

Association for Information Systems

AIS Electronic Library (AISeL)

ICIS 2024 Proceedings

International Conference on Information
Systems (ICIS)

December 2024

My Fault, Not AI's Fault. Self-serving Bias Impacts Employees' Attribution of AI Accountability

Guangyu Du

Karlsruhe Institute of Technology, guangyu.du@kit.edu

Sebastian Lins

Karlsruhe Institute of Technology, sebastian.lins@kit.edu

Ivo Blohm

University of St. Gallen, ivo.blohm@unisg.ch

Ali Sunyaev

Karlsruhe Institute of Technology, sunyaev@kit.edu

Follow this and additional works at: <https://aisel.aisnet.org/icis2024>

Recommended Citation

Du, Guangyu; Lins, Sebastian; Blohm, Ivo; and Sunyaev, Ali, "My Fault, Not AI's Fault. Self-serving Bias Impacts Employees' Attribution of AI Accountability" (2024). *ICIS 2024 Proceedings*. 27.

<https://aisel.aisnet.org/icis2024/aiinbus/aiinbus/27>

This material is brought to you by the International Conference on Information Systems (ICIS) at AIS Electronic Library (AISeL). It has been accepted for inclusion in ICIS 2024 Proceedings by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact elibrary@aisnet.org.

My Fault, Not AI's Fault. Self-serving Bias Impacts Employees' Attribution of AI Accountability

Completed Research Paper

Guangyu Du

Karlsruhe Institute of Technology
Karlsruhe, Germany
guangyu.du@kit.edu

Sebastian Lins

Karlsruhe Institute of Technology
Karlsruhe, Germany
sebastian.lins@kit.edu

Ivo Blohm

University of St. Gallen
St. Gallen, Switzerland
ivo.blohm@unisg.ch

Ali Sunyaev

Karlsruhe Institute of Technology
Karlsruhe, Germany
sunyaev@kit.edu

Abstract

The widespread diffusion of Generative Artificial Intelligence (GAI)-based systems offers many opportunities, but it is also accompanied by public scandals causing harm to individuals, markets, and society. Practice and research are calling for ensuring AI accountability. However, accountability attributions are challenging because multiple actors are involved in the development, operation, and usage. It remains unclear who do employees hold accountable when using a GAI-based system. This study examined how employees attribute accountability based on their tendencies to enhance and protect their self-concept. We ran an experiment with 466 participants and compared their successful outcomes and failures when using GAI-based systems. Our experiment revealed that under success and failure scenarios, employees attribute accountability predominantly to themselves but also share attribution with other actors. Our study enriches understanding of employees' perceptions of accountability from a multi-actor's perspective and aids in the establishment of accountability frameworks.

Keywords: Accountability, Generative Artificial Intelligence, Self-serving Bias

Introduction

The emergence and rapid development of artificial intelligence (AI) is shifting power in education, business, and healthcare (Yu et al., 2018). The leverage of deep learning models and the progress of machine learning algorithms has driven the development of Generative AI (GAI) (Jovanovic & Campbell, 2022). With its ability to generate novel texts, images, and sound based on user prompts, GAI has a stunning impact on research, daily life, and business (Jovanovic & Campbell, 2022). GAI tools such as ChatGPT, GitHub Copilot, and DALL-E have attracted widespread attention and show great potential at work (e.g., marketing and sales, human resource management, and employee optimization (Brynjolfsson et al., 2023)). Given these opportunities, employees will collaborate more strongly with AI during day-to-day business activities, like brainstorming, generating, or refining solutions to solve problems (Fui-Hoon Nah et al., 2023). For example, GitHub Copilot, a GAI tool generating code suggestions for programmers, has achieved reliable performance on technical coding questions and can provide natural language explanations of how the code it produces works (Nguyen & Nadi, 2022). Meanwhile, ChatGPT gives chances to increase work efficiency and productivity significantly (Peres et al., 2023).

Conversely, GAI can manifest legal, security, and ethical issues, such as discrimination (Fui-Hoon Nah et al., 2023) and hallucinations (e.g., fictitious information or factual errors (Dwivedi et al., 2023)). For example, ChatGPT showed gender biases when used by employees to write recommendation letters for job applications (Wan et al., 2023). Similarly, potential GAI misuse can happen, such as employees offering GAI-based recommendations without proper validation or oversight (Fui-Hoon Nah et al., 2023). To achieve long-term sustainable use and mitigate these issues, it becomes increasingly important to meet accountability in GAI-based systems (Thiebes et al., 2021). In general, AI accountability requires actors to explain and justify their development, use, or decisions regarding the AI system and respond to interrogations by forums (e.g., regulatory bodies) (Bovens, 2007; Nguyen et al., 2024; Wieringa, 2020). AI accountability then ensures that sanctions can be applied to punish actors as well as provide redress for those adversely affected, including compensation or reparation (Busuioc, 2021). When AI fails, and there is a lack of accountability, it is challenging to address the root causes of AI failures, provide solutions for current failures, and prevent similar failures in the future. Hence, ensuring accountability becomes essential in today's world relying on GAI.

However, constructing accountability within the realm of GAI is challenging due to various barriers. The so-called problem of “*many hands*” is one of the main barriers to accountability in AI and computing in general (Cooper et al., 2022). The problem of many hands refers to the difficulty of attributing accountability for outcomes caused by multiple actors (Nissenbaum, 1996). The prevalent opacity of GAI strengthens this problem and leads to novel accountability challenges (Cooper et al., 2022; Nguyen et al., 2024). Specifically, various actors are involved in the development, operation, and use of GAI, leading to a divergent understanding of who can be held accountable. For instance, when ChatGPT provides incorrect information, it is difficult to isolate the actors and define who is accountable if employees rely on this information. The many hands problem thus results in employees underestimating their roles in and responsibility for GAI usage and may end up in situations where no one feels accountable for an error or damage. Lacking accountability can result in serious problems, such as inaccurate diagnoses in healthcare (Sendak et al., 2019), privacy violations, and data breaches (Mittelstadt et al., 2016).

As the focus of creating accountable GAI depends on a wide range of actors involved (Slota et al., 2023), it prompts us to consider who is accountable for consequences that result from GAI usage (Fui-Hoon Nah et al., 2023; Jovanovic & Campbell, 2022). Clarifying who is accountable for the potential harm of GAI usage in business settings is crucial for three key reasons. First, academia and practice consider AI accountability as essential when handling AI-based systems, for example, expressed by prominent frameworks like FATE (i.e., Fairness, Accountability, Transparency, Ethics (Memarian & Doleck, 2023)). Second, individuals' attributions of AI accountability have proven to shape their perceptions and behavior (e.g., fostering employees' trust and adoption (Shin, 2021)). Finally, if employees are not aware of their accountability or are not willing to be held accountable for their use of GAI-based systems, they may follow GAI blindly or blame GAI for potential risks or failure (i.e., the computer as a scapegoat (Moon & Nass, 1998)). Such GAI overreliance during human-AI collaboration (i.e., adopting GAI's information without rationalization or verification) can create automation bias and impede employees' critical thinking and problem-solving skills (Fui-Hoon Nah et al., 2023). In contrast, if employees hold themselves accountable, they will act more conservatively and carefully (Tetlock & Kim, 1987). Therefore, it calls for attention to if GAI fails, who do employees hold accountable and how accountability attributions impact their perceptions and behavior to foster reliable and mindful GAI usage.

There has been growing research interest in AI accountability and the benefits of perceived accountability among information systems (IS) researchers. Many studies have stated that accountability is important for AI systems (e.g., Khan and Vice (2022)) and call for accountability in AI systems (e.g., Nguyen et al. (2024)). Related AI accountability research predominantly focuses on accountability of AI developers, for example, how much accountability developers self-attribute to them and how much accountability they perceive others attribute to them (Schmidt et al., 2023). In contrast to related research on general accountability demonstrating that accountability impacts individuals' judgment processes and decision-making (e.g., motivate people to think more analytically (Lerner & Tetlock, 1999; Simonson & Nye, 1992)), studies examining how AI accountability shapes individuals' perceptions and behavior remain scarce. Notable exceptions with a focus on users' perceived accountability found evidence that perceived accountability leads to higher levels of trust in AI systems (Shin, 2021), can foster resource sharing with AI systems (León et al., 2021), and encourage system adoption (Choudhury & Asan, 2022).

While prior research provides insightful contributions, discusses accountable actors (e.g., Nguyen et al. (2024)), and highlights the consequences of AI accountability attributions (e.g., León et al. (2021)), there has been limited research regarding employees' perception of AI accountability when various accountable actors are involved. Indeed, prior research has focused on solely examining accountability of single actors, like users (e.g., Choudhury and Asan (2022)) or algorithms and their organizations (e.g., Shin and Park (2019)), or examined perceptions of general accountability (e.g., Shin (2021)). Given the crucial problem of many hands when using GAI and diverse involved actors like users, developers, GAI algorithms, and the organization, there is a need for further research that examines who do employees hold accountable when using GAI-based systems. Uncovering who employees hold accountable enables richer and more nuanced theorizing, for example, by comparing and integrating the impact of accountability attributions to different actors on individuals' behavior. Discerning who employees hold accountable paves the way to mitigate the many hands problem in AI contexts. We thus seek to answer the following research question (RQ): *Who do employees hold accountable when using a GAI-based system?*

We build on the theoretical framework of attribution theory and self-serving bias to answer this RQ. We develop a research model to examine how employees will attribute AI accountability to actors involved: Employees themselves or other actors, including the GAI-based system, developers, and organizations operating the system. Notably, we contrast two outcomes of system usage, success and failure, to compare employees' accountability attributions. We test our model using an experiment with 466 participants.

The results of our study revealed that in case of success, employees attribute accountability predominantly to themselves to preserve or enhance their self-concept. Surprisingly and against our propositions, employees also attribute the most accountability to themselves in case of failure. In addition, our results show that there is a rise in accountability attribution to other actors from failure to success outcomes. Performing detailed qualitative post-hoc analyses provides evidence that employees share accountability across actors to counteract the negative consequences of self-presentation. Our study, therefore, enriches our understanding of employees' perceptions of accountability from a multi-actor perspective and aids in the establishment of accountability frameworks within GAI systems. For practice, we encourage multiple stakeholders, including developers, organizations, and employees using GAI, to be aware of accountability of GAI systems and act more carefully to foster less biased and more reliable GAI-based systems.

Background

Generative AI at Work

This paper focuses on an important class of AI: GAI, referring to AI that can generate human-like text and creative content (e.g., text, music, and images) as well as consolidate data from different sources for analysis (Dasborough, 2023). Unlike traditional information technology, GAI can understand the patterns in each type of data, sample new items of that type, and produce original outputs (Goldstein et al., 2023). The integration of GAI into work has proven beneficial (e.g., fostering greater accessibility, efficiency, and cost-saving (Chen et al., 2023)). The leverage of GAI further drives business transformation and digital evolution, for example, in HR management (Goldstein et al., 2023). For employees, GAI enhances operational efficiency and problem-solving ability (Saengrith et al., 2022). Similarly, employees' work on processes, such as job description creation and intelligent search, can be augmented using GAI, transforming processes and empowering professionals to innovate and focus on higher-value work (Goldstein et al., 2023).

AI Accountability and the Problem of 'Many Hands'

Ensuring AI accountability becomes increasingly important to address the existing uncertainties and challenges employees face when using AI-based systems (e.g., ethical issues such as discriminating AI predictions). Both academia and practice consider accountability as key when handling AI-based systems, for example, expressed by prominent frameworks like FATE (Memarian & Doleck, 2023) or trustworthy AI (Thiebes et al., 2021). Accountability is most often referred to as "a relationship between an actor and a forum, in which the actor has an obligation to explain and to justify his or her conduct, the forum can pose questions and pass judgment, and the actor may face consequences" (Bovens, 2007, p. 450). Wieringa (2020) contextualized this prominent definition of accountability and defines AI accountability as creating an account for a socio-technical AI system that involves multiple actors (e.g., decision-makers, developers,

users) who have an obligation to explain and justify their use, design, or decisions concerning the AI system and subsequent consequences. These actors may be held accountable by several types of fora (e.g., external to the organization) for aspects of the system (e.g., the software code) or the entirety of the system (Nguyen et al., 2024).

However, achieving accountability in the context of AI is challenging because multiple actors (e.g., developers, users, organizations) are involved. This ambiguity to identify accountable actors is commonly referred to as the problem of “many hands” (Cooper et al., 2022) and has been identified as one of the leading barriers for achieving (AI) accountability (Nissenbaum, 1996). In its essence, the many hands problem refers to the difficulty of individuals to attribute moral responsibility for outcomes caused by multiple moral actors. In this study, we engage with this problem and examine employees' attributions of accountable actors involved in GAI-based systems. In general, there are divergent views on who should be accountable for the results of AI. For instance, Diakopoulos (2016) argues that the providers of algorithms should be held accountable for the consequences of their AI. On the contrary, designers and managers of AI systems are eventually held accountable as well for the outcomes and influence an AI system might have on stakeholders and society (Shin & Park, 2019). These mixed views of AI accountability attribution are even further complicated by related research engaging in inconsistent discussions on whether the AI algorithm can be an accountable entity (e.g., Kahn et al. (2012)). Legal studies mostly conclude that there is no way to hold algorithms accountable under current regulations because they still lack legal personhood (e.g., Deibel (2021)). In contrast, research on human-computer interaction has found compelling evidence that individuals will treat computers and technologies as social actors and therefore apply social rules and expectations to them (Jia et al., 2022; Moon, 2003). These divergent views on AI accountability attribution provide compelling evidence for the existence of the many hands problem. To further disentangle the problem of many hands and address calls for clarifying who is accountable for the positive and negative consequences when AI works correctly and incorrectly (Martin, 2018), we aim to empirically distinguish whom employees perceive as accountable actors when using GAI systems.

Related Research on AI Accountability Attribution

Related research on AI accountability has been increasing in recent years and can be summarized in three key streams (Table 1): (1) examining general accountability mechanisms, and how AI accountability shapes users' (2) perceptions and (3) behavior.

Related Streams	Description	Exemplary Studies
General accountability mechanisms	Effects of accountability on human thought processes and decision-making.	Tetlock and Kim (1987)
Users' perception of AI accountability	Users' perception of AI accountability affects their feelings toward and perceptions of AI systems.	Trust (Shin, 2021); satisfaction of AI systems (Shin & Park, 2019)
How AI accountability shapes behavior	Perceived AI accountability leads users to share resources with AI systems and use systems.	Choudhury and Asan (2022); León et al. (2021)
Table 1. Summary of Related Research Streams.		

Prior research demonstrates that individuals deeply engage with accountability, impacting their judgment and decision-making processes. Accountability influenced individuals by requiring them to specify who must answer to whom, for what, and under what ground rules they need to justify decisions (Tetlock & Kim, 1987). Attributing accountability to actors and resolving the problem of many hands thus is an essential part of accountability as it can increase thinking complexity, motivate individuals people to think more analytically and comprehensively, ultimately optimizing judgment process and improving final judgment results (Tetlock & Kim, 1987).

Related AI research similarly shows that users' perception of accountability affects their feelings and perceptions of AI systems. Accountability is ranked the top given the importance among ethical requirements of AI (e.g., fairness, transparency (Choung et al., 2023)). Accountability of AI systems has a strong influence on trust (Shin, 2021) and satisfaction of AI (Shin & Park, 2019). Furthermore, accountability changes users' behavior and leads users to react differently. For example, when users need to justify their actions, they tend to share more resources with AI systems and think more (León et al., 2021). Research has also explored the negative consequences of the problem of many hands. For instance, a lack of accountability impedes clinicians decide to use AI in clinical practices (Choudhury & Asan, 2022).

These research findings underline the importance of individuals' elaboration on who is accountable and the need to mitigate the problem of many hands.

Altogether, previous studies have provided compelling insights into how accountability shapes individuals' perceptions and behavior when using AI systems. However, extant has mostly focused on examining accountability attributions to single actors. For example, León et al. (2021) have examined how individuals behave when they are accountable, whereas Shin and Park (2019) examined accountability of algorithms and their organizations. This narrow focus on single accountable actors in prominent research is problematic because we may disregard beneficial and adverse outcomes of accountability attributions when theorizing the impact of AI accountability on individuals' behavior. For instance, if employees do not perceive themselves accountable, they may tend to blame the GAI for potential risks or failure, as witnessed in the past when users treated the computer as a scapegoat (Moon & Nass, 1998). In addition, positive consequences of feeling accountable cannot be accounted for when theorizing (e.g., acting more conservatively and carefully (Tetlock & Kim, 1987)). Further research is required that examines the problem of many hands in AI accountability and uncovers which actors individuals hold accountable for the outcome. To do so, we turn to the theoretical framework of attribution theory because identifying actors for accountability involves a process of attributing causes for actions.

Attribution Theory and Self-Serving Bias

Attribution theory explains how individuals make causal explanations about situations and illustrates how they will behave based on theoretical explanations (Serenko, 2007). Attribution theory highlights that when individuals' behavior is consistent with expectations, individuals will attribute causes of actions to internal characteristics (e.g., ability), but when behavior is inconsistent with their prior expectations, causes of actions will be attributed to external situational characteristics (e.g., bad luck) (Jones & Nisbett, 1987).

Attribution theory is an appropriate theoretical perspective for our research endeavors due to three key reasons. First, attribution theory provides theoretical explanations to examine "how people arrive at causal inferences, what sort of inferences they make, and what the consequences of these inferences are" (Folkes, 1988, p. 548). We leverage these explanations to examine how and why employees attribute accountability to different AI actors. Second, one of the most prominent theoretical mechanisms rooted in attribution theory is the self-serving bias that explains when and why individuals attribute an outcome to themselves or other actors depending on whether the outcome is a failure or success. Examining the presence of self-serving bias seems promising to distinguish between positive and negative consequences of GAI usage and examine who employees will hold accountable in each case. Finally, self-serving bias and related literature on attribution theory fits well with the basic premises of (AI) accountability: Self-serving bias manifests when individuals seek to maintain their self-concept (Snyder et al., 1976), which is similar to individuals' desire to attribute accountability during their decision-making processes to be evaluated favorably and avoid criticism by other actors (Lerner & Tetlock, 1999).

Self-serving bias leads employees to attribute dispositional and internal factors to their success, while external and uncontrollable factors are used to explain failure (Moon & Nass, 1998). Much research in a variety of fields supports the self-serving bias (e.g., Lerner and Tetlock (1999); Miller and Ross (1975)). Self-serving bias argues that a positive outcome usually leads individuals to claim credit for the success, while a negative outcome usually leads participants to blame either situational factors or other participants for the failure. In the case of a situation involving human-computer interaction, the self-serving bias leads to the prediction that users will tend to claim credit for successful outcomes and will tend to blame the computer for negative outcomes (Hong et al., 2020). Previous research also showed support for applying self-serving bias in AI contexts. For example, people attributed nearly the same level of blame to autonomous robots as to humans in case of failure (Furlough et al., 2021). Similarly, in the case of navigation systems helping human drivers, drivers attributed the driving performance more to the system when the system's autonomy was high (Brown et al., 2015). We next outline our research model explaining who do employees hold accountable in case of success and failure when using a GAI-based system.

Research Model

Our research model builds on the theoretical foundation of self-serving bias rooted in attribution theory. In its essence, the self-serving bias reflects employees' tendencies of self-enhancement and self-protection,

which entails employees taking more credit than they deserve for positive outcomes (i.e., self-enhancement), as well as failing to accept adequate responsibility for negative outcomes (i.e., self-protection) (Shepperd et al., 2008). We compare employees' accountability attribution to themselves and other actors. Considering that GAI typically follows a machine learning pipeline that starts with problem formulation and, in commercial settings, finally results in the deployment and continued monitoring of a trained model, we particularly compare employees' accountability attribution to *themselves* and multiple groups of *other actors* involved (Cooper et al., 2022): The GAI-based system itself, developers, and organizations operating the system. We thereby also consider the problem of many hands.

Self-Enhancement: Employees Taking Accountability for Success

Self-enhancement refers to individuals' motivation to maintain or enhance their sense of self-worth (Snyder et al., 1976). Prior research has shown that individuals strongly establish self-serving bias because of the benefits to self-worth (Shepperd et al., 2008). We similarly argue that if employees achieve expected outcomes and success when using GAI, they will hold themselves more accountable than other actors to enhance self-worth. The principles of self-serving bias and related research provide support: Employees attribute successful outcomes more strongly to internal factors, such as (1) unique skills and personal characteristics, (2) perceived control for the outcome, and (3) efforts (Cui et al., 2020; Miller & Ross, 1975).

Specifically, when using highly automated systems, individuals attribute success to their personal skills (Brown et al., 2015). Successfully engaging with GAI-based systems requires a certain degree of AI literacy and knowledge of GAI prompt engineering (i.e., creating accurate and effective prompts to generate optimal results (Dwivedi et al., 2023)). When usage of GAI-based systems leads to successful outcomes, employees are confirmed in their high qualifications and attribute credits to themselves. Then, employees will hold themselves accountable for the outcome to enhance self-worth and feel good about themselves (Meyer et al., 2022). For example, when employees use GAI like ChatGPT for their work and succeed with their tasks, they would attribute the success to personal extensive knowledge of ChatGPT and their wise use of it.

In addition, individuals' perceived control over the outcome is a strong condition for the manifestation of self-serving bias. The outcome of using a GAI-based system is dependent on employees because they are in control of prompting and engaging with the system to generate content. Related research has shown that if individuals feel like they are in control and have contributed to the outcome, they attribute the responsibility for the outcome more strongly to themselves to enhance their self-concept (Meyer et al., 2022; Puntoni et al., 2021). Employees' feelings of control over the GAI will then foster the emergence of self-serving bias because it contributes to employees' self-worth by emphasizing employees' contribution to success. Related AI research substantiates this proposition by highlighting that users take ownership of the outcomes of a human-AI collaboration depending on the degree to which they have realized their own agency (Sundar, 2020).

Finally, individual's perceived effort is also a strong antecedent of self-serving bias (Campbell & Sedikides, 1999; Jia et al., 2022). Employees may deeply engage with the GAI-based system to mitigate the risk of GAI hallucinations, among others (Dwivedi et al., 2023). If they then succeed because the GAI-based system provided reliable answers, employees will attribute the outcome to their usage efforts, further strengthening the self-serving bias and employees' accountability attribution to themselves.

In summary, employees feel good about themselves and achieve self-enhancement by taking personal credits for the success that is based on their attributes and (laborious) actions when using GAI-based systems. Therefore, we assume that employees using GAI successfully act in a self-serving manner and attribute accountability more strongly to themselves. We hypothesize the following:

H1: Employees will be more likely to attribute accountability to *themselves* when the outcome of using a GAI-based system is positive, compared to attributing accountability to *other actors* involved.

Self-Protection: Employees Denying Accountability for Failures

Literature on attribution theory and self-serving bias argues that individuals frequently attribute failure to external factors (Jones & Nisbett, 1987). "In an attempt to escape this uncomfortable state [of failure], individuals will make self-serving attributions" (Campbell & Sedikides, 1999, p. 26) and try to find someone else to blame (Na et al., 2023). Individuals then engage in self-protection to defend themselves from a

negative self-concept (Campbell & Sedikides, 1999). When encountering undesirable and potentially damaging feedback (e.g., failures), individuals will diminish the meaning and implications of the feedback to maintain a positive self-view. Individuals take self-protection seriously as an adaptive function for maintaining and protecting self-esteem and promoting a positive self-concept (Wang et al., 2017). We propose that negative outcomes using GAI-based systems will invoke employees' desire for self-protection, and they will hold other actors accountable to protect self-esteem and control damage to self-image.

Related research revealed that individuals showed self-protection in their interaction with computers as they become more autonomous and independent (Kling, 1996). Individuals have particularly developed a tendency to blame computers for their own mistakes and to hold computers morally responsible for errors (Kim & Hinds, 2006; Moon & Nass, 1998). Most research concludes that automated systems, such as AI systems, are held more accountable than other machines (e.g., a vending machine (Kahn et al., 2012)). Employees treat systems as social actors even when they know that systems do not possess feelings or intentions (Moon, 2003). Prior research proposes that if unfavorable outcomes appear, "a blame game" is likely to emerge (Jia et al., 2022). For example, employees tend to blame AI technologies for failing to meet individual user needs (Longoni et al., 2019) and are less likely to attribute undesirable AI-related outcomes to themselves (Meyer et al., 2022). In addition, the autonomous and black box features of GAI make the performance of GAI systems unpredictable and uncontrollable (Bélisle-Pipon et al., 2022), which encourages employees to claim "*it's the system's fault*" to evade accountability for harm. By blaming GAI-based systems, employees avoid feeling bad but instead feel that it is not their fault and thereby protect themselves from damage to self-image.

Besides attributing accountability for failures to the GAI-based system, employees can also blame other actors related to the system due to the many hands problem, such as AI developers. For example, when ChatGPT showed significant gender biases while being used by employees to write recommendation letters for job applications (Wan et al., 2023), employees may have concluded that developers have performed faulty model training and are therefore accountable for the outcome. Related research lends support for this proposition, arguing that such external attributions typically result if individuals perceive uncooperative other actors (Campbell & Sedikides, 1999). Similarly, employees can hold organizations accountable because they are responsible for operating and monitoring the system. (G)AI's changing behavior resulting from systems' intelligence, learning capabilities, and adaptability to contexts can lead to negative outcomes like performance reductions over time (Batarseh et al., 2021). Employees may suspect that organizations have failed to detect and mitigate such negative consequences, thereby protecting their self-concept and increasing organizations' accountability attributions. Since employees cannot control organizations' actions, they more strongly attribute responsibility for the failure to the external environment (Miller & Ross, 1975). Hence, by blaming other actors for the pitfalls and failure of GAI-based system usage, employees protect self-esteem. Based on these reasons, we suggest:

H2: Employees will be more likely to attribute accountability to *other actors* when the outcome of using GAI-based systems is negative, compared to attributing accountability to *themselves*.

Research Method

We conducted a scenario-based experiment to test our hypotheses. We introduced participants to a scenario in which they should attribute accountability to actors involved in the pipeline of using a GAI-based system. We examined participants' perception of accountability by manipulating the outcome of system usage, being either a success or failure. We chose a scenario-based experiment to ensure that the interactions were identical except for the respective outcome manipulations. We then could better control for confounding influences to achieve high internal validity (Meyer et al., 2022). We also asked participants to reason their accountability attributions, thereby combining qualitative and quantitative data collection and analyses.

Experiment Design

Our experiment aligns with the recommended design for testing self-serving biases: Participants perform a task, then randomly receive success or failure feedback, and finally make attributions for the task outcome (Campbell & Sedikides, 1999). We employed a between subject's experimental design consisting of six steps.

First, we welcomed our participants and provided a brief description of our study and information to ease onboarding to GAI, including ChatGPT and Sora as examples. We measured participants' AI knowledge to ensure they meet our data collection requirements. Next, we asked questions to control the potential influence of participants' general attitude toward AI and self-esteem. Afterward, we introduced our experiment scenario: We asked participants to imagine they were employed in a large coffee company. As a task, we told them that they should write a project proposal for expansion into new markets in Asia. Senior managers would decide if the proposal was good enough. If they approve, participants should expect to receive a budget to run a detailed feasibility study and additionally receive a salary bonus of \$10,000. We then told them that the company is operating 'WorkGPT' as a GAI-based assistant being able to provide information and data, give market-expanding recommendations, and generate an outline or template for the project proposal, among others. We asked them to assume that they use WorkGPT as an assistant to write the project proposal (i.e., performing the task). We purposely aligned WorkGPT's capabilities to similar GAI like ChatGPT because they are commonly used in real workplace settings today, to achieve high scenario realism, and harness participants' familiarity with ChatGPT. After reading the scenario, we provided three example screenshots of their assumed interaction with WorkGPT to simulate their system usage and increase immersion in the scenario. Each screenshot contains a prompt they inserted into WorkGPT and its response (Figure 1 (a)). To generate realistic responses, we used ChatGPT and adjusted its results to fit the scenario context of the coffee company. After engaging with the examples, we asked participants to select up to five prompts from a list. With this multiple-choice question, we aimed to (1) continue simulating the system usage (i.e., increasing participants' immersion) and (2) give participants the impression that they can make decisions on how to work on the task. Afterward, we measured participants' willingness to rely on WorkGPT's information and perceived risk of system usage. Then WorkGPT provides them two executive summaries (Figure 1 (b)) based on their previous choice of prompts. They needed to choose one to include in their final project proposal (i.e., again making choices related to the task, increasing their scenario immersion).

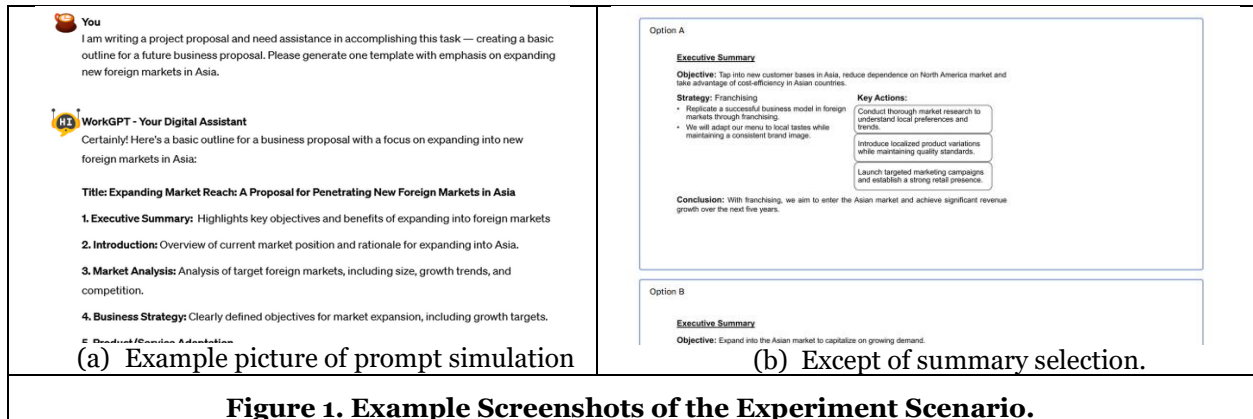


Figure 1. Example Screenshots of the Experiment Scenario.

Next, we performed our manipulation and either showed a feedback message telling them that the managers approved their proposal and they receive a bonus (i.e., success treatment), or disapproved their proposal and they have no bonus (i.e., failure treatment). Participants were assigned randomly because previous research found that when outcomes of experiments are random without considering the performances of participants, it is especially evident for participants to show self-serving bias tendency (Campbell & Sedikides, 1999). Hence, our approach aligns with typical self-serving bias research (e.g., Jia et al. (2022); Moon (2003)). To ensure that participants have read the outcome description and therefore our manipulation succeeded, we next performed an attention check. If they failed, they had to re-read the scenario outcome. We asked participants to make attributions of whom is held accountable for the outcome. Afterwards, they should provide a brief reasoning for their choices by answering open-ended questions (i.e., the complementary qualitative data collection). Then, we added a debriefing section. We measured how participants perceived the realism of the scenario, with a result showing a perception score of 5.68 (on a seven Likert point scale). We then told participants that the scenario and WorkGPT were fictional examples and outcomes were randomly assigned. In the end, we asked demographic questions (e.g., age, gender).

Measurement Instrument and Piloting

Following self-serving bias research on measuring outcome attributions (e.g., Moon (2003); Wang et al. (2017)), we asked participants to rate whether *they are accountable* (i.e., “Please think about whether you are or are not accountable for the outcome?”), and whether *other actors are accountable* for the scenario outcome (i.e., “Please think about who you perceive as accountable for the outcome?”). To consider the many hands problem, we included WorkGPT, its developers, and the operating organization as other actors (e.g., “The developers or designers of WorkGPT are accountable for the outcome.”). Items were measured using seven-point Likert scales (strongly agree – strongly disagree). To gather additional qualitative data on accountability attributions, we required participants to explain their attribution choices by answering two open-ended questions (i.e., “Please explain briefly your choice about why you are or are not accountable for the outcome” and “Please explain briefly your choice about who is accountable for the outcome”). The maximum number of characters for these open-ended questions is unlimited. We also measured users' general attitude towards AI (adapted from Schepman and Rodway (2020)), AI knowledge (adapted from Flynn and Goldsmith (1999)), self-esteem (adapted from Rosenberg (2015)), willingness to rely on WorkGPT's information (adapted from Soh et al. (2009)), and perceived risks when using WorkGPT (adapted from Pavlou and Gefen (2004)) to assess whether our treatment groups are structurally equivalent.

We pretested the experiment design, the two treatments, and measurement instrument with 10 faculty members and incorporated their feedback. We also conducted a pilot test with 59 participants using a panel provided by Amazon Mechanical Turk. Based on the feedback from the pilot test, we made changes to improve the comprehensibility and readability of the items and the treatment texts.

Participants

We used the online panel provider Cint to acquire participants for our experiment. We sampled participants based on two criteria. First, participants must reside in the United States to reduce cultural biases. We asked for a census-based sample representative of the US population concerning age, gender, and regions. Second, participants must be employed, at least sometimes use computers at their daily work, and have some knowledge and experience with AI, ensuring that participants are familiar with AI systems.

Cint administered the experiment to 500 participants, since we aimed to obtain about 250 participants for each group. We recorded the time spent on each page, ensuring participants paid sufficient attention. We removed 34 participants who did not pay attention to our survey (17 participants), rushed through the survey (6 participants), were unengaged or showed irregular answer patterns (11 participants), resulting in 466 valid responses. Of the final participants, 45.5% were women, 53.9% were men, and 0.6% preferred to self-describe their gender. They were, on average, 42 years old (minimum 18, maximum 64). They work as senior managers (23.6%), middle managers (26.8%), technicians (11.8%), professional staff (21.9%), administrative (9.7%), and other professions (6.2%). Participants work in various industries, with the largest proportion of employees working in information industries (13.5%) and educational services (10.1%). They held a high school diploma (18.7%), a bachelor's degree (48.3%), a master's degree (26.4%), a Ph.D. (2.4%), or another level of education (4.2%). They actively use AI in their work, 21.5% multiple times a day, 36.9% daily, 26.8% once a week, and 10.5% monthly. Most participants (69.5%) used AI text generators, with additional usage of AI image generators (37.8%), chatbots (63.3%), recommendation systems (54.3%), supportive AI tools like DeepL (32%), digital assistants like Siri (64.6%), and smart home technology (40.6%). Participants stated to generally have a positive attitude towards AI (mean = 5.28) and perceive moderate risks of AI (mean = 3.84).

Hypothesis Testing

First, we assessed the normality of the measurement items. All the items had the absolute kurtosis value below the acceptable threshold of 10.0 for kurtosis and skewness below the threshold of 3 (Kline, 2023). The highest kurtosis value was 2.95 and the highest skewness value was 1.45 for self-esteem₃. We conclude that the distributions of our measurement items do not deviate substantially from normality. Second, we assessed whether our treatment groups are structurally equivalent based on control variables and demographic data. We first checked reliability, convergent validity, and discriminant validity of our control variables via a confirmatory factor analysis in SPSS AMOS. Then, we conducted a one-way MANOVA that

found no statistically significant differences between treatment groups, $F(11, 454) = 1.093$, $p = 0.365$, Wilk's $\Lambda = 0.974$. We therefore conclude that treatment groups are structurally equivalent and differences in accountability attributions should (largely) stem from our outcome manipulations.

To test our hypotheses, we conducted paired samples t-tests. Considering the multi-actor and many hands nature of GAI systems, we particularly compared participants' accountability attributions to themselves versus attributions to AI developers, the organization, and WorkGPT. We also combined the attributions for the latter actors as "other actors" being accountable.

For H1, in the success treatment, a paired samples t-test revealed a statistically significant difference between employees' attribution of accountability to themselves and other actors ($p < 0.001$; mean difference (M-Diff) = 1.49). Specifically, employees were significantly more likely to attribute accountability to themselves compared to AI developers ($p < 0.001$; M-Diff = 1.71), the organization ($p < 0.001$; M-Diff = 1.15), and WorkGPT ($p < 0.001$; M-Diff = 1.60). It showed that the patterns of attribution of accountability remain consistent when we consider the actors separately or collectively. Thus, H1 is supported: Employees will be more likely to attribute accountability to themselves when the outcome of using a GAI system is positive, compared to attributing accountability to other actors.

Conversely, in the failure treatment, even though a paired samples t-test revealed a statistically significant difference between employees' attribution of accountability to self and attribution to other actors, contrary to our hypothesis, employees were significantly more likely to attribute accountability to themselves compared to other actors when the outcome was a failure ($p < 0.001$, M-Diff: 2.13). Similarly, multi-actor sensitivity analysis revealed the same pattern: Employees were significantly more likely to attribute accountability to themselves compared to AI developers ($p < 0.001$; M-Diff = 2.22), the organization ($p < 0.001$; M-Diff = 1.90), and WorkGPT ($p < 0.001$; M-Diff = 2.28). H2 is not supported.

We compared our treatment groups in detail. A one-way MANOVA showed a statistically significant difference between the treatment groups on the combined accountability attributions ($F(5, 460) = 7.292$, $p < .001$, Wilk's $\Lambda = .927$). Post-hoc univariate ANOVAs show no statistically significant difference between the treatment groups for attribution to oneself ($F(1, 464) = 2.888$, $p = 0.090$) whereas significant differences for attribution to developers ($F(1, 464) = 14.257$, $p < 0.001$), organization ($F(1, 464) = 27.417$, $p < 0.001$), and WorkGPT ($F(1, 464) = 22.840$, $p < 0.001$). We observed an increase in the level of attribution in the success scenario compared to failure (**Figure 2**). In the failure scenario, the mean score for attribution to oneself is 5.77 (SD = 1.35) and 3.63 (SD = 1.87) to other actors. In contrast, in the success scenario, the mean score for attribution to oneself increased to 5.96 (SD = 1.09) and to others increased to 4.47 (SD = 1.67). Reflecting these surprising findings, we turned to our qualitative data to better understand why participants attribute accountability differently to self and others, in case of failure and success.

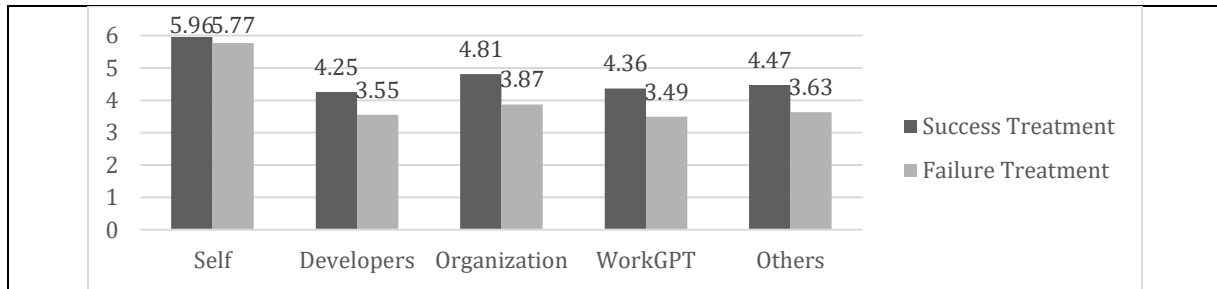


Figure 2. Mean Comparisons of Accountability Attributions.

Post-hoc Qualitative Analyses

We analyzed participants' responses to our open-ended questions to reason their accountability attributions. We thereby complement our quantitative analyses with qualitative analyses to gain deeper insights into the data and derive explanations for our surprising findings. In total, participants provided 7144 words in total and 15.23 words in average as reasoning for attribution to themselves, and 7083 words in total and 14.82 words in average for attributing others. We applied coding techniques commonly used in IS research: open and axial coding (Myers, 2013). We complemented these coding techniques by

performing constant comparisons of our treatments and literature findings. Two authors coded the qualitative data independently and discussed and aggregated their results. The author team then discussed the qualitative findings and iteratively improved the coding results.

First, we performed open coding (Myers, 2013) to identify the reasons for holding someone accountable. We assigned labels as codes to each of the responses. We created 76 codes related to 1387 textual segments through this way of coding. Each code briefly described the participants' reasons and the treatment label (i.e., SUCCESS, FAILURE) and the attributed actor (i.e., oneself, developers, organizations, WorkGPT, and shared accountability). For example, we coded the text segment *"I am accountable for my outcome because [...] I did not put the right amount of effort into it"* with the code "FAILURE_Self_bad performance". Second, we engaged in axial coding (Myers, 2013). We analyzed each individual code in detail and compared its relationships with (related) codes. We aggregated 26 codes because they contained similar reasons. We also assigned a label to each code determining whether the reason argues for or against the accountability of an actor (e.g., "FAILURE_Self_accountable_bad performance"), to better distinguish why employees hold or do not hold certain actors accountable. In total, we identified 29 reasons in the success and 21 in the failure treatments (summarized in Table 2). Finally, we engaged in constant comparison of our findings. We compared reasons across our two treatments, mapped our qualitative findings to participants' quantitative accountability attributions, and turned back to the literature on self-serving bias and attribution theory to explain our findings.

Our qualitative findings further substantiate H1 because participants strongly argued that they are accountable for success. *"I am ultimately deciding what information I gather from AI, what information is used in my presentation from AI, and how to compile all the information together cohesively to best fit the specific project proposal."* In line with self-serving bias (Shepperd et al., 2008), participants attribute the successful outcomes to internal factors, particularly their efforts, abilities, prompt engineering skills, and oversight actions (e.g., validating AI-generated information and cross-checking with additional data sources). Particularly, oversight is important in GAI to achieve positive outcomes. *"I would be double checking and editing all the work. I would take the time to change any parts I felt were substandard to my expectations, etc."* Notably they also claim accountability since they are in charge of decision making, including the decision to use (or not use) the GAI-based system and finally submit the proposal. Making the decisions and emphasizing that oneself is in charge of the task strengthens participants' feelings of control, which has been a key source of self-serving bias (Meyer et al., 2022; Puntoni et al., 2021) and important to make attributions to oneself according to attribution theory (Jones & Nisbett, 1987).

Actor	Key reasons for accountability attribution (#text segments coded)	
	Success treatment	Failure treatment
Employee	- A: confidence in own work (111) - A: making decisions (91) & submitted (9) - A: performed oversight (48) - A: choice for selecting GAI (44) - A: employment responsibility (36) - A: prompting (35) - A: efforts (31) & ability (23)	- A: making decisions (118) & submitted (68) - A: bad performance of self (54) - A: performed oversight (45) - A: employment responsibility (45) - A: choice for selecting GAI (43) - A: WorkGPT is just a tool (33) - A: involvement (25) - A: prompting (21) - A: overreliance on GAI (21)
WorkGPT	- A: provided support (59) - A: a primary actor (24) - A: involved in proposal creation (7) - NA: WorkGPT is just a tool (35) - NA: not a human (5)	- A: choice for selecting GAI (43) - A: WorkGPT is just a tool (33) - A: involvement (25) - A: prompting (21) - A: overreliance on GAI (21)
Developer	- A: good programming (4)	- A: bad performance (7)
Organization	- A: decision to adopt GAI (19) - A: operating GAI (4)	- A: reliable adoption of GAI (5)
Shared	- A: collective contributions (25)	- A: collective contributions (14)
Note: A = accountability attributed, NA = no accountability attributed		
Table 2. Summary of Key Reasons for Accountability Attribution		

Turning to H2, our qualitative findings from the failure treatment provide deeper insights into the surprising, quantitative results. Participants generally hold themselves accountable because they believe they are in control. Particularly, participants stressed more often that they decided to submit the report (i.e., 68 times stated as reason in case of failure vs. 9 times in case of success). Similarly, WorkGPT was

regarded as a tool; although supportive, it is not accountable because it only supports gathering and aggregating data. Related research has similarly argued that if technology is perceived by individuals as only providing simple information or fulfilling utilitarian tasks, they perceive themselves as more (or solely) responsible for the outcome (Jia et al., 2022). Instead of protecting their self-concept, participants actually argued that their performance was not sufficient (e.g., *"I am accountable for not using much of my own mind to create this project."*) or lacked oversight over AI-generated information (e.g., *"WorkGPT was just a tool I ultimately had to decide which information was good enough"*). Provided reasons in case of failure thus mirrors participants' reasons for success attributions. Some even argued for an over-reliance on AI resulting in mindless usage of generated information: *"[...] it was my fault for relying on the program and not using my own brain"*. In addition, participants attribute more accountability to themselves because of employment responsibility (e.g., *"am accountable because I was given the job not WorkGPT"*). In conclusion, qualitative analysis provides further support that participants attributed AI accountability in case of failure to themselves because they perceive to have the control and are much more open to self-critical thinking than expected.

While participants mostly attributed accountability to themselves, we also witnessed that participants generally shared accountability with other actors in both treatments. Participants attributed varying accountability to the other actors since they are involved (to a certain degree) in and made contributions to the system development and operation. A participant nicely summarizes the idea of shared accountability: *"[...] accountability for AI outcomes is distributed among users, developers, organizations, regulators, and society. Each stakeholder has a role in ensuring that AI technologies are developed and deployed responsibly, ethically, and in the best interest of all stakeholders involved"*. These shared accountability attributions substantiate the problem of many hands in (G)AI contexts and also align with the core idea of AI accountability requiring that each actor needs to justify their use, design, or decisions (Wieringa, 2020).

Perceptions of shared accountability particularly manifested in the failure treatment: Participants stressed that WorkGPT, developers, and the organization are accountable as well because their inadequate performance contributes to diminished AI system qualities. *"The AI Work GPT put up financial information that got me rejected"* and *"Especially if it is a paid AI, then you would assume it has some accuracy etc. The organization is accountable somewhat as they chose the product AI to use"*. Denying accountability through sharing is similar to blaming computers for failures, a common belief found in self-serving bias literature (e.g., Moon and Nass (1998); Serenko (2007)). Although participants' accountability attributions are lower for the other actors, these comments provide weak support for employees' self-protection tendencies inherent in the self-serving bias. Nevertheless, since employees reported to be in charge of oversighting AI-generated results, they attribute a larger share of accountability to themselves.

In case of success, participants also shared accountability with the other actors. They particularly reported that the GAI tool was supportive and took an active role in the development of the proposal. Interestingly, participants in the success treatment attributed significant higher accountability to other actors compared to the failure treatment, as evident in the quantitative results (**Figure 2**). Related literature provides reasoning by highlighting that self-presentation (i.e., explaining own success to others) moderates self-serving bias. Although attributing success to oneself can enhance employees' self-concept (Forsyth & Schlenker, 1977), excessive self-attribution can lead to the perception of being a blowhard, resulting in disapproval from others (Bradley, 1978; Miller & Schlenker, 1985). Consequently, individuals may present themselves rather modestly to prevent potential negative consequences. Our open-ended questions may have led to a situation of self-presentation (i.e., determining attribution and explaining their reasons to the authors). We believe that the difference in accountability attribution then results from employees' calibration of their success attributions among all actors that they perceive to be involved in the GAI-based system. This calibration manifests through sharing accountability across all involved actors, thereby providing credit to their individual contribution to the successful outcome. *"I used WorkGPT to help assist me so I feel like I only deserve about half the credit on the project instead of full credit"*. *"The owners of the company made the decision to purchase and use WorkGPT and ultimately are accountable for it [...]"*. At the same time, they frequently stressed that they were in control to highlight that their share of accountability is larger compared to the other actors, as result of the self-serving bias. *"[WorkGPT] Did help me get to that place but I still submitted the final proposal"*. Such a calibration can also take more extreme forms: We observed that some participants in the success treatment even discredit the other actors to enhance their self-concept. For example, they stated that the GAI-based system is *"just a tool"* and *"only as good as the wielder of it."* *"I entered the parameters into the bot, so it didn't have done anything I didn't"*

input". Although acknowledging the accountability of other actors, those participants may wanted to make sure that success is still largely attributed to themselves and based on their abilities. "*It is still my proposal*" [emphasis added]. Israeli (2020) similar proposes that employees seek to be perceived as accountable for their work and to enjoy their colleagues' appreciation.

Discussion

Principal Findings

In this study we examined employees' accountability attributions based on the self-serving bias which is rooted in attribution theory. We conducted an experiment to test who do employees hold accountable in case of success and failure. The quantitative results support H1, which indicates that when employees succeed, they will hold themselves more accountable for the positive outcome than developers of GAI, organizations, and the GAI itself to maintain or enhance their sense of self-worth. Qualitative findings substantiate these results and highlight that participants attribute the successful outcomes based on their efforts and abilities such as prompt engineering skills and oversight actions.

We found no support for H2. Contrary to our expectations and the rationales of self-serving bias that employees will be more likely to attribute accountability to other actors when the outcome is negative, our analysis showed that they attributed more accountability to themselves than to other actors separately and collectively when the outcome of using GAI was a failure. Qualitative analyses explain that employees attribute accountability in case of failure to themselves because they remain in control, make the (final) decisions on how to use AI-generated information. Open responses also reflect that employees engaged in rather self-critical thinking instead of self-protection. Besides, these attribution tendencies may result from the study being placed in a work context because employees strongly feel accountable due to their employment and job responsibilities. "[...] *As a professional, I take ownership of my actions and decisions, especially when they impact the success or failure of a project. By acknowledging my accountability, I am motivated to work diligently, and take responsibility for the consequences. This mindset fosters trust and reliability within the team and ensures that I remain focused on achieving the desired outcomes.*" This comment also provides evidence that attributions of AI accountability to oneself can lead to more mindful engagement with GAI-based systems, in line with prior accountability research (e.g., Tetlock and Kim (1987)).

Interestingly, we identified that employees share accountability attribution among all actors involved in the GAI-based system development, operation, and usage. Our findings of shared accountability highlight the many hands problems in the realm of GAI and emphasized the need for all actors to justify their actions. We particularly found that the levels of attribution in the success outcome to all actors are higher than the levels of attribution in failure scenario. Self-serving bias explains that such a shared, increased accountability attribution results from employees protecting themselves against negative consequences of self-presentation (e.g., showing-off).

Contributions to Research and Practice

This study makes several contributions to research. First, there is a general lack of research on AI accountability's impact on individuals within the realm of (G)AI. With unique characteristics like being a black box, GAI has an incomprehensible nature of how it makes decisions or process information, which raised significant concerns about accountability. While various researchers call for AI accountability (e.g., Fui-Hoon Nah et al. (2023); Khan et al. (2023); Memarian and Doleck (2023)), particularly given recent AI scandals, only few studies have examined how AI accountability shapes individuals' perceptions and behavior (e.g., León et al. (2021); Shin (2021)). In this study, we examine user's accountability perceptions of GAI systems and reveal that in the context of GAI, employees tend to hold themselves more accountable in failure scenarios, which contrasts with the findings of non-GAI systems (Brown et al., 2015; Jia et al., 2022). Hence, our study provides insights that GAI's opaque nature leads to a more complex accountability dynamic and is valuable for establishing AI accountability frameworks.

Second, prior research has mostly focused on examining accountability attributions to single actors (León et al., 2021; Shin & Park, 2019). However, in (G)AI systems, there are multiple actors involved and focusing solely on single actors underestimates the complexity of the many hands problem. Our study addressed this

limitation by considering multiple actors involved in GAI systems. We identify that employee attribute accountability to self and other actors, with a higher-level attribution to themselves but also medium attribution to other actors. We thus highlight that employees' accountability attributions are more diverse and extensive than assumed by prior research. This study's findings thereby enrich research's understanding about employees attributing and sharing accountability to various actors involved in the development and operation of (G)AI-based systems, which is valuable for comprehending the complexities of the many hands problem inherent in (G)AI. Our findings inform upcoming accountability research: We advise researchers to clearly identify potential accountable actors in AI use context and apply a corresponding multi-actor accountability measurement to comprehensively assess individuals' accountability attributions. Then, researchers could engage in more deeper theorizing, for example, by comparing the impact of accountability attributions of different actors on individuals' behavior.

Finally, prior research only examines AI accountability when using AI systems resulted in success (e.g., León et al. (2021); Shin (2021)) and failure separately (e.g., Furlough et al. (2021); Na et al. (2023)). In this study, we considered both outcomes simultaneously, not only verifying previous research but also revealing surprising findings: Employees attribute accountability at higher levels to themselves in case of success than failure. By considering both conditions, success and failure simultaneously, we provide researchers with novel knowledge about different attribution behavior under different scenarios.

Our study also informs practice by offering insights for employees using GAI at work, developers of GAI, and organizations operating GAI. First, we encourage employees to act more carefully when using GAI at work instead of following GAI blindly. Accepting accountability can foster mindful GAI usage and improve personal critical thinking, and problem-solving abilities, as evident in prior research (León et al., 2021). Second, we suggest that developers should be cautious when designing GAI-based systems and align with expectations and needs of employees. Our study emphasizes that users of GAI will hold developers accountable for their involvement in GAI systems to some extent. Developers should thus accept this shared accountability to drive their motivation in developing reliable and suitable GAI-based systems. If developers better fulfill employees' needs, GAI systems can work in a manner that resonates with employees' accountability perceptions, promoting trust and adoption of GAI at work (e.g., Choudhury and Asan (2022); León et al. (2021); Shin (2021)). Likewise, our study found that employees also perceive organizations operating GAI systems accountable for both failure and success outcomes. Hence, we want to arise organizations' attention that they must carefully consider their accountability before launching a new GAI-based systems and elaborate on possible outcomes when employees use these systems.

Limitations and Future Work

This study has limitations that that open valuable avenues for future research. First, the generalizability of our findings may be limited given that our experiment involved solely US employees and focused on a GAI system at work similar to ChatGPT. Future research should validate our findings across various GAI contexts and systems with varying (interaction) capabilities, such as GitHub Copilot, and expand samples to include participants from diverse cultural backgrounds. Second, our experiment did not consider potential moderators or other factors that can influence attributions of accountability as highlighted by prior research on self-serving bias (e.g. self-esteem (Shepperd et al., 2008) or outcome expectations (Campbell & Sedikides, 1999)). Considering them could provide deeper insights into the process of how self-serving bias influence accountability attributions. Also, there are other possible factors that could influence individuals' locus of control and their attribution behavior (e.g., personal knowledge of GAI and the underlying task, personal experience). Future research could examine these moderating impacts and factors. Third, the surprising findings regarding H2 can be further explored by other methodological approaches (e.g., interviews or non-intrusive accountability measurements) to gain deeper insights into sharing accountability attributions. Fourth, our findings are based on a lab experiment, which increased the controllability of our study but may not fully represent a realistic work situation and tasks (e.g., stress and pressures at work). Attributions may differ depending on the task complexity and how the task is embedded in the work settings (e.g., team dynamics). Our feedback mechanisms work solely as text explanations, which can be improved by immersed interaction. The results should be verified with field experiments to observe actual behavior and real-life consequences. Lastly, our study is a single-case experiment, which may overlook the dynamics of accountability attributions. Future longitudinal studies are needed to see how individuals may change accountability attributions over time, such as after facing multiple successes and failures.

Conclusion

In this study, we examine who do employees hold accountable when using GAI-based systems. Building on the self-serving bias rooted in attribution theory, we developed a research model and tested our proposed relationships in an online experiment with 466 participants. Our findings demonstrate that in case of success and failure employees will hold themselves more accountable than developers of GAI, organizations, and the GAI-based system itself. This study's findings enrich research's understanding of employees' accountability attributions and highlight that a multi-actor perspective is needed when examining AI accountability for GAI-based systems.

Acknowledgements

This study is funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – project no. 471168026.

References

- Batarseh, F. A., Freeman, L., & Huang, C.-H. (2021). A survey on artificial intelligence assurance. *Journal of Big Data*, 8(1), 60.
- Bélisle-Pipon, J.-C., Monteferrante, E., Roy, M.-C., & Couture, V. (2022). Artificial intelligence ethics has a black box problem. *AI & SOCIETY*, 38, 1507-1522.
- Bovens, M. (2007). Analysing and assessing accountability: A conceptual framework 1. *European law journal*, 13(4), 447-468.
- Bradley, G. W. (1978). Self-serving biases in the attribution process: A reexamination of the fact or fiction question. *Journal of Personality and Social Psychology*, 36(1), 56-71.
- Brown, M., Houghton, R., Sharples, S., & Morley, J. (2015). The attribution of success when using navigation aids. *Ergonomics*, 58(3), 426-433.
- Brynjolfsson, E., Li, D., & Raymond, L. R. (2023). Generative AI at work. National Bureau of Economic Research.
- Busuioc, M. (2021). Accountable artificial intelligence. *Public administration review*, 81(5), 825-836.
- Campbell, W. K., & Sedikides, C. (1999). Self-threat magnifies the self-serving bias: A meta-analytic integration. *Review of general Psychology*, 3(1), 23-43.
- Chen, B., Wu, Z., & Zhao, R. (2023). From fiction to fact: the growing role of generative AI in business and finance. *Journal of Chinese Economic and Business Studies*, 21(4), 471-496.
- Choudhury, A., & Asan, O. (2022). Impact of accountability, training, and human factors on the use of artificial intelligence in healthcare. *Human Factors in Healthcare*, 2, 100021.
- Choung, H., David, P., & Ross, A. (2023). Trust and ethics in AI. *AI & SOCIETY*, 38(2), 733-745.
- Cooper, A. F., Moss, E., Laufer, B., & Nissenbaum, H. (2022). Accountability in an Algorithmic Society: Relationality, Responsibility, and Robustness in Machine Learning. *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, 864-876.
- Cui, G., Wang, Y., Wang, X., Zheng, L., Li, L., Li, P., Zhang, L., Guo, Y., Chen, Y., & Sun, Z. (2020). Static and dynamic functional connectivity of the prefrontal cortex during resting-state predicts self-serving bias in depression. *Behavioural Brain Research*, 379, 112335.
- Dasborough, M. T. (2023). Awe-inspiring advancements in AI: The impact of ChatGPT on the field of Organizational Behavior. *Journal of Organizational Behavior*, 44(2), 177-179.
- Deibel, T. (2021). Back to (for) the Future: AI and The Dualism of Persona and Res in Roman Law. *European Journal of Law and Technology*, 12(2), 1-27.
- Diakopoulos, N. (2016). Accountability in algorithmic decision making. *Communications of the ACM*, 59(2), 56-62.
- Dwivedi, Y. K., Kshetri, N., Hughes, L., Slade, E. L., Jeyaraj, A., Kar, A. K., Baabdullah, A. M., Koohang, A., Raghavan, V., & Ahuja, M. (2023). "So what if ChatGPT wrote it?" Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy. *International Journal of Information Management*, 71, 102642.
- Flynn, L. R., & Goldsmith, R. E. (1999). A short, reliable measure of subjective knowledge. *Journal of business research*, 46(1), 57-66.

- Folkes, V. S. (1988). Recent attribution research in consumer behavior: A review and new directions. *Journal of Consumer Research*, 14(4), 548-565.
- Forsyth, D. R., & Schlenker, B. R. (1977). Attributing the causes of group performance: Effects of performance quality, task importance, and future testing 1. *Journal of Personality*, 45(2), 220-236.
- Fui-Hoon Nah, F., Zheng, R., Cai, J., Siau, K., & Chen, L. (2023). Generative AI and ChatGPT. *Journal of Information Technology Case and Application Research*, 25(3), 277-304.
- Furlough, C., Stokes, T., & Gillan, D. J. (2021). Attributing Blame to Robots: I. The Influence of Robot Autonomy. *The Journal of the Human Factors and Ergonomics Society*, 63(4), 592-602.
- Goldstein, J. A., Sastry, G., Musser, M., DiResta, R., Gentzel, M., & Sedova, K. (2023). Generative language models and automated influence operations. arXiv Preprint arXiv:2301.04246.
- Hong, J.-W., Wang, Y., & Lanz, P. (2020). Why is artificial intelligence blamed more? Analysis of faulting artificial intelligence for self-driving car accidents in experimental settings. *International Journal of Human-Computer Interaction*, 36(18), 1768-1774.
- Israeli, T. (2020). Losing information is like losing an arm: employee reactions to data loss. *Behaviour & Information Technology*, 39(12), 1297-1307.
- Jia, H., Wu, M., & Sundar, S. S. (2022). Do we blame it on the machine? Task outcome and agency attribution in human-technology collaboration. *Proceedings of the 55th Hawaii International Conference on System Sciences (HICSS)*, 388-397.
- Jones, E. E., & Nisbett, R. E. (1987). The actor and the observer: Divergent perceptions of the causes of behavior. In E. E. Jones, D. E. Kanouse, H. H. Kelley, R. E. Nisbett, S. Valins, & B. Weiner (Eds.), *Attribution: Perceiving the causes of behavior* (pp. 79-94). Lawrence Erlbaum Associates, Inc.
- Jovanovic, M., & Campbell, M. (2022). Generative Artificial Intelligence: Trends and Prospects. *Computer*, 55(10), 107-112.
- Kahn Jr, P. H., Kanda, T., Ishiguro, H., Gill, B. T., Ruckert, J. H., Shen, S., ... & Severson, R. L. (2012). Do people hold a humanoid robot morally accountable for the harm it causes? *Proceedings of the seventh annual ACM/IEEE international conference on Human-Robot Interaction*, 33-40.
- Khan, A. A., Akbar, M. A., Fahmideh, M., Liang, P., Waseem, M., Ahmad, A., Niazi, M., & Abrahamsson, P. (2023). AI Ethics: An Empirical Study on the Views of Practitioners and Lawmakers. *IEEE Transactions on Computational Social Systems*, 10(6) 1-14.
- Khan, M. M., & Vice, J. (2022). Toward accountable and explainable artificial intelligence part one: theory and examples. *IEEE Access*, 10, 99686-99701.
- Kim, T., & Hinds, P. (2006). Who should I blame? Effects of autonomy and transparency on attributions in human-robot interaction in ROMAN 2006 - The 15th IEEE International Symposium on Robot and Human Interactive Communication, 80-85.
- Kline, R. B. (2023). Principles and practice of structural equation modeling. Guilford publications.
- Kling, R. (1996). Computerization and controversy: Value conflicts and social choices. Elsevier.
- León, G. A., Chiou, E. K., & Wilkins, A. (2021). Accountability Increases Resource Sharing. *International Journal of Human-Computer Interaction*, 37(5), 434-444.
- Lerner, J. S., & Tetlock, P. E. (1999). Accounting for the effects of accountability. *Psychological bulletin*, 125(2), 255-275.
- Longoni, C., Bonezzi, A., & Morewedge, C. K. (2019). Resistance to medical artificial intelligence. *Journal of Consumer Research*, 46(4), 629-650.
- Martin, K. E. (2018). Ethical Implications and Accountability of Algorithms. *Journal of business ethics*, 160(4), 833-850.
- Memarian, B., & Doleck, T. (2023). Fairness, Accountability, Transparency, and Ethics (FATE) in Artificial Intelligence (AI), and higher education: A systematic review. *Computers and Education: Artificial Intelligence*, 5(4), 100152.
- Meyer, N., Schwede, M., Hammerschmidt, M., & Weiger, W. H. (2022). Users taking the blame? How service failure, recovery, and robot design affect user attributions and retention. *Electronic Markets*, 32(4), 2491-2505.
- Miller, D. T., & Ross, M. (1975). Self-serving biases in the attribution of causality: Fact or fiction? *Psychological bulletin*, 82(2), 213-225.
- Miller, R. S., & Schlenker, B. R. (1985). Egotism in group members: Public and private attributions of responsibility for group performance. *Social Psychology Quarterly*, 48, 85-89.
- Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. *Big Data & Society*, 3(2), 1-21.
- Moon, Y. (2003). Don't blame the computer. *Journal of Consumer Psychology*, 13(1-2), 125-137.

- Moon, Y., & Nass, C. (1998). Are computers scapegoats? Attributions of responsibility in human–computer interaction. *International Journal of Human-Computer Studies*, 49(1), 79-94.
- Myers, M. D. (2013). *Qualitative Research in Business and Management*. SAGE Publications.
- Na, G., Choi, J., & Kang, H. (2023). It's not my fault, but I'm to blame. *International Journal of Human–Computer Interaction*, 1, 1-15.
- Nguyen, L. H., Lins, S., Renner, M., & Sunyaev, A. (2024). Unraveling the Nuances of AI Accountability: A Synthesis of Dimensions Across Disciplines. *Proceedings of the 32nd European Conference on Information Systems (ECIS)*.
- Nguyen, N., & Nadi, S. (2022). An empirical evaluation of GitHub copilot's code suggestions. *Proceedings of the 19th International Conference on Mining Software Repositories*, 1-5.
- Nissenbaum, H. (1996). Accountability in a computerized society. *Science and engineering ethics*, 2, 25-42.
- Pavlou, P. A., & Gefen, D. (2004). Building effective online marketplaces with institution-based trust. *Information Systems Research*, 15(1), 37-59.
- Peres, R., Schreier, M., Schweidel, D., & Sorescu, A. (2023). On ChatGPT and beyond. *International Journal of Research in Marketing*, 40(2), 269-275.
- Puntoni, S., Reczek, R. W., Giesler, M., & Botti, S. (2021). Consumers and artificial intelligence: An experiential perspective. *Journal of Marketing*, 85(1), 131-151.
- Rosenberg, M. (2015). *Society and the adolescent self-image*. Princeton university press.
- Saengrith, W., Viriyavejakul, C., & Pimdee, P. (2022). Problem-Based Blended Training via Chatbot to Enhance the Problem-Solving Skill in the Workplace. *Emerging Science Journal*, 6, 1-12.
- Schepman, A., & Rodway, P. (2020). Initial validation of the general attitudes towards Artificial Intelligence Scale. *Computers in human behavior reports*, 1, 100014.
- Schmidt, J.-H., Bartsch, S. C., Adam, M., & Benlian, A. (2023). Accountability incongruence and its effects on AI Developers' Job Satisfaction. *Proceedings of the 31st European Conference on Information Systems (ECIS)*.
- Sendak, M., Gao, M., Nichols, M., Lin, A., & Balu, S. (2019). Machine learning in health care: a critical appraisal of challenges and opportunities. *EGEMs*, 7(1), 1.
- Serenko, A. (2007). Are interface agents scapegoats? *Interacting with Computers*, 19(2), 293-303.
- Shepperd, J., Malone, W., & Sweeny, K. (2008). Exploring causes of the self-serving bias. *Social and Personality Psychology Compass*, 2(2), 895-908.
- Shin, D. (2021). The effects of explainability and causability on perception, trust, and acceptance: Implications for explainable AI. *International Journal of Human-Computer Studies*, 146, 102551.
- Shin, D., & Park, Y. J. (2019). Role of fairness, accountability, and transparency in algorithmic affordance. *Computers in Human Behavior*, 98, 277-284.
- Simonson, I., & Nye, P. (1992). The effect of accountability on susceptibility to decision errors. *Organizational behavior and human decision processes*, 51(3), 416-446.
- Slota, S. C., Fleischmann, K. R., Greenberg, S., Verma, N., Cummings, B., Li, L., & Shenefiel, C. (2023). Many hands make many fingers to point. *AI & SOCIETY*, 38(4), 1287-1299.
- Snyder, M. L., Stephan, W. G., & Rosenfield, D. (1976). Egotism and attribution. *Journal of Personality and Social Psychology*, 33(4), 435-441.
- Soh, H., Reid, L. N., & King, K. W. (2009). Measuring trust in advertising. *Journal of advertising*, 38(2), 83-104.
- Sundar, S. S. (2020). Rise of machine agency: A framework for studying the psychology of human–AI interaction (HAI). *Journal of Computer-Mediated Communication*, 25(1), 74-88.
- Tetlock, P. E., & Kim, J. I. (1987). Accountability and judgment processes in a personality prediction task. *Journal of Personality and Social Psychology*, 52(4), 700-709.
- Thiebes, S., Lins, S., & Sunyaev, A. (2021). Trustworthy artificial intelligence. *Electronic Markets*, 31(2), 447-464.
- Wan, Y., Pu, G., Sun, J., Garimella, A., Chang, K.-W., & Peng, N. (2023). "kelly is a warm person, joseph is a role model": Gender biases in llm-generated reference letters. arXiv preprint arXiv:2310.09219.
- Wang, X., Zheng, L., Li, L., Zheng, Y., Sun, P., Zhou, F. A., & Guo, X. (2017). Immune to situation: The self-serving bias in unambiguous contexts. *Frontiers in Psychology*, 8, 822.
- Wieringa, M. (2020). What to account for when accounting for algorithms. *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, 1-18.
- Yu, K.-H., Beam, A. L., & Kohane, I. S. (2018). Artificial intelligence in healthcare. *Nature Biomedical Engineering*, 2(10), 719-731.