



Cascade Effect in AI-Human Interactions: Study of Cognitive Biases, Personality Traits and Decision Making Patterns

Lekshmi Parvathy A^{*}, Aadithyan C.A², Nimisha M³

1. Lekshmi Parvathy A, Assistant professor, Kings Cornerstone International College, Chennai.
2. Aadithyan C.A, PhD Scholar, Department of Social Work, Pondicherry University.
3. Nimisha M, Mphil PSW scholar, Dept. of Psychiatric Social work, NIMHANS.

***Correspondence to:** Lekshmi Parvathy A, Assistant professor, Kings Cornerstone International college, Chennai.

Copyright

© 2025 **Lekshmi Parvathy A**, This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Received: 04 Mar 2025

Published: 13 Mar 2025

DOI: <https://doi.org/10.5281/zenodo.15044403>

Abstract

This qualitative study examines the cascade effect in artificial intelligence (AI)-human interactions, where repeated exposure to AI predictions creates feedback loops that incrementally influence human decision-making. Through purposive sampling, the research engages 15 young adults (aged 18–30) who regularly use AI systems, employing semi-structured interviews to explore how this cascade effect manifests in daily decision-making processes. The study utilizes thematic analysis and enables the investigation of how cognitive biases (anchoring, confirmation, and automation) develop and cascade through repeated AI interactions and how personality traits, particularly assertiveness and agreeableness, moderate this cumulative influence. The research aims to uncover patterns in trust development, bias awareness, and AI reliance over time while identifying effective strategies for maintaining autonomous decision-making capabilities. The study revealed that cognitive bias creates a cascade effect in AI-human interaction, where cognitive bias affects trust and dependency, with assertiveness moderating the bias-trust relationship and agreeableness moderating the bias-dependency relationship

Keywords: *cascade effect, cognitive bias, human-AI interaction, personality traits, decision-making.*

1. Introduction

The growing integration of artificial intelligence (AI) into decision-making processes across diverse domains ranging from healthcare diagnostics to personnel selection has brought substantial improvements in decision efficiency and accuracy. However, these benefits are accompanied by unintended cognitive and behavioral consequences. Human interaction with AI often triggers and amplifies cognitive biases, including anchoring bias (the tendency to rely too heavily on initial information) and confirmation bias (the predisposition to validate pre-existing beliefs), and automation bias (an over-reliance on AI outputs without adequate scrutiny). (Nourani et al., 2021) These biases do not act in isolation but frequently interact in cascading and reinforcing ways, altering decision heuristics, compounding errors, and fostering dependency on AI systems.

A slight ripple in our thoughts, feelings, or actions can start an immense chain reaction that reshapes our lives and the world around us; this is the fascinating phenomenon known as the cascade effect in psychology. (Neurolaunch, 2024). The cascade effect in the context of humans interacting with AI, where cognitive biases are involved, refers to a self-reinforcing cycle where initial human-AI interactions amplify certain cognitive

biases. This occurs when AI systems, designed based on historical data or patterns, provide feedback that aligns with or reinforces pre-existing human biases, leading to a loop where the biases are magnified over time.

In such scenarios, the AI can inadvertently validate a user's biased perceptions or decisions, leading the human to trust the system more and become more reliant on its outputs. As a result, the user may ignore or undervalue alternative perspectives or critical thinking. This loop can create a cognitive cascade where errors in judgment or decision-making become more pronounced and cognitive biases such as confirmation bias, anchoring, automation bias, reducing objectivity, and increasing the risk of poor decisions.

According to Salter (2002) and Kapponi and Novak (1995), assertiveness is a personal quality that can be characterized as autonomy, independence from outside judgments and influences, and the capacity to autonomously control one's own behavior. Characteristics of an assertive personality include: orientation to real-world events, where both the past and the future make sense; separateness of behavior and values from outside influences; autonomy in forming one's own life views; openness and freedom of expression of one's own potential; confidence in oneself and trust in others (Kapponi and Novak, 1995). Being assertive entails being self-initiating and self-regulating, according to Bandura (1986). The confidence in one's own efficacy, which is founded on self-worth and self-esteem, is a sign of assertiveness.

As per the Big Five Factor Theory, agreeableness relates to the degree to which an individual maintains pleasant, harmonious interpersonal relationships and acts in a prosocial manner towards others. Compassion (as opposed to disregard for others), civility (as opposed to hostility), and trust (as opposed to suspicion of others) are important aspects of agreeableness. People with high agreeableness are more likely to treat others with respect, aid them, and forgive them; people with low agreeableness are more likely to belittle others, argue, and harbour resentment. (Soto,2016).Studies have found that agreeableness plays a vital role in establishing trust and cognitive load conditions (Zhou, 2020)

1.1 Cognitive Bias Cascades in AI-Human Interaction

AI interactions often exacerbate anchoring bias, where human decision-makers use initial AI outputs as fixed reference points for subsequent judgments, leading to systematic errors. (Nourani et al. 2021) explored how early positive impressions of an AI system's strengths anchor user trust, which in turn reduces scrutiny and fosters automation bias. Similar findings were observed by Echterhoff et al., who found that decision-makers often anchor to sequential AI recommendations in decision tasks, with significant impacts on outcome consistency. Moreover, the reliance on these anchors triggers confirmation bias, as individuals are more likely to seek out or emphasize information that validates AI suggestions aligned with their pre-existing beliefs. This

phenomenon is particularly prominent in high-stakes or tightly time-constrained contexts, as demonstrated by Rosbach et al.,(2024), who quantified confirmation bias driven by AI-induced false confirmations in a pathology context. Once anchoring and confirmation biases take hold, automation bias further compounds the problem by encouraging unchecked trust in AI, often at the expense of critical reasoning. Cabitza et al. conceptualized this dynamic as “technological dominance,” wherein users place undue reliance on AI recommendations, even when external evidence or context contradicts those outputs. These interactions highlight how cognitive biases triggered by AI can cascade into compounding errors, gradually embedding themselves into users' decision-making processes and diminishing independent judgment.

1.2 Mitigation Strategies and the Role of Explainable AI (XAI)

The rise of explainable AI (XAI) offers promising avenues to address cognitive bias cascades by improving user understanding and fostering scrutiny of AI-generated recommendations. Studies have explored how XAI can mitigate the effects of anchoring, confirmation, and automation biases through increased transparency. For instance, Haag et al.(2024) demonstrated that combining XAI explanations with AI recommendations significantly reduces anchoring bias in purchase-related decision-making, while providing confidence scores and localized explanations helps calibrate trust in AI systems. (Zang, 2020) These interventions ensure that users appropriately reconcile AI recommendations with their own domain expertise, thereby reducing reliance on anchors and curbing automation bias. However, the effectiveness of XAI remains mixed and context-dependent. Schemmer et al. observed that while XAI explanations mitigated automation bias in certain decision-making scenarios, they sometimes inadvertently increased reliance on flawed AI models in the absence of accessible error cues. This duality highlights the need for carefully designed XAI frameworks that account for task context, user expertise, and individual traits.

In addition to XAI, other mitigation strategies targeting specific biases have been examined. For example, Echterhoff et al. (2023) proposed real-time algorithmic interventions, such as reordering decision options, to reduce sequential anchoring bias, which improved decision accuracy in both experimental and real-world applications. Adaptive trust calibration strategies, which dynamically adjust user confidence in AI outputs based on both human and AI likelihoods of correctness, have also proven effective in reducing automation bias in tasks requiring collaborative decision-making (Okamura et al.,2020), (Wang et al., 2023). However, while these approaches demonstrate measurable success, they often fall short of addressing the cascading nature of multiple biases or user-specific vulnerabilities.

1.3 Personality Traits and Individual Differences:

Individual differences, particularly personality traits like assertiveness and agreeableness, play a significant role in moderating the relationship between AI interaction, cognitive biases, and decision dependency. Assertiveness, characterized by confidence and resistance to external inputs, has been associated with reduced susceptibility to automation bias but increased vulnerability to confirmation bias. For instance, a study designed two approaches to communicate the decisions of an intelligent agent for breast cancer diagnosis with different tones: a suggestive (non-assertive) tone and an imposing (assertive) one. It was found that assertiveness plays an important role in how information is perceived. (Calisto et al., 2023) Agreeableness, on the other hand, is linked to higher trust in AI, making agreeable individuals more prone to automation bias. However, these users are also more accepting of corrective interventions and de-biasing efforts, such as feedback-based trust calibration (Kupfer et al., 2023). People in the UK population sample who scored higher in agreeableness are more likely to have positive attitudes toward AI (Babiker, 2024). In a study that examined how an individual's personality trait relates to attitude towards AI. It was found that agreeableness was associated with both positive and negative emotions, and it was positively associated with sociality and functionality (Park et al., 2021) Despite these findings, the influence of personality traits remains underexplored in quantitative models of bias cascades, leaving significant gaps in tailoring AI designs to specific user profiles.

1.4 Dependency Formation and Feedback Loops

Prolonged interaction with AI systems compounds cognitive biases and fosters dependency through feedback loops, whereby repeated AI recommendations progressively diminish user independence. This phenomenon, referred to as “cognitive offloading,” occurs when humans transfer decision-making responsibilities to AI systems, particularly in high-stakes or complex tasks where the perceived reliability of AI reduces psychological discomfort or uncertainty. The Bayesian model of information cascade discusses information cascades where agents follow others' decisions, potentially overriding their own information. (Srivathsan et al., 2021) In AI contexts, this can lead to users making decisions based on observed behaviors rather than personal insights, impacting collective outcomes and decision-making processes. Studies such as Cabitza et al. and Hondrich et al. have shown that dependency is reinforced in environments where users perceive AI as an “uncertainty buffer.” However, as dependency deepens, users are less likely to scrutinize AI outputs, amplifying automation bias and minimizing corrective opportunities. Feedback dynamics, such as the iterative nature of AI-human interaction observed by Echterhoff et al. (2023), further exacerbate this problem by embedding cognitive biases like anchoring into decision heuristics over time. These findings underscore the

importance of designing AI systems that actively disrupt these feedback loops through interventions such as re-alignment cues, periodic reliance checks, and adaptive trust recalibration (Okamura et al., 2020). . A study by xiu et al. (2023), has discussed the feedback effect, where unhelpful interactions lead to reduced future engagement, creating a feedback loop that diminishes user satisfaction and interaction diversity over time. However, there is a gap in the studies in addressing the cognitive biases

1.5 Justification for the present study

Despite the growing body of literature, significant gaps remain in understanding and addressing cognitive biases in AI-driven decision-making. First, the role of personality traits as moderators of bias dynamics and trust is underexplored, with limited integration of traits like assertiveness, agreeableness, and prior beliefs into computational models or experimental designs (Gurney et al., 2023). Second, while studies like Cabitza et al. and Nourani et al. (2021) examine bias cascades over iterative sessions, comprehensive longitudinal studies investigating dependency formation and bias persistence are sparse. Finally, the contextual impact of AI design, particularly the challenges posed by black-box models, opacity, and user perceptions of system reliability, warrants further exploration to ensure ethical and effective human-AI collaboration (Schemmer, 2022).

Understanding these cascading effects, the moderating influence of personality traits like assertiveness and agreeableness, and methods for mitigating bias is essential for developing ethical and robust AI-human collaboration frameworks. The present study aims to explore:

1. What role does cognitive bias play in the decision-making process of AI-human interaction?
2. How does cognitive bias create a cascade effect?
3. How do assertiveness and agreeableness moderate the relationship between cognitive bias and trust?
4. How do assertiveness and agreeableness moderate the relationship between cognitive bias and dependence?
5. What are the ways to mitigate the cognitive biases that arise due to AI-human interaction?

2. Method

2.1 Participants

The study comprised 15 young adults aged 18–30 years, all of whom were regular or frequent users of AI technology. A purposive sampling technique was utilized to identify and recruit participants who actively use AI assistance in their professional work and academic pursuits. This sampling approach ensured that participants could provide rich, detailed accounts of their experiences with AI systems. The participants were recruited from three South Indian states-Tamil Nadu, Kerala, and Karnataka-providing a regional perspective

on AI usage patterns. The sociodemographic profiles of the participants are presented in Table 1.

Table 1. Demographic characteristics of participants

Sl.No	Unique ID	Age	Sex	Education	Occupation	Residence
1	ID_01	23	M	Post Graduate	Fellowship	Puducherry
2	ID_02	26	F	Post Graduate	Unemployed	Kerala
3	ID_03	24	M	Post Graduate	Student	Kerala
4	ID_04	21	M	Higher Secondary	Student	Kerala
5	ID_05	24	F	Post Graduate	Assistant Professor	Tamil Nadu
6	ID_06	25	F	Undergraduate	Infrastructure Engineer	Kerala
7	ID_07	23	M	Post Graduate	Social worker	Kerala
8	ID_08	19	F	Higher Secondary	Student	Tamil Nadu
9	ID_09	26	M	Post Graduate	Unemployed	Kerala
10	ID_10	25	F	Post Graduate	Assistant Professor	Tamil Nadu
11	ID_11	26	F	Post Graduate	Assistant Professor	Tamil Nadu
12	ID_12	23	F	Higher Secondary	Student	Tamil Nadu
13	ID_13	24	F	Post Graduate	Scholar	Karnataka
14	ID_14	25	F	Post Graduate	Financial Manager	Karnataka
15	ID_15	23	F	Post Graduate	Scholar	Karnataka

2.2 Inclusion criteria

1. Participants must be regular or frequent users of AI technologies, either in their personal or professional lives.

2. Participants were required to be young adults between the ages of 18 and 30.
3. Participants must be able to communicate clearly in either English or Malayalam

2.3 Procedure

Semi-structured interviews served as the primary data collection method. Three pilot studies were conducted, and the interview questions were modified according to the objective of the study. Three independent reviewers evaluated the interview questions to validate their alignment with the study's aims. The questions for measuring assertiveness were adapted from the Assertiveness Inventory (Alberti et al.) and modified to suit the purpose of the study. Participants were asked to self-report their responses to each question on a scale of 0 (Never) to 3 (Always). At the end, they were also asked, "On average, how assertive are you in your life?". Similarly The questions for measuring agreeableness were adapted from the Big Five Factor Inventory and modified according to the purpose of the study, where participants self-reported their responses to each question on a scale of 1 (strongly disagree) to 5 (strongly agree), followed by a final question: "On average, how well do you get along with others and have a successful relationship in your life?". The recruitment process utilized social media channels, where digital announcements and flyers were circulated. Potential participants who showed interest were subsequently contacted for further screening. The interviews were conducted in English and Malayalam, depending on the participant's preference, to ensure comfortable and authentic communication. Among 15 participants, three interviews were conducted face-to-face, while the rest 11 interviews were telephonic interviews. Each interview session lasted approximately 20–30 minutes and was digitally recorded with participant consent. All interviews were conducted during the month of January 2025. The interview protocol was designed to explore participants' experiences, perceptions, and attitudes toward AI technology while remaining flexible enough to pursue emerging themes and insights. With consent, the interview sessions were audio recorded and then manually transcribed. In order to maintain participant privacy, their names were hidden by giving every individual a distinct ID.

2.4 Data Analysis

Six-phase data analysis methods proposed by Braun & Clarke (2006) were followed in the study. Three separate coders worked on the data analysis procedure. First, to familiarize themselves with the data and concepts, each coder thoroughly studied the transcribed data independently. Second, using the conventional paper-pen method, coders began classifying the data by carefully reading over the verbatim transcripts, noting pertinent quotes, and simultaneously assigning codes. Eight themes and 5 sub-themes were identified. The study has incorporated two scores for assertiveness and agreeableness: 1) The average of item scores

calculated by the researchers, and 2) The self-reported average score given by the participant at the end is considered. These scores are then compared and analysed

2.5 Ethical Consideration

The study implemented ethical safeguards through a verbal informed consent process. Each participant received a clear explanation of the study's voluntary nature and their right to discontinue participation at any point. To protect participant privacy, no identifying information was collected, and all data was anonymized. Participants were assured of data confidentiality, with access restricted to the research team and usage limited to research purposes only.

3. Results

Thematic analysis was conducted, and the results were organized into themes, sub-themes, and codes. Each theme is explained with its operational definition and illustrative quotations from the data in the result section. Verbatim responses from participants are labeled with unique IDs, age, and gender, where “M” represents male and “F” represents female

3.1 Cognitive bias

This theme captures the systemic inaccuracy in thinking that influences how we absorb information, perceive others, and make decisions. The study has identified three main cognitive biases: Automation bias, Anchoring bias and confirmation bias

3.1.1 Automation bias

Automation bias is the predisposition for humans to favor ideas from automated decision-making systems and to reject contrary information generated without automation, even if it is right. Participants have reported that they rely on AI information as it saves time and effort regarding making their job easy. Participants have also reported that AI was helpful in diagnosis before the doctor could do so.

AI was able to diagnose that I had an IBD before the doctor was able to do it. (ID_04/21/M)

While I am in Chennai, I had some problem with my eyes, and based on that, a doctor diagnosed me with auto-immune disease, like because I had redness in the eyes for 3 times repeatedly, so the doctor said it will not be like that for any other reason; it will be some auto-immune issues. After that, I asked Chat GPT about its symptoms and the thing that I was diagnosed with and asked why it is, and Chat GPT also said it is an auto-immune disease. (ID_10/25/F)

3.1.2 Anchoring bias.

This theme highlights the people's tendency to depend excessively on the first piece of information they receive about an issue. Participants have reported that they have used the information given by AI without cross-verifying on many occasions, such as related to work, drafting a mail, correcting grammar, etc.

Like not only me, all my coworkers also, if they have some blocks somewhere or if they need to take a further step, immediately they are using not Google but Chat GPT, so its usage is high nowadays. (ID_06/25/F)

3.1.3 Confirmation bias

This theme captures the human tendency to process information by seeking out or interpreting details that align with their pre-existing beliefs. Participants have reported that they have changed the prompts to get their desired response. They reported feeling temporary relief from seeking confirmation, but realizing the AI's initial information was wrong led to distress. This reinforced their belief and created a cycle of repeatedly seeking confirmation.

I thought I would rather know if somebody liked me and I repeatedly asked the AI, was it sure that the guy liked me? I asked it multiple times till I got an answer that seemed exactly like mine. (ID_12/23/F)

Recently my friend invited me for his marriage and I didn't know what to say. I was getting in touch with him after a long time, so at that time I didn't know what to say, so when I asked AI for a reply first, it gave an answer, and then I gave an extra prompt like this is my friend, so I don't want a formal reply like that. (ID_09/26/M)

3.2 Cascade effect

This theme focuses on the small shift in human thoughts, emotions, or actions that can trigger a significant chain of events, transforming lives and the perception of the world. Automation bias initiates a chain reaction, leading to anchoring and confirmation biases, which subsequently result in negative psychological outcomes like reduced self-esteem, self-doubt, decreased problem-solving ability, and distress. Participants have reported that relying on the information provided by AI has impacted their overall problem-solving ability and self-esteem.

I have to give my student some reading activities, like some chapters from the text book or page number-wise,

and it should be related to the topics covered on each day, so I have to create a plan as per this need. So to create this plan, I was not able to read the entire book, so I uploaded the book to AI and asked it to give me the page numbers according to the topics and titles; and for that, AI has given me, say, if it is a 32-session-long activity, then it gave me a 32-session-long activity, and after that, when I compared it with the book, it was entirely different from it. I have already submitted the docs to my boss and I was so scared that It might lose my job and was really in a distress (ID_ 11/26/F)

3.3 Awareness

This theme explores participants' awareness of the role AI plays in their lives and how much it influences their decision-making processes. Participants reported using AI assistance for daily tasks, particularly for formal writing, and often accepting the AI-generated responses without much scrutiny. Many respondents did not recognize this as part of a decision-making process and seemed to be in denial about the extent of AI's influence. Those who acknowledged AI's role described it as an inevitable part of their routine, suggesting a sense of reliance and normalization of AI in their day-to-day activities.

When a decision-making situation comes, like I said that I am using it for work-related things and there are some processes and workflow that need to be followed there, so since I have these processes and I have a clarity that I have to follow these processes, I was not really dependent on AI for the decision-making (ID_06/25/F).

Like many things I asked AI what to do, and most of those cases it influenced me; there was a treasure hunt event in my college as part of the tech fest, so for that I gave a general idea, like it was a big prompt, and with that prompt it gave me so many suggestions, so yes, it does influence my decision-making (ID_04/21/M).

3.4 Trust and skepticism

This theme explores the level of trust placed in AI by the participants. It was found that most of the participants are skeptical towards the suggestions given by AI, and this is due to their personal experiences of getting wrong information from AI, especially while doing mathematical calculations and answering logical questions, some responses are:

When I was preparing for exams, I asked some questions to AI, and the responses I received were not even in the options that were provided with the questions. Then after that, I again put the questions with the options, and these. So I feel it is not that trustworthy (ID_01/23/M).

Almost 75% of the time it was correct. Sometimes I use it to answer question papers for competitive exams, and sometimes for the same question, one time it will give one answer and another time it will give another answer, so that time I had to use search engines like Google (ID_03/24/M).

In the case of less trustability, recently I asked a math reasoning question to AI and the answer it gave was wrong, and then I asked it like this is wrong, so then it said that yes, it is wrong; this is the right answer like that. So there are mistakes in it; I lost the truest when it made a mistake in math (ID_09/26/M).

Even though some respondents have high agreeability and low assertiveness, they are also not trusting AI 100%; they have also shown a slight skepticism but still believe the chances of AI making such a mistake are much rarer than a human.

If I know an answer to a question very well, like in math, I know what it says my judgement is correct. If we take a hypothetical scenario where the AIs answer is different than mine, so since I know the concept very well, I will ask it why you responded like that, and its response is not justifiable because I know the topic, I can understand what it says. If it can factually contradict me, it is right; if it misinterprets my prompt, I can correct it (ID_04/21/M).

3.5 Dependency

This theme explores the dependency of participants towards the AI models, and it was understood that every participant on a certain level depends on AI for their tasks in personal and professional lives, and a relationship between the participants' trust in AI systems and their dependency on AI systems was able to be observed. To justify this relationship, the participants were divided into 4 groups: 1) Participants with low trust and low dependency 2) Participants with low trust and high dependency 3) Participants with high trust and low dependency 4) Participants with high trust and high dependency.

By classifying participants into these four groups based on their trust and dependency scores, it was found out that people with high trust in AI systems are highly dependent on AI, and people with low trust in AI are less dependent on AI. There was also a case of people with low trust and high dependency on AI, and researchers were able to identify 2 participants in this criteria, but it was justified as their full-time jobs demand seeking the support of AI models to complete tasks in their professional lives. And one more notable finding here is that there was no respondent in the high trust and low dependency category, which indicated that people who have a high level of trust in AI systems are dependent upon AI models for their daily activities.

3.6 Moderating role of personality traits

This theme highlights how assertiveness moderates the relationship between cognitive bias and trust or dependency, offering insight into how individuals' confidence in expressing their thoughts may influence their susceptibility to biases and their reliance on artificial intelligence. Assertiveness is the quality of expressing one's thoughts, feelings, or needs in a confident, direct, and respectful manner while maintaining consideration for the rights and opinions of others

It also explores how agreeableness moderates the relationship between cognitive bias and trust or dependency. Agreeableness is a core personality trait that reflects an individual's abilities to get along well with others and their concern for social harmony.

3.6.1 Moderating role of Assertiveness

It was found that participants who exhibited high assertiveness, even though they had cognitive bias, had low trust towards the AI models compared to those with low assertiveness. This relationship suggests that individuals with higher assertiveness traits maintain their independent decision-making capabilities and are more likely to question AI recommendations, regardless of any existing cognitive biases. An illustrative verbatim of a participant with high assertiveness and low assertiveness is as follows:

I only have a certain level of trust towards AI; I don't have 100% blind trust towards it, and the information we get from AI is the collective of a lot of information, and I feel it is capable of doing that. Also, I only use the information from AI after double-checking, and if it is not accurate, I don't use it, so I don't have a blind trust in it. (ID_07/23/M) (Avg. Assertiveness score : 3)

I trust it more than a human telling something, but I know for a fact that it is not mostly right because if there is a programming-related thing I can figure out it is outdated, because AI is trained on 6 months previous data, so anything that happened during that time, any updates in programs, everything certain packages change that stuff, AI is not fully updated, but there are certain AIs for coding and medical purposes so they will have up-to-date data; they are quite good. An AI that is dedicated to a particular field is the same as a PHD student because it is fine-tuned and trained for that specific task to get a 10/10; it is trained like that.(ID_04/23/M) (Avg. Assertiveness score : 1)

3.6.2 Moderating role of Agreeableness

It was found that participants who exhibited high agreeableness, even though they had cognitive bias, had high trust and dependency towards the AI models when compared to those with low agreeableness. This indicates

that highly agreeable individuals tend to be more accepting of AI suggestions and are more likely to develop reliance on AI systems, despite the presence of cognitive biases in their decision-making process.

I have noticed so many changes; I think I am overly dependent on AI to make my decisions. Like some decisions, I will ask AI's suggestions before I make them. (ID_10/25/F) (Avg. Agreeableness score: 4)

Over time, I don't feel like it's getting more like I am able to trust AI more. One because there are like fewer glitches, and now that I am using AI more, I am kind of sure I know what input to give to get more of an accurate result, and I also know which paths to cross-verify and like which commands to give to so that I get more of an answer I am willing to use. (ID_08/19/F) (Avg. Agreeableness score : 2)

3.7 Impact on mental health

This theme examines how intensive AI use affects participants' mental wellbeing. Participants reported feeling inferior to AI models, particularly regarding language and writing abilities. They experienced self-doubt, diminished self-esteem, and sought constant validation. Some noted reduced social skills and growing dependence on AI for conversation and problem-solving, reducing and replacing human interaction.

Created a level of self-doubt and also lowered my self-esteem, overall vocabulary, and memorizing capacity; everything has lowered. Before that, I used to learn new words to make my English language more attractive, but now it is not there, and memorising something that all has been so difficult now because we have sources for that. (ID_10/25/F)

I ask Chat GPT for answers, or for even instances when I am down; I even share my feelings, and it gives me guidance too, and also programming; it helps me in my programming and apps developing everything like that. It is more secure and more private than an actual therapist, technically, and it is also free (ID_04/21/M).

AI is actually a good tool, because sometimes if we feel that we are alone, while chatting and all, we will have someone there to talk, so I felt that. Also, we get so much information from AI regarding every aspect of a topic (ID_02/26/F).

3.8 Strategies

Participants employed distinct strategies for maintaining independent judgment while interacting with AI systems, with their approach correlating to their level of trust in the technology. Those expressing skepticism implemented verification-focused strategies to maximize accuracy, while participants with higher trust levels

oriented their strategies toward obtaining personally appealing responses. The predominant strategies that emerged from participant responses were cross-referencing information and developing deep subject matter understanding. Additionally, participants frequently cited consulting domain experts and clearly defining their own requirements as important complementary approaches.

Whenever we are searching for something, we must have a basic understanding about it. If we are only taking information, fine, but if we want to take a decision based on that, before checking with the AI, we must have some understanding of ourselves about what we want. If the AI is deviating that route, ok, let's see if we can use it, but ultimately it should be our choice based on what we want (ID_09/26/M).

I guess cross-processing is bumping, better prompt, but AI, perhaps using two or three different kinds of AI, different things, books that I have, or perhaps if it's like a decision-making thing, perhaps asking people who are more experienced than me. (ID_11/26/F).

4. Discussion

This study aims to explore areas in social psychological practice: how AI-human interaction creates a cognitive bias in humans and how personality traits moderate it. The study contributes to the knowledge in the area of socio-psychological aspects of AI-human interaction by exploring how human personality traits like agreeableness and assertiveness moderate the relationship between cognitive bias and AI dependency.

The participants' reliance on AI-generated content appears to reflect a pattern, suggesting that positive initial experiences with AI may encourage a less skeptical and more dependent approach to health information seeking. As the digital world has flourished, individuals increasingly rely on online sources for information, embodying the concept of having "the world at one's fingertips." A notable shift has emerged, with a preference for AI-based searches over traditional resources such as books or evidence-based studies, indicating a growing dependency on AI-generated content. Tasks that once required human effort, such as drafting letters, correcting grammar, or completing assignments, are now frequently delegated to AI through simple commands, potentially diminishing essential human competencies. Findings of this study indicate that participants frequently turned to AI as a convenient tool to save time and effort in understanding their health conditions, particularly when doctors were unable to provide a clear diagnosis. This behaviour underscores the growing curiosity and self-reliant approach of individuals in health-related decision-making, fostering an attitude toward self-diagnosis. These results align with Nourani et al. (2021), who found that early positive impressions of an AI system's capabilities establish user trust, which subsequently reduces critical evaluation and fosters automation bias. Participants frequently described AI as influencing their decisions by serving as

a primary source of information and guidance. Nourani et al. (2021) highlight how AI interactions often exacerbate anchoring bias, where initial AI outputs become fixed reference points for decision-making, leading to systematic errors. Similar findings were observed by Echterhoff et al., who found that decision-makers frequently anchor to sequential AI recommendations during decision-making tasks, significantly impacting outcome consistency. These insights emphasize the need for critical engagement with AI to balance its convenience with the preservation of human judgment and skills.

The findings of this study reveal an intriguing pattern in participants' interactions with AI, particularly when they encountered incongruent information. Initial mistrust in AI-generated content prompted participants to reframe their queries from different perspectives to obtain a desired response. This aligns with the literature on anchoring and confirmation biases, where individuals fixate on initial information and seek evidence to validate their pre-existing beliefs. Once these cognitive biases take hold, automation bias can further reinforce this tendency, leading to unchecked trust in AI, even at the expense of critical reasoning. Cabitza et al. (2021) conceptualized this phenomenon as "technological dominance," where users place undue reliance on AI-generated recommendations, even when external evidence or context contradicts those outputs. This reliance was evident in participants' reported behaviors, such as altering prompts until they received confirmation aligned with their expectations. One participant exemplified this pattern by persistently asking whether someone liked them until the AI provided a response that matched their desired belief. While this behavior initially offered temporary relief, participants reported distress upon realizing that the AI's initial information was inaccurate. This reinforced a cyclical pattern of seeking confirmation, contributing to a loop of dependency and mistrust. These findings underscore the psychological implications of AI interactions, highlighting the complex interplay between user cognition, emotional responses, and AI-driven information-seeking behaviour. They emphasize the importance of fostering digital literacy and critical thinking to navigate AI-generated content responsibly, as well as the need for further exploration into strategies that help users break free from confirmation-seeking cycles. The findings of this study illustrate a cascade effect in participants' interactions with AI. Initial experiences of incongruent or inaccurate information generated feelings of mistrust, prompting participants to modify their prompts in search of more satisfying answers. This iterative process aligns with Cabitza et al.'s (2021) concept of "technological dominance," where users place undue reliance on AI despite contradictory evidence. As participants continued seeking confirmation, cognitive biases such as anchoring and confirmation biases further entrenched this behavior, leading to automation bias and diminishing critical reasoning. This cascade effect was evident in reports where participants repeatedly sought AI validation, even for subjective questions such as whether someone liked them, until receiving a response that aligned with their desires. Although this behavior offered temporary

relief, participants later experienced distress when realizing the initial AI-generated information was inaccurate, perpetuating a cycle of dependency and dissatisfaction. These findings underscore how AI interactions can trigger a chain of cognitive and emotional responses, reinforcing user reliance on AI-generated content. Addressing this issue requires fostering digital literacy and self-regulation strategies to mitigate the cascade effect and promote responsible engagement with AI. This theme also demonstrates how small shifts in behaviour, such as initial trust in AI, can set off a cascade effect with significant psychological consequences. In this study, participants' dependence on AI triggered a series of cognitive biases, including anchoring and confirmation bias, which compounded the problem of automation bias, ultimately resulting in negative psychological outcomes such as reduced self-esteem, self-doubt, and distress.

The study explores how people are becoming more aware of AI's influence on their daily lives and decision-making. Many use AI for routine tasks like writing and often accept its suggestions without much thought. Some don't realize how much AI affects their decisions, while others acknowledge its impact and rely on it for everyday tasks. A key difference emerged: some individuals use AI only for specific tasks, such as following established rules in work-related activities, thereby limiting its influence on their decision-making. In contrast, others rely heavily on AI for more creative tasks like brainstorming and idea generation, showing how AI can significantly shape decisions in diverse ways. This highlights the complex relationship between AI and decision-making, emphasizing the need for individuals to be more mindful of AI's subtle yet significant impact on their choices. Some participants' experiences with misleading information from AI, many began to cross-check the generated content to ensure its accuracy. This shift towards verification reflects a growing skepticism towards AI's reliability as users became more aware of its limitations. While AI can provide valuable insights and assistance, participants recognized the importance of validating the information independently to mitigate the risk of making decisions based on inaccurate or incomplete data. This highlights the evolving relationship between users and AI, where trust is contingent on the ability to critically assess and verify the information it provides.

The findings of this study reveal a significant relationship between participants' trust in AI systems and their dependency on these systems. Notably, every participant exhibited some level of dependency on AI in their personal and professional lives. Participants were categorized into four groups based on their trust and dependency scores to explore further. The results showed that individuals with high trust in AI systems tended to be highly dependent on them, whereas those with low trust exhibited lower dependency. Two participants fell into the category of low trust and high dependency. However, this apparent paradox was attributed to the demands of their full-time jobs, which required them to rely heavily on AI models to complete tasks. Another striking finding was the absence of participants in the high trust and low dependency category. This suggests

that individuals who have a high level of trust in AI systems inevitably become dependent on them for their daily activities. This study highlights the significant role of assertiveness in shaping individuals' perceptions and interactions with AI-generated information. Assertive individuals tend to question AI's validity, whereas agreeable individuals exhibit higher trust in AI but are also more receptive to corrective interventions. Interestingly, assertiveness and agreeableness were positively correlated. Highly assertive participants demonstrated lower trust in AI models, suggesting they maintain independent decision-making capabilities. A participant's response exemplifies this cautious approach: "I don't have 100% blind trust towards AI... I only use the information from AI after double-checking." This underscores the influence of personality traits on trust and decision-making processes. This study's findings align with previous research (Babiker, 2024), which discovered that individuals with higher agreeableness scores in a UK population sample tended to have more positive attitudes toward AI. Similarly, this study revealed that participants with high agreeableness exhibited high trust and dependency on AI models, despite cognitive biases. This consistency suggests that agreeableness is a significant predictor of positive attitudes and trust in AI, highlighting the importance of considering personality traits in the development and implementation of AI systems. Excessive AI use was found to negatively impact mental well-being, particularly self-esteem and confidence. Participants felt inferior to AI models, especially in language and writing abilities, leading to reduced self-worth and motivation for self-improvement. This reliance on AI also diminished human interaction and social skills, fostering a sense of inadequacy and social isolation. Participants employed distinct strategies to maintain independent judgment when interacting with AI. Skeptical individuals cross-referenced information and consulted experts to ensure accuracy, while those with higher trust in AI sought confirmation or appealing answers. A common strategy was having a foundational understanding of the subject matter before consulting AI, emphasizing the importance of maintaining an independent framework of knowledge and critical thinking. This study offers new insights into the relationship between trust, dependency, and personality traits in the context of AI interactions. One notable finding is that trust and dependency on AI content are directly proportional—participants with higher trust in AI models also demonstrated higher dependency on the information provided. This relationship underscores how early positive experiences with AI can foster reliance, echoing previous findings by Nourani et al. (2021). However, the present study adds a nuanced understanding by examining the moderating roles of assertiveness and agreeableness.

Participants who exhibited high agreeableness, despite cognitive biases, demonstrated higher trust and dependency on AI models compared to those with low agreeableness. This suggests that agreeable individuals are more inclined toward harmonious and trusting interactions with AI, making them more susceptible to automation bias. Agreeableness, characterized by a strong concern for social harmony and a tendency to avoid

conflict, appears to make individuals more accepting of AI-generated content without rigorous evaluation. Conversely, assertiveness emerged as a protective factor, moderating the relationship between cognitive bias and AI trust or dependency. Assertive individuals, defined by their ability to express thoughts, feelings, and needs confidently while respecting others' perspectives, maintained a greater degree of independent decision-making. They were more likely to question AI recommendations, even in the presence of cognitive biases, highlighting their resilience against automation bias.

These findings contribute to a deeper understanding of how personality traits can shape user-AI interactions. They emphasize the importance of fostering assertiveness to encourage critical engagement with AI-generated content, while recognizing that agreeable individuals may benefit from interventions aimed at enhancing their skepticism and decision-making independence.

4.1 Limitation and Future Direction

This study has several limitations. First, as a qualitative investigation with only 15 participants, the findings may not be broadly generalizable. While the study provided initial insights, deeper exploration is needed to fully understand the phenomena observed. Future research should consider quantitative approaches to examine the moderating effects of personality traits like assertiveness and agreeableness on cognitive bias and AI dependency. Additionally, the relationship between age and AI dependency merits investigation, as does the interplay between social interaction patterns and reliance on AI systems

4.2 Implication

This pioneering study is one of the few that examines three critical aspects of human-AI interaction, like AI-induced cognitive biases, their cascade effects, and the moderating role of personality traits, using a qualitative approach. As AI continues to rapidly evolve and integrate into all facets of human life, these findings provide valuable groundwork for future research in AI-human interaction and dependency patterns. The study's insights into how personality traits like agreeableness and assertiveness moderate cognitive biases and cascade effects offer a foundation for developing quantitative models and larger-scale investigations. These findings have significant implications for both AI development and mental health fields, particularly in understanding and addressing the psychological impacts of AI integration in daily life

5. Conclusion

The study revealed several key findings about AI dependency and human behavior. Cognitive biases significantly influenced participants' decision-making processes, with a notable cascade effect in their AI

interactions. Personality traits emerged as important moderating factors: assertiveness moderated the relationship between cognitive bias and trust in AI systems, while agreeableness influenced how cognitive biases translated into AI dependency. Participants with higher agreeableness showed greater susceptibility to developing trust and dependency on AI systems when affected by cognitive biases. The findings also highlighted the detrimental effects of cognitive biases and AI dependency on user behavior and decision-making capabilities

References

1. Do personality traits impact the attitudes towards artificial intelligence? (2024, August 16). IEEE Conference Publication | IEEE Xplore. <https://ieeexplore.ieee.org/abstract/document/10780777>
2. Echterhoff, J. M., Sen, B., Ren, Y., & Gopal, N. (2023, September 29). Should you make your decisions on a WhIM? Data-Driven Decision making using a What-If Machine for Evaluation of Hypothetical Scenarios. arXiv.org. <https://arxiv.org/abs/2309.17364>
3. Gurney, N., Miller, J. H., & Pynadath, D. V. (2023). The Role of Heuristics and Biases during Complex Choices with an AI Teammate. *Proceedings of the AAAI Conference on Artificial Intelligence*, 37(5), 5993–6001. <https://doi.org/10.1609/aaai.v37i5.25741>
4. Haag, F., Stingl, C., Zerfass, K., Hopf, K., & Staake, T. (2024, May 8). Overcoming anchoring bias: The potential of AI and XAI-based decision support. arXiv.org. <https://arxiv.org/abs/2405.04972>
5. Kupfer, C., Prassl, R., Fleiß, J., Malin, C., Thalmann, S., & Kubicek, B. (2023). Check the box! How to deal with automation bias in AI-based personnel selection. *Frontiers in Psychology*, 14. <https://doi.org/10.3389/fpsyg.2023.1118723>
6. Ma, S., Lei, Y., Wang, X., Zheng, C., Shi, C., Yin, M., & Ma, X. (2023). Who Should I Trust: AI or Myself? Leveraging Human and AI Correctness Likelihood to Promote Appropriate Trust in AI-Assisted Decision-Making. *CHI '23: Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, 1–19. <https://doi.org/10.1145/3544548.3581058>
7. McCarthy, M. H., Wood, J. V., & Holmes, J. G. (2017). Dispositional pathways to trust: Self-esteem and agreeableness interact to predict trust and negative emotional disclosure. *Journal of Personality and Social Psychology*, 113(1), 95–116. <https://doi.org/10.1037/pspi0000093>

-
8. Nourani, M., Roy, C., Block, J. E., Honeycutt, D. R., Rahman, T., Ragan, E., & Gogate, V. (2021a). Anchoring Bias Affects Mental Model Formation and User Reliance in Explainable AI Systems. *ACM Digital Library*, 340–350. <https://doi.org/10.1145/3397481.3450639>
 9. Nourani, M., Roy, C., Block, J. E., Honeycutt, D. R., Rahman, T., Ragan, E., & Gogate, V. (2021b). Anchoring Bias Affects Mental Model Formation and User Reliance in Explainable AI Systems. *IUI '21: Proceedings of the 26th International Conference on Intelligent User Interfaces*, 340–350. <https://doi.org/10.1145/3397481.3450639>
 10. Okamura, K., & Yamada, S. (2020). Adaptive trust calibration for human-AI collaboration. *PLoS ONE*, 15(2), e0229132. <https://doi.org/10.1371/journal.pone.0229132>
 11. Park, J., & Woo, S. E. (2022). Who Likes Artificial Intelligence? Personality Predictors of Attitudes toward Artificial Intelligence. *The Journal of Psychology*, 156(1), 68–94. <https://doi.org/10.1080/00223980.2021.2012109>
 12. Postolati, E., a & Moldova State University, Chişinău, Republic of Moldova. (2017). ASSERTIVENESS: THEORETICAL APPROACHES AND BENEFITS OF ASSERTIVE BEHAVIOUR. In *Journal of Innovation in Psychology, Education and Didactics: Vol. Vol. 21 (Issue No. 1, pp. 83–96)*. https://jiped.ub.ro/wp-content/uploads/2017/09/JIPED_21_1_2017_7.pdf
 13. Rosbach, E., Ammeling, J., Krügel, S., Kießig, A., Fritz, A., Ganz, J., Puget, C., Donovan, T., Klang, A., Köller, M. C., Bolfa, P., Tecilla, M., Denk, D., Kiupel, M., Paraschou, G., Kok, M. K., Haake, A. F. H., R, D. K. R., Sonnen, A. F.-., . . . Aubreville, M. (2024, November 1). When two wrongs don't make a right" -- Examining confirmation bias and the role of time pressure during Human-AI Collaboration in Computational Pathology. *arXiv.org*. <https://arxiv.org/abs/2411.01007>
 14. Schemmer, M., Kühl, N., Benz, C., & Satzger, G. (2022a, April 19). On the Influence of Explainable AI on Automation Bias. *arXiv.org*. <https://arxiv.org/abs/2204.08859>
 15. Schemmer, M., Kühl, N., Benz, C., & Satzger, G. (2022b, April 19). On the Influence of Explainable AI on Automation Bias. *arXiv.org*. <https://arxiv.org/abs/2204.08859>
 16. Soto, C. J., Kronauer, A., & Liang, J. K. (2015). Five-Factor Model of Personality. *The Encyclopedia of Adulthood and Aging*, 1–5. <https://doi.org/10.1002/9781118521373.wbeaa014>
 17. Srivathsan, S., Cranefield, S., & Pitt, J. (2022). A Bayesian model of information cascades. In *Lecture*

notes in computer science (pp. 97–110). https://doi.org/10.1007/978-3-031-16617-4_7

18. Trapezanides, A. (2024). Agreeableness: Dimension of personality or social desirability artifact? Canterbury.

https://www.academia.edu/1138715/Agreeableness_Dimension_of_personality_or_social_desirability_artifact

19. Xiu, Z., 1, Cheng, K.-C., 1, Sun, D. Q., 1, Lu, J., 1, Kotek, H., 1, Yuhan Zhang, McCarthy, P., 1, Klein, C., 1, Pulman, S., 1, Williams, J. D., 1, Apple, & Department of Linguistics, Harvard University. (2023). Feedback Effect in User Interaction with Intelligent Assistants: Delayed Engagement, Adaption and Drop-out. arXiv:2303.10255v2 [cs.HC] 18 Apr 2023.

<https://typeset.io/pdf/feedback-effect-in-user-interaction-with-intelligent-2fakfmdx.pdf>

20. Zhang, Y., Liao, Q. V., & Bellamy, R. K. E. (2020). Effect of confidence and explanation on accuracy and trust calibration in AI-assisted decision making. FAT* '20: Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency. <https://doi.org/10.1145/3351095.3372852>

21. Zhou, J., Luo, S., & Chen, F. (n.d.). Effects of personality traits on user trust in human–machine collaborations. *Journal on Multimodal User Interfaces*, 14(4), 387–400. <https://doi.org/10.1007/s12193-020-00329-9>.



Medtronic