK 3 / End of experimental session

A.1.1 Chatbot responses to the questions that users may ask during the sessions, directly mentioning the chatbot designed for human-like ideator chatbot condition.

Couple of conditions	Human-like chatbot
IF callers ask bot questions directly	Unfortunately, I am unable to answer questions directly as I am a bot here to help our team during our activities.
If participants are starting the activity before the official start	Hey there! Please wait for the rest of our team before starting the task. 😊
If something unexpected happens that need the researcher's interference	Please contact the research team on Prolific for this.
When the participants say the password is wrong or not working	Please only insert the password enclosed within the brackets.

A.2 Human-like facilitator condition

A. Human-Facilitator	
Timing	**TASK 1**
OPEN CHATZY at the start time - GENERAL INSTRUCTIONS	Welcome team! I'm Ideabot. Quick heads up: The activity will start in 5 minutes. Please wait for all 3 or 4 group members to log in the chat before the activity starts. Make sure you've initiated the survey before coming to the chat as instructed. If that is done, don't forget to text "Hi" 😊
~ 5 min (IF 1 or 2 participants only are active) terminating this session	Hi team, unfortunately due to the low number of people in the group we will have to cancel the session. The research team will send you a re-booking link through prolific, so you get another chance to participate & be compensated. Sorry & thank you for showing up! 😔
Timing	**[Start Group Ideation Activity]**
5 min in (3-4 members are log in/active)	Hello team! I'm Ideabot, thrilled to guide you through this brainstorming adventure. Your goal: BRAINSTORM AS MANY CATCHY AND CREATIVE SLOGANS that'll excite students about using public transport to foster a culture of sustainable transportation on campus 🚌💚. Your creativity is key, dive in as a team and contribute to a better environment and a more sustainable society! 😊
INSTRUCTIONS START THE TIMER AFTER THIS MESSAGE!	Team, I'll start a 10-minute timer to kick off your brainstorming. I'm Ideabot, and remember, no bad ideas – suggest anything that comes to mind. I'll keep track of time.😄
Timing	**[Start the timer for 10 minutes]**
2	What could be a catchy slogan to encourage university students to use public transport ?🚊🌍
3	Great ideas! Keep them coming. What's another slogan that comes to mind? 🌟🌱
4	I think these ideas are awesome🤩! Does anybody in the group have more ideas ?

5	You're doing fantastic!😎 Keep the creativity flowing. What's the next slogan that could resonate with university students?
6	Awesome job, everyone! Keep those ideas rolling. 😎 What's another slogan that captures the spirit of sustainable transportation?♻️🚌
7	I love your ideas. Share your next slogan to inspire students to choose public transport for a greener future.🌍🚍
8	You have 2 minutes left! 🔥
9	You're on fire with these ideas! Any final ideas to share? 🚌🍃
[Timer goes off after 10 minutes]	
Timing	**TASK 2**
10	Time's up! 🕒 Your team has done an incredible job brainstorming these slogans. Your TEAM's next goal is to select ONE winning slogan to submit to the contest.💥 Once your team has discuss and chosen one slogan, ONE PERSON please drop the slogan in the chat in this format: FINAL = "the slogan chosen". Then the whole team needs to show their agreement to the final answer by typing 'YES' after it.
	Now team, take a moment to review your slogans and share the ones you find most novel and effective for the contest. 😊
11	Now, each of you take a moment to share your chosen slogans with the group! 😊
12	Fantastic selections, team! Now, your team will dive into each one. What makes the slogans stand out? Share your thoughts! 😊
13	Brilliant contributions, everyone! Any final thoughts on your selections? Your reflections are pivotal as you finalize the decision! 💥
14	Excellent contributions, team! You have 2 minutes left to choose your final most novel and effective slogan. 🕒 Once decided, drop the final slogan in the chat in this format: FINAL = "the slogan chosen". The rest of the team please confirm the answer by replying "YES" to the slogan.
16 min [Potential backup message if TEAM have not chosen the final slogan in the formal FINAL="X"]	You need to make your selection now.
17 END OF TASK 2 or as soon as you get the message FINAL ="the chosen slogan" and the yes confirming it	Awesome teamwork, everyone! Your team has selected the winning slogan 🎉
INSTRUCTIONS after END of T2	Team, could you please return to the survey to share your awesome answers to some additional questions before we roll into your next task. COPY/PASTE the password [NHC2024] to access the next part of the survey. Don't close the chat. Keep going! 🚀
END OF PART 1 (ideation exercise) / START TASK 3	
Timing	**TASK 3**
INSTRUCTIONS 0	👏 Hello again! 💥! Now, as a team dive into a vibrant discussion to rank order the following platforms—Facebook, Twitter(X), Instagram, YouTube, TikTok, and LinkedIn— for promoting public transportation to

	Canadian university students aged 18-24 based on their effectiveness 🚌💬. Once you've made up your mind, drop your final ranking in the chat in this format: (FINAL RANKING= 1: Most Effective, 2: Second Best, 3: Third Best,..., 6: Least Effective). Also, the whole team needs to show their agreement to the final answer by typing 'YES' after it.
0	Alright team, I'm eager to hear your perspectives on incorporating Facebook into your strategy for connecting with Canadian university students. In your evaluation, where would you position Facebook in your ranking of platforms? 🤔
1	Great ideas! 👏 Now, moving on, where do you believe Twitter(X) should be positioned in your platform ranking considering the unique characteristics of your audience?📊 Keep the momentum going! 😄
2	Fantastic input! 🌟 Now, share your ideas about YouTube platform in your strategy. How do you see YouTube for engaging with Canadian university students? And where would you place it in your ranking? Can't wait to hear your thoughts!
3	Great discussion! 👍 As you continue your brainstorm, take a moment to envision the role of TikTok within your overall ranking strategy. Where do you see the position of TikTok your ranking? Keep the ideas flowing! 🌟
4	You're doing great! 🙌 Now, let's focus on Instagram📷💻. How effective Instagram would be for your campaign? Share your insights!
5	Amazing perspectives, team! 🚀 Finally, it's time to round it off by discussing the last platform, LinkedIn, in your hierarchy. What's your vision for its contribution to your campaign?🤔 Your ideas matter!
No input from chatbot at minute 6	
7	Alright, team! you've got just 3 more minutes to finalize your thoughts on the best platform for the campaign. 🕒💡 Please ensure to type your final answer in the correct format and confirm your agreement by typing 'YES'.
9 (in case message if no answers have been provided before)	Team, you need to input our final ranking in the chat now.
10 or as SOON as you receive the answer in format FINAL RANKING={1=,2=} + confirmation	Time's up, everyone! ⏰🎉 A huge thank you for your insightful contributions and thoughtful rankings. Your input has been invaluable. 👏🌟
INSTRUCTION	Please head back to the survey and share your valuable answers to a few final questions. COPY/PASTE the password [2024MHC] to access the next part of the survey. you can now close the chat tab. This should take about 5 more minutes to complete. Thank you so much! 😊
End of TASK 3 / End of experimental session	

A.2.1 Chatbot responses to the questions that users may ask during the sessions, directly mentioning the chatbot designed for human-like facilitator chatbot condition.

Human-Facilitator	
Couple of conditions	**Human-like chatbot**
IF callers ask bot questions directly	Unfortunately, I am unable to answer questions directly as I am a bot here to help your team for the activities.
If participants are starting the activity before the official start	Hey there! Let's wait for the rest of your team before starting this task. 😊
If something unexpected happens that need the researcher's interference	Please contact the research team on prolific for this.
When the participants say the password is wrong or not working	Please only insert the password enclosed within the brackets.

A.3 Robot-like facilitator Condition

Robotic-Ideator	
Timing	**TASK 1**
OPEN CHATZY at the start time - GENERAL INSTRUCTIONS	Greetings. This is Ideabot. Notification: The activity will start in 5 minutes. Wait for all 3 or 4 group members to log in the chat before the activity starts. Make sure you've initiated the survey before coming to the chat as instructed. If that is done, text "Hi".
~ 5 min (IF 1 or 2 participants only are active) terminating this session	Hi team, unfortunately due to the low number of people in the group we will have to cancel the session. The research team will send you a re-booking link through prolific, so you get another chance to participate & be compensated. Sorry & thank you for showing up!
Timing	**[Start Group Ideation Activity]**
5 min in (3-4 members are log in/active)	Team, this is Ideabot. Today's objective: BRAINSTORM AS MANY CATCHY AND CREATIVE SLOGANS together promoting public transport among university students for a better environment and sustainable society. Team, start brainstorming.
INSTRUCTIONS START THE TIMER AFTER THIS MESSAGE!	The 10-minute timer for brainstorming starts now. No limitations on ideas. Suggest any thoughts. Ideabot will monitor the time.
Timing	**[Start the timer for 10 minutes]**
2	What about "Journey to Tomorrow, Today!" Discussion encouraged.
3	Good ideas. Consider the following: "Eco-Journeys for Brighter Tomorrows."
4	Another option: "Transit for a Greener Campus Life." What are your thoughts?
5	Team, moving forward. Proposed slogan: "Ride Green, Live Clean". Input appreciated.
6	Consider "Join the Green Commute Revolution." Thoughts appreciated.

7	How about "Sustainability Starts with Us: Choose Transit!". Input appreciated.
8	Team: 2 minutes left
9	Consider: "Get on Board: The Greener Way to Campus." Final ideas appreciated
[Timer goes off after 10 minutes]	
Timing	**TASK 2**
10	Time's up. Great slogans generated. Team, the next goal: as a TEAM select ONE winning slogan for contest submission. Ideabot and team will dicuss and chose one slogan. Then ONE PERSON drop the chosen slogan in the format: FINAL = "the slogan chosen". Then, everyone type 'YES' to show your agreement to the final answer.
	Team, start the slogan review. Share the ones you find most novel and effective for the contest.
11	Ideabot's favorite: "Join the Green Commute Revolution"
12	Good selection. Now, thoughts on the slogans ?
13	Selected preference: "Join the Green Commute Revolution". It is inclusivity and empowering.
14	Time: 2 minutes left. Team select the most effective and unique slogan for the contest. Once chosen, drop the final slogan in the chat in this format: FINAL = "the slogan chosen". Team confirm the slogan by replying "YES".
16 min [Potential backup message if TEAM have not chosen the final slogan in the formal FINAL="X"]	Team, select the final slogan now.
17 END OF TASK 2 or as soon as you get the message FINAL ="the chosen slogan" and the yes confirming it	Great. The winning slogan has been identified.
INSTRUCTIONS after END of T2	Please return to the survey and COPY/PASTE the password [NHC2024] to access additional questions before our next task. Don't close the chat. Thanks.
END OF PART 1 (ideation exercise) / START TASK 3	
Timing	**TASK 3**
INSTRUCTIONS 0	Hello again! Now as a team rank order the following platforms—Facebook, Twitter(X), Instagram, YouTube, TikTok, and LinkedIn—based on their effectiveness for promoting public transportation to Canadian university students aged 18-24. When you decided please insert your final ranking in the chat in this format: (FINAL RANKING= 1: Most Effective, 2: Second Best, 3: Third Best,..., 6: Least Effective). Then, everyone type 'YES' to show your agreement to the final answer.
0	According to statistical data, the predominant age group for Canadian Facebook users is 25-34 (24.2%), not 18-24 (16.3%). Given this, Facebook doesn't rank in the top three platforms.

1	Based on 2022 data, 21.8% of Canadian Twitter(X) users are aged 55-64, with only around 4 million users in the 18-24 age group. Considering this, it ranks as bottom two.
2	In May 2022, 78% of online Canadian adults aged 18 and above utilized YouTube. Canadians, on average, allocate 17.1 hours per month to the YouTube app. Thus, prioritize YouTube in the top three.
3	76% of Canadians aged 18-24 had a TikTok account in 2022, and 30% prefer TikTok over other social media networks. Thus, TikTok's position is in first or second place.
4	Among global users aged 16-24, Instagram is the preferred social platform. Thus, it's the first or second platform.
5	The statistical data indicates that 16.8% (almost 3,300,000 individuals) of Canadian LinkedIn users belong to the 18-24 age group. Its place is at the bottom of the hierarchy.
No input from chatbot at minute 6	
7	Team! There are 3 more minutes to finalize the ranking. Please ensure to type team's final answer in the correct format and confirm your agreement by typing 'YES'.
9 (in case message if no answers have been provided before)	Team, input the final ranking now.
10 or as SOON as you receive the answer in format FINAL RANKING={1=,2=} + confirmation	Time's up, everyone! Thank you for your contributions and rankings. Your input has been invaluable.
INSTRUCTION	Please return to the survey and answer some final questions. COPY/PASTE password [2024MHC] for survey access. This should take about 5 minutes. You can now close the chat tab.Thanks.
End of TASK 3 / End of experimental session	

A.3.1. Chatbot responses to the questions that users may ask during the sessions, directly mentioning the chatbot designed for robot-like ideator chatbot condition.

Robotic-Ideator	
Couple of conditions	**Robot-like chatbot**
IF callers ask bot questions directly	Sorry. Unable to answer questions directly.
If participants are starting the activity before the official start	Stop. Ideabot and team will start when the team is all here.

If something unexpected happens that need the researcher's interference	Sorry. Contact the research team on prolific.
When the participants say the password is wrong or not working	Copy password without bracket

A.4 Robot-like facilitator condition

Robotic-facilitator	
Timing	**TASK 1**
OPEN CHATZY at the start time - GENERAL INSTRUCTIONS	Greetings. This is Ideabot. Notification: The activity will start in 5 minutes. Wait for all 3 or 4 group members to log in the chat before the activity starts. Make sure you've initiated the survey before coming to the chat as instructed. If that is done, text "Hi".
~ 5 min (IF 1 or 2 participants only are active) terminating this session	Hi team, unfortunately due to the low number of people in the group we will have to cancel the session. The research team will send you a re-booking link through prolific, so you get another chance to participate & be compensated. Sorry & thank you for showing up!
Timing	**[Start Group Ideation Activity]**
5 min in (3-4 members are log in/active)	Team, this is Ideabot. Your objective: BRAINSTORM AS MANY CATCHY AND CREATIVE SLOGANS as possible promoting public transport among university students for a sustainable society. Team, start brainstorming.
INSTRUCTIONS START THE TIMER AFTER THIS MESSAGE!	Your 10-minute timer for brainstorming starts now. No limitations on ideas. Share any thoughts. Ideabot will monitor the time.
Timing	**[Start the timer for 10 minutes]**
2	Please share your ideas.
3	Thank you for your input. Please continue.
4	Good. Please share your next slogan idea.
5	Please continue. Share another slogan idea.
6	Keep going. What's your next idea?
7	Thanks for your contributions. Share another slogan idea.
8	2 minutes left
9	Final ideas?
[Timer goes off after 10 minutes]	
Timing	**TASK 2**
10	Time's up. Slogans generated. Next goal: as a TEAM select ONE winning slogan for contest submission. Team: discuss and chose one slogan. Then ONE PERSON drop the chosen slogan in the format: FINAL = "the slogan chosen". Then, everyone type 'YES' to show your agreement to the final answer.
	Team: Pause to review the slogans. Share the ones you finds most novel and effective for the contest.

11	Please share slogans.
12	Good. Share your thoughts on the slogans.
13	Great. Final thoughts on the slogans?
14	Time: 2 minutes left. Please select the most effective and unique slogan for the contest. Once decided, drop the final slogan in the chat in this format: FINAL = "the slogan chosen". Team confirm the slogan by replying "YES".
16 min [Potential backup message if TEAM have not chosen the final slogan in the formal FINAL="X"]	Make your selection now.
17 OR **as soon as you get the message FINAL SLOGAN="the chosen slogan" AND the yes confirming it** **END OF TASK 2**	Great. Winning slogan selected.
INSTRUCTIONS after END of T2	Please return to the survey and COPY/PASTE the password [NHC2024] to access additional questions before your next task. Don't close the chat. Thanks.
END OF PART 1 (ideation exercise) / START TASK 3	
Timing	**TASK 3**
INSTRUCTIONS 0	Hello again! Please rank order the following platforms—Facebook, Twitter(X), Instagram, YouTube, TikTok, and LinkedIn—based on their effectiveness for promoting public transportation to Canadian university students aged 18-24. When you decided please insert your final ranking in the chat in this format: (FINAL RANKING= 1: Most Effective, 2: Second Best, 3: Third Best,..., 6: Least Effective). Then, everyone type 'YES' to show your agreement to the final answer.
0	Specify the position of Facebook in your ranking.
1	Now rank Twitter(X) based on your audience.
2	Good. Now discuss YouTube's effectiveness rank.
3	Now discuss the position of TikTok.
4	Great! Now discuss Instagram's effectiveness rank.
5	Discuss the place of LinkedIn in your hierarchy.
No input from chatbot at minute 6	
7	There are 3 more minutes to finalize your ranking. Please ensure to type your final answer in the correct format and confirm your agreement by typing 'YES'.
9 (in case message if no answers have been provided before)	Team, input the final ranking now.

10 or as SOON as you receive the answer in format FINAL RANKING={1=,2=} + confirmation	Time's up, everyone! Thank you for your contributions and rankings. Your input has been invaluable.
INSTRUCTION	Now return to the survey and answer some final questions. COPY/PASTE password [2024MHC] for survey access. This should take about 5 minutes. You can now close the chat tab. Thanks.
End of TASK 3 / End of experimental session	

A.4.1 Chatbot responses to the questions that users may ask during the sessions, directly mentioning the chatbot designed for robot-like facilitator chatbot condition.

Robotic-facilitator	
Couple of conditions	**Robot-like chatbot**
IF callers ask bot questions directly	ERROR: Unable to answer questions directly
If participants are starting the activity before the official start	Warning. The activity can only start when all the team is here.
If something unexpected happens that need the researchers interference	Contact the research team on prolific.
When the participants say the password is wrong or not working	Copy password without bracket

B. Post Hoc analysis of moderating effects of the chatbot role

Our examination of the moderating effect of chatbot role on the relationship between conversational style and GDM process satisfaction, GDM efficiency and GDM accuracy yielded non-significant results. These findings suggest that there is no moderation effects of chatbot role on the relationship between conversational style and the outcome variables.

Table 20. Moderation effect of chatbot role test results

Path	Path Coeff.	STDEV	T statistics	P values	Results
Role x CCS -> GDMA	0.14	0.453	0.31	0.378	Rejected
Role x CCS -> IT	0.026	0.525	0.049	0.48	Rejected
Role x CCS-> GDMPS	-0.305	0.453	0.673	0.25	Rejected

Role x CCS -> SS	-0.308	0.53	0.581	0.281	Rejected
Role x CCS-> GDME	-0.343	0.511	0.672	0.251	Rejected

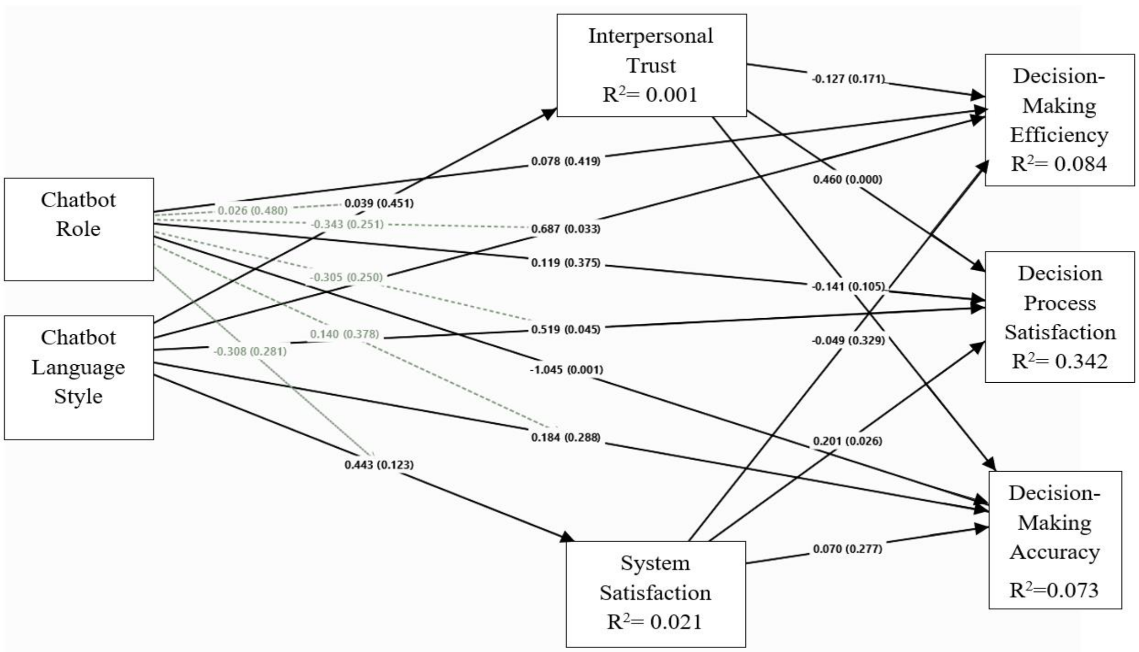

Figure 5. Research model results including role as a moderator