A Conceptual Model of Trust in Generative AI Systems

Nargess Tahmasbi
Pennsylvania State University
nvt5061@psu,edu

Elham Rastegari
Creighton University
elhamrastegari@creighton.edu

Minh Truong
Creighton University
minhtruong@creighton.edu

Abstract

Generative Artificial Intelligence (GAI) significantly impacts various sectors, offering innovative solutions in consultation, self-education, and creativity. However, the trustworthiness of GAI outputs is questionable due to the absence of theoretical correctness guarantees and the opacity of Artificial Intelligence (AI) processes. These issues, compounded by potential biases and inaccuracies, pose challenges to GAI adoption. This paper delves into the trust dynamics in GAI, highlighting its unique capabilities to generate novel outputs and adapt over time, distinct from traditional AI. We introduce a model analyzing trust in GAI through user experience, operational capabilities, contextual factors, and task types. This work aims to enrich the theoretical discourse and practical approaches in GAI, setting a foundation for future research and applications.

Keywords: Generative AI, trust, sector-specific applications

1. Introduction

Generative Artificial Intelligence (GAI) has emerged as a transformative force, unlocking vast potential across various sectors. Its applications span from healthcare advice (Abbasian et al., 2024) to self-education on specialized topics (Hadi et al., 2023) and even extend to creativity in fields like fashion and multimedia content creation (Cheng et al., 2023). Unlike traditional Artificial Intelligence (AI), which predominantly operated within organizational frameworks, the advent of GAI has democratized access, empowering individuals with newfound control over its utilization (Hadi et al., 2023).

While the proliferation of GAI promises substantial benefits, it also introduces challenges and risks. Unlike deterministic systems, GAI lacks theoretical guarantees regarding the correctness of generated outcomes, raising concerns about transparency and accountability. The opacity of AI processes complicates the ability to understand how decisions are made, posing potential hazards, including bias, discrimination, and healthcare inaccuracies (Bach et al., 2024). This seems to be of a greater issue in GAI, where the abundance of generated data modalities amplifies the potential for over trust or mistrust, eroding user confidence in GAI (Cheng et al., 2023).

Central to addressing these challenges is the concept of trust — a foundational element for the widespread adoption of GAI. Trust not only influences user acceptance but also underpins ethical considerations in AI design and deployment. Recognizing its pivotal role, various entities, including governmental bodies, have emphasized the importance of integrating trust into AI development frameworks (Glikson & Woolley, 2020).

Trust in the context of AI is a socio-technical construct, deeply intertwined with the interaction between users and AI systems (Benk et al., 2022). As these systems grow increasingly complex, understanding their inner workings becomes more challenging for users. This lack of transparency can compromise users' sense of control and, consequently, their ability to trust the technology.

The literature introduces various terms for trust that contribute to a rich theoretical discourse on what constitutes "trust in AI." (Langer et al., 2021; Yang et al., 2020; Schlicker & Langer, 2021). However, the absence of a unified definition hinders comparability between studies and delays progress in the field.

Despite extensive discussions around trust in AI, there remains a significant gap in understanding how trust specifically interacts with the unique attributes of GAI. However, existing research on trust in AI predominantly offers generic insights, failing to address the nuanced dynamics of GAI across diverse sectors. The differentiation between trust in traditional AI systems and GAI systems lies in their operational capabilities and the type of output they produce. Traditional AI systems often operate within defined parameters and predictable boundaries to provide recommendations or facilitate tasks, whereas GAI systems create novel, previously unseen outputs. This aspect of surprise and novelty may affect user trust differently, potentially enhancing it when the outcomes are perceived as beneficial and diminishing it when they are not. Furthermore, the ability of GAI to learn and adapt from iterative interactions introduces another layer of complexity. Users must trust not only what the system produces now but also its capacity to evolve and improve over time.

This fundamental difference impacts how user trust is evaluated, suggesting that trust in GAI may encompass additional layers of complexity related to the system's creative potential and the unpredictability of its outputs. Moreover, sector-specific studies are scarce, leaving a critical knowledge gap about how trust dynamics may vary across different fields such as healthcare, finance, and creative industries.

Therefore, this paper seeks to define a comprehensive understanding of trust in GAI, tailored to sector-specific sensitivities and implications. Our research endeavors to address the following key questions:

RQ 1. How is user trust in GAI defined?

RQ 2. What factors influence trust in GAI systems?

In summary, this paper seeks to bridge the gap in current research by offering sector-specific insights into trust in GAI and leveraging the adopted definition of trust to investigate the nuances of user expectations and judgments in scenarios of uncertainty or vulnerability. By examining the commonalities and divergences in trust measures between traditional AI and GAI systems, this study seeks to contribute to a deeper understanding of trust dynamics in the age of generative technologies.

2. Background and Conceptual Model

2.1. Trust Definition

Trust in AI encompasses the degree to which AI can be relied upon to make decisions (Pal et al., 2023). It involves cognitive trust, related to the perception of the AI's ability, reliability, and integrity; and affective trust, concerning the AI's perceived benevolence and disposition to do good (Kim et al., 2021). Trust also includes the reliance that the AI will not undertake actions harmful to the user's well-being (Hancock et al., 2011, p. 24), confidence in the AI's reliability and integrity (Morgan & Hunt, 1994), and the user's judgment or expectations about the AI system's assistance in situations of uncertainty or vulnerability (Amoozadeh, Daniels, Nam, Kumar, Chen, Hilton, Ragavan, et al., 2024; Cheng et al., 2023; Vereschak et al., 2021). Previous research acknowledges that there are multiple definitions of trust but selecting the most appropriate trust definition to depict trust in a specific context should be the focus instead of comparing definitions of trust (Bach et al., 2024).

For this research, we adopt a specific definition of trust as “the user’s judgement or expectations about how the AI system can help when the user is in a situation of uncertainty or vulnerability” (Amoozadeh et al., 2024). This definition is particularly resonant with the socio-technical system perspective, emphasizing the significance of trust in the interaction between humans and technology within real-world applications.

Moreover, current research often treats trust as a pre-existing condition, overlooking the dynamic nature of trust formation and the factors that facilitate or hinder its development. Thus, we also consider specific factors that contribute directly and indirectly to users’ trust in GAI in this research.

2.2. The Four-Party Framework

Previous research emphasizes the critical role of user trust in AI-enabled systems, advocating for a human-centric approach and identifying key influencing factors, including socio-ethical considerations, technical features, and user characteristics, to enhance trust and foster adoption (Bach et al., 2024). In the same line, a literature review conducted by Tamò-Larrieux et al. (2023) on trust identifies three key areas influencing trust in AI: the sphere of the trusting party (the human), the sphere of the trusted party (e.g., the artificial agent), and the sphere of the trusted act. This three-dimension framework has been utilized in human trust in AI studies with various focus areas such as studies concerning privacy, bias, and risk (Li et al., 2024) and trust in automated production management systems (Saßmannshausen et al., 2021).

We adopt the previous literature and integrate task type as a fourth dimension, emphasizing the importance of task type and condition suggested by the literature (Bach et al., 2024; Kaplan et al., 2023). The type of task significantly impacts users' trust in AI systems, although findings can be mixed. Klein et al. (2023) investigated the impact of task type on trust in AI by comparing high-impact and low-impact AI decisions in everyday work situations. The study found no significant difference between the two groups. However, in a recent study investigating users’ trust in GAI especially in sensitive areas such as healthcare, Choudhury & Shamszare (2023) found that trust significantly influences both users' intent to use and their actual use of ChatGPT. The research emphasizes the risks of overreliance on ChatGPT for health advice, suggesting the need for systems to direct such queries to human experts. This underscores the importance of considering task type when studying trust in AI, as different uses (e.g., healthcare vs. entertainment) involve varying levels of risk and trust requirements.

Our Four-Party framework encompasses the following dimensions: (1) the Trustee (The GAI system); (2) the Trustor (The human); (3) the Environment, Situation, and Condition in which the two interact, (4) the Task Type. In the following, we delve into each of these dimensions and explore the factors influencing human trust in GAI leveraging the literature studying the human trust in AI.

2.2.1. Party 1: The Trustee

Operational Capabilities

The **robustness** and **reliability** of an AI system are critical aspects of its competence, which in turn influence user trust. The ability of an AI system to maintain consistent performance and stability across varying conditions and input changes is crucial for fostering user trust. Users are more likely to trust an AI system that reliably handles different scenarios without a loss of functionality or performance. Lack of robustness can lead to unreliable predictions and unpredictable outputs which significantly undermine user trust (Hendrycks and Dietterich, 2019). Robustness-enhancing techniques can improve neural network performance in the face of anxiety and corruption, directly connecting robustness to increased user trust (Hendrycks and Dietterich, 2019).

In our model, we identify these elements as key predictors of trust, asserting that a system's consistent performance across various conditions enhances user confidence and trust. Thus, we hypothesize that enhanced robustness and reliability in GAI systems are positively correlated with greater user trust.

Ensuring **integrity** and **accountability** in AI systems is crucial for fostering user trust. When users are aware that an AI system, particularly its algorithms, is ethically created and consistently maintained, they are more likely to feel comfortable using the system. This awareness can significantly increase trust, as users are reassured that the system will not violate ethical principles or exploit individuals. Despite frequent references to the concept of *responsible AI*, the concepts of responsibility and accountability are often not clearly defined. Jobin et al. (2019) highlight this ambiguity in their analysis of global AI ethics guidelines. They emphasize that recommendations for responsible AI typically involve acting with integrity and clearly defining responsibility and legal liability, either upfront in contracts or by focusing on remedies.

Promoting diversity and integrating ethics into STEM education are also highlighted as key aspects of responsible AI. Ensuring that AI systems operate within ethical guidelines and are accountable for their actions has a significant impact on user trust. Transparency about the development of AI systems and assurances that these systems do not exploit, or harm users are critical for users to feel comfortable and ethically secure in their use of the technology (Jobin et al., 2019). Furthermore, the fear of users that AI cannot be held accountable for their actions prevents users' trust in AI (Li et al., 2024). Moreover, integrity relating to legal, ethical, and moral principles establishes a direct impact on users' trust (Cheng et al., 2023; Tamò-Larrieux et al., 2023). Similarly, in the context of GAI, we argue that users will trust and rely on the answers generated by GAI less when they are concerned about GAI's integrity and ability to be held accountable for its actions. In the consideration of the relationship between trust, accountability, and integrity above, we hypothesize that the concerns about the lack of accountability and integrity in GAI negatively impact users' trust in GAI systems.

We consider the factors of robustness and reliability as well as integrity and accountability a measure of a GAI's system's operational capabilities. Therefore, our first hypothesis is:

***H1**: Operational capabilities of a GAI system directly impact user's trust in the system.*

Novelty and Creativity: Unlike traditional AI systems, GAI systems are designed to create **novel** and **creative** content, which inherently involves a level of unpredictability. This unpredictability can be both a strength and a challenge. While it enables innovative applications, it also means that the outputs may sometimes be unexpected or inappropriate. We argue that if a GAI system generates creative solutions that consistently meet user needs, it builds trust. However, if the outputs are too unpredictable or occasionally problematic, trust can be undermined.

***H2a:** The ability of GAI systems to consistently deliver creative and novel content that meets user expectations positively impacts user trust. Conversely,*

***H2b:** Unpredictability in the creativity and novelty of outputs that do not align with user needs or expectations diminishes trust.*

Explainability and Transparency: GAI systems often operate using complex neural networks, such as those found in large language models (LLMs), which generate content that can be difficult to trace back to specific inputs or rules. This complexity can obscure the decision-making process, making it harder for users to understand why the AI made a particular decision or generated a specific output. Thus, the role of **transparency** and **explainability** seem to be far more important in the context of GAI compared to traditional AI. Transparency encompasses the transparency of the algorithm used, data sources, ethical guidelines, and regulations that the system adheres to, as well as the ability to explain why the system generates certain answers and whether it can provide answers for specific inputs due to the legality of the input.

AI's transparency and explainability are deeply intertwined. Research on Explainable AI (XAI) seeks to combat the black-box nature of AI and make algorithmic decision-making processes more interpretable and understandable. Understanding how and why AI makes decisions is crucial for user trust (Shin, 2021). However, this is challenging due to the opacity of neural networks and algorithm complexity (Li et al., 2024). Transparent AI models, which clearly explain their decision-making processes, help users trust the system's capabilities and accuracy and decreases users' judgments on who controls the system or its ability to represent human-characteristics (Alonso, 2018).

Ethical Considerations

Traditional AI follows predefined rules with predictable outputs, making it less prone to unexpected, harmful content. In contrast, GAI systems generate unpredictable content from vast datasets, increasing the risk of perpetuating **biases** and ethical concerns related to **fairness** and harm. This generative nature necessitates rigorous oversight and responsible development practices to mitigate these risks. Ethical considerations are important in understanding the trust dynamics in AI-based systems (Omrani et al., 2022). Fairness in GAI systems entail ensuring that these technologies do not perpetuate or exacerbate existing biases, which is crucial to maintaining social equity. A source of bias could be related to biases in training data in LLMs (Hadi et al., 2023). This issue underscores the importance of designing GAI systems that are not only technically proficient but also socially conscious, ensuring that they treat all user groups equitably. According to Li et al. (2024), transparency in AI design and implementation aids in identifying potential sources of bias, enabling developers to address these issues and ensure equitable and fair treatment of all users. Similarly, (Shin, 2021) found that explainability influences perceived bias, which in turn impacts trust.

Therefore, we propose the following:

H3: *Transparency and explainability positively impact user's perception of fairness in GAI systems.*

Data **privacy** and **security** are exceptionally important in the context of GAI, especially in sensitive applications like virtual financial advisors (Li et al., 2024). Users need to trust that their personal and financial information is safe when interacting with AI systems. Strengthening data protection mechanisms and ensuring robust security protocols are in place is essential for maintaining user trust. GAI systems that handle user data with high levels of security and compliance with data protection laws are more likely to be trusted by users. This is particularly important in sectors like finance, where the potential for misuse of data is high, and the impact of security breaches can be severe.

Ensuring the safety and ethical operation of GAI systems is crucial to prevent harm and build user trust. Omrani et al. (2022) stress the importance of designing GAI systems that are safe and non-maleficent, addressing ethical concerns such as discrimination and accountability to enhance trust. Transparent practices and the development of ethical guidelines are essential for minimizing risks and ensuring that these systems do not cause physical, psychological, or social harm. Moreover, Tamò-Larrieux et al. (2023) argue that well-structured legal frameworks, which enforce fairness, accountability, and transparency, are vital for fostering trust in AI technologies. This approach includes robust measures for fairness, privacy, security, and **harm mitigation**, underpinned by existing **regulations** to ensure ethical compliance and enhance the trustworthiness of GAI systems. We measure the ethical considerations factor through measuring fairness, privacy and security, harm mitigation, and existing regulations.

We propose the following hypothesis.

H4*: Perceived ethical considerations positively impact user's trust in GAI systems.*

2.2.2. Party 2: The Trustor

User perception encompasses several factors that influence users' interactions and perceptions of AI systems. Factors such as previous interactions or experiences with AI/GAI, self-efficacy, hedonic motivation, trust propensity, anthropomorphism, sense of control, language barriers, perceived benefits, perceived risks, and personalization all contribute to shaping users' responses to AI. Each of these elements plays a critical role in how users perceive, interact with, and ultimately trust AI technologies, affecting their overall satisfaction and engagement with these systems. In the following paragraphs, we will explore each factor in detail, providing a thorough explanation of how they contribute as measures of user experience in the context of user trust in GAI systems. This exploration aims to clarify how these elements serve as integral measures of user perception, influencing the degree of trust users place in GAI technologies. By comprehensively examining these factors, we seek to highlight their individual and collective impact on user interactions, perceptions, and overall satisfaction with AI systems. We hypothesize the following.

H5*: Cumulative positive user perception correlates with higher user's trust in GAI systems.*

Previous interactions or experiences with GAI play a crucial role in shaping users' expectations and beliefs about AI systems and directly impact user perceptions. Understanding users' trust in GAI systems requires examining how past interactions influence perceptions and expectations. Research indicates that users' expectations are shaped by their historical interactions with AI, significantly impacting their trust in the technology (Li et al., 2024). Negative experiences can lead to an overreaction to errors, reinforcing distrust (Vereschak et al., 2021).

The importance of prior user experiences is crucial, as demonstrated by how users' trust can be affected by others' reviews (Cheng et al., 2023). For example, negative feedback can deter potential users, highlighting the role of social learning in trust dynamics. Furthermore, AI systems that adapt and learn from user interactions enhance trust by showing the ability to evolve and meet users' needs (Glikson & Woolley, 2020).

Self-efficacy significantly impacts user perception and trust in AI and GAI systems. Defined as an individual's belief in their ability to succeed in specific situations (Li et al., 2024), self-efficacy influences how users perceive and interact with technology. Users with high self-efficacy are more likely to trust and adopt new technologies, as they feel competent and capable of managing complex systems. This trust is crucial, particularly when interacting with AI systems that may appear daunting due to their complexity. The perception of a system's usability and effectiveness directly impacts its acceptance, with less complex systems being more readily embraced by users (Venkatesh et al., 2012). Furthermore, users with a strong belief in their technological capabilities are more likely to see AI technologies as useful and manageable, which enhances trust and promotes the integration of these systems into daily activities (Bandura et al., 1999). Thus, understanding and addressing factors that influence self-efficacy can help in designing AI systems that are perceived as more user-friendly and trustworthy.

Hedonic motivation plays a pivotal role in enhancing user experience and trust within the context of GAI systems. This motivation arises from the intrinsic pleasure and satisfaction users derive from interacting with technology, rather than from any tangible outcome. When AI systems efficiently resolve challenges—thereby saving time and reducing the need for user problem-solving—this contributes significantly to user enjoyment and satisfaction. Such positive emotional responses enhance users' perceptions of the technology, fostering trust and encouraging repeated use. The ability of AI to provide quick and accurate solutions directly enhances the user experience, making it more enjoyable and hassle-free.

Zhang (2008) underscores the importance of hedonic motivation, highlighting that information and communication technology designs that maximize motivational affordances—those that enhance enjoyment and pleasure—play a critical role in fostering user engagement and trust. When applied to GAI, these principles not only streamline problem-solving but also create an engaging and pleasurable user experience. This successful integration of motivational affordances into GAI ensures that the technology meets users' hedonic needs, reinforcing trust through positive and satisfying interactions.

Trust propensity, which is an individual's innate inclination to trust others, plays a crucial role in the adoption and effective utilization of GAI systems. This inherent psychological trait deeply influences how users perceive and engage with AI technologies, significantly impacting their readiness to rely on these systems for critical decision-making. High trust propensity often leads users to quickly embrace GAI, apply it across various applications, and advocate for its benefits. Such predisposition not only accelerates the initial adoption of AI technologies but also enhances their integration into everyday activities (Li et al., 2024). Research has underscored the importance of trust propensity in shaping how users interact with new technologies. Individuals with high trust propensity tend to approach AI with less skepticism, which facilitates easier acceptance and promotes positive endorsements of the technology (Mcknight et al., 2011). This trait also affects how users respond to the AI's functionality and relationship dynamics, influencing overall satisfaction and likelihood of continued use. In the dynamic landscape of GAI, where system outputs can be highly variable and sometimes unpredictable, trust propensity becomes particularly significant. It dictates users' tolerance for errors and their readiness to adjust the system to better meet their needs.

Anthropomorphism significantly enhances user trust in AI and GAI systems by making interactions more human-like and relatable. This concept involves equipping AI systems with human-like traits that make them appear more like human beings. Research has shown that when users interact with AI systems that display such human-like characteristics, their trust levels increase, akin to how they would trust more familiar and intuitive human interactions, such as those with Siri or Alexa (Troshani et al., 2021). This humanization of AI not only makes the technology more accessible and easier to interact with but also enables the development of a more personable connection over time, which can deepen trust and user engagement (Van Brummelen et al., 2023). By adopting anthropomorphic features, GAI systems can become more engaging, fostering a trusting relationship that encourages continued use and acceptance.

Sense of agency, or control, is particularly crucial when it comes to user experience and trust in GAI systems, more so than in traditional AI. This concept refers to users' perception that they can oversee, influence, and directly impact the outcomes of AI operations. The ability to control, monitor, and correct the AI's actions enhances transparency and accountability, which are especially vital in GAI due to its ability to generate new and unexpected outputs. Such capabilities heighten the need for users to feel they can trust the system's decision-making processes (Li et al., 2024).

In GAI systems, a sense of agency is essential because these systems often operate with a higher degree of autonomy and complexity compared to standard AI applications. Enabling users to understand, modify, and intervene in the AI system's decisions and outputs through adjustable settings, transparent decision-making mechanisms, and override options is critical. These functionalities empower users, giving them a substantial role in managing the AI, thereby ensuring it acts in alignment with their expectations and needs. Providing a robust sense of control not only reduces user anxiety and uncertainty associated with the use of advanced automated technologies but also significantly boosts trust. Users feel more secure and confident in their capacity to manage the GAI, leading to increased trust as they can ensure the system operates transparently and accountably. This enhanced sense of agency is paramount in GAI environments, where the stakes of unexpected or uncontrolled outputs are potentially higher, making user control even more critical for fostering trust and acceptance.

Language barriers play a significant role in shaping user experience and ultimately affect trust in AI and GAI systems. These barriers emerge when users struggle to fully comprehend conversational agents or accurately interpret the responses provided by the technology, leading to misunderstandings and potential mistrust (Hadi et al., 2023). The advent of large-scale pretrained language models such as GPT-3 and GPT-4 has enhanced the capabilities of AI systems to perform various natural language processing (NLP) tasks, including answering questions and translating languages with high accuracy (Hadi et al., 2023). Despite these advancements, the effectiveness of these models can be compromised if the training data lack diversity. This oversight can lead to representation bias, resulting in the dissemination of missing or inaccurate information (Abbasian et al., 2024). To address this, it is crucial that virtual agents are designed to accommodate a wide array of cultural preferences, encompassing different languages, communication patterns, and even facial features that resonate with specific ethnic groups (Glikson & Woolley, 2020).

Furthermore, research indicates that users who interact with conversational agents in their native language tend to perceive these agents as more accurate and human-like. A study showed that participants reported a higher sense of correctness and relatability when using their first language compared to a secondary language when interacting with systems like Alexa (Van Brummelen et al., 2023). This finding underscores the impact of language proficiency on the user's perception of AI and GAI systems and highlights the importance of overcoming language barriers to build trust. Thus, effectively managing language barriers not only enhances the user's ability to understand and engage with AI systems but also significantly influences their trust levels. By ensuring that AI and GAI systems can effectively communicate across language divides, developers can foster a more inclusive and trustworthy environment for all users.

In summary, while language barriers are significant in both AI and GAI, their impact on trust can be more critical in GAI due to its advanced capabilities, higher user expectations, and broader application areas. Addressing these barriers effectively is crucial for ensuring that GAI systems are perceived as trustworthy and reliable.

Perceived Benefits and Risks Traditional AI systems typically focus on specific tasks such as data analysis and automation, where the benefits and risks are directly related to data privacy and algorithm biases. In contrast, GAI can produce highly complex and diverse outputs, including potentially misleading deepfakes, introducing new types of risks and benefits that necessitate a deeper understanding and careful management. When users recognize the advantages provided by algorithmic recommendation systems, such as enhanced efficiency, accuracy, and personalization capabilities, they are more likely to engage deeply with the technology. This recognition not only fosters a sense of algorithmic equity but also enhances social trust (Wu et al., 2024). Consequently, an awareness of these benefits can significantly boost user trust in the system. Conversely, an awareness of potential risks—such as concerns over privacy, security, potential errors, and the unpredictability of AI behavior—can adversely affect trust. Understanding how users perceive these benefits and risks is crucial for integrating these factors into a model of user trust in GAI.

Users' perceptions directly influence their attitudes and behaviors towards GAI systems. An increase in perceived benefits tends to enhance trust, while an increase in perceived risks tends to diminish it. These perceptions are shaped by factors such as users' previous experiences with AI, their knowledge about AI, and their individual risk tolerance levels.

Personalization in AI, particularly in GAI, plays a crucial role in enhancing user experience and building trust. Unlike traditional AI, which often provides basic recommendations based on historical data, GAI offers a more advanced level of personalization that can generate contextually relevant content. This capability significantly shapes user expectations and boosts satisfaction by catering more precisely to individual needs. Effective personalization in AI systems must account for a variety of demographic factors that influence trust. Research by Cheng et al. (2023) indicates that demographic characteristics such as gender, age, race, and regional access to technology can profoundly impact how users trust AI systems. For example, disparities in trust levels have been observed between different genders and among individuals from regions with varying levels of technological advancement. Amoozadeh et al. (2024) further demonstrate that factors like class standing, and gender differences can affect trust in AI, suggesting a nuanced approach to personalization is required to address these variations effectively. Moreover, the design of the user interface significantly contributes to how personalization affects user trust. An interface that is user-friendly, easy to navigate, and responsive to user needs not only makes the system more accessible but also enhances user confidence in the technology. A well-designed interface that accommodates individual preferences and needs can lead to increased user satisfaction, higher levels of trust, and greater likelihood of continued use of the AI system.

In summary, GAI's ability to deliver personalized experiences tailored to the specific preferences and circumstances of individual users is crucial for fostering trust.

*2.2.3. Party 3: The Environment (*Contextual Factors)

Socio-cultural factors play a significant role in shaping trust in AI and GAI systems, as these factors are deeply rooted in cultural differences and beliefs (Hadi et al., 2023). The impact of these differences can be profound, particularly in how users perceive and interact with AI technologies. Different cultural backgrounds influence users' expectations and interactions with AI systems. For instance, if a user poses a question related to a specific cultural or religious context, and the AI model has not been trained with adequate data from that cultural perspective, it might generate responses that appear biased or inappropriate. This occurs because the AI's responses are limited by the scope of its training data, which may not adequately represent all cultural or religious nuances (Hadi et al., 2023).

The sensitivity of AI systems to socio-cultural dynamics is crucial because inappropriate or culturally insensitive responses can erode trust and reduce the likelihood of users engaging with the technology. Ensuring that AI systems are trained on diverse datasets and are capable of understanding and respecting cultural differences is essential for building and maintaining trust. Socio-cultural factors often serve as mediating factors in the relationship between users' initial perceptions of AI and GAI systems (user experience) and their eventual trust in these technologies. These factors can influence how other attributes of the AI system, such as its communication style, functionality, or perceived intelligence, are interpreted by users from different cultural backgrounds. For example, expectations about privacy and data handling can vary significantly across cultures, affecting how security features are perceived and impacting trust indirectly. We hypothesize:

H6*: socio-cultural factors impact the users' perceptions of GAI systems.*

Specifically, the adequacy of an AI system's response to culturally or religiously specific queries—shaped by the system's training on diverse cultural data—significantly impacts trust. Users from diverse cultural backgrounds are likely to trust AI systems that accurately understand and respectfully respond to their culturally specific needs, while a lack of cultural sensitivity in AI responses reduces trust and user engagement.

Information and power asymmetries play crucial roles in shaping user trust in AI and GAI systems. These asymmetries arise from differences in users' knowledge and the organizational context in which AI systems are deployed. Users who possess a strong technological background exhibit less information asymmetry, which often translates into higher levels of trust and a greater willingness to adopt AI/GAI technologies. This phenomenon is based on the principle that a deeper understanding of technology reduces uncertainty and skepticism about its functionalities and outcomes (Tamò-Larrieux et al., 2023). In contrast, users with limited technological knowledge may feel more apprehensive and less confident, leading to lower trust levels. Power asymmetry also significantly influences trust, particularly concerning the size and reputation of the organizations that deploy AI/GAI systems. Systems operated by large, well-established organizations are often perceived as more credible and reliable due to the organizations' reputations and resources. Such entities are typically viewed as having higher stakes in maintaining user trust through responsible AI practices, thus fostering a greater degree of trust among users compared to smaller, less well-known organizations (Tamò-Larrieux et al., 2023). Information and power asymmetry directly impact user trust in AI/GAI systems in significant ways:

Information Asymmetry: Users with a deep understanding of AI technology tend to trust the system more, feeling confident in their ability to use and comprehend its operations. Conversely, users lacking this understanding may distrust AI due to feelings of vulnerability and uncertainty, fearing potential misuse.

Power Asymmetry: The trust level is also influenced by the organization's credibility behind the AI system. Users generally place greater trust in larger, well-established organizations as they are perceived to have more at stake in maintaining ethical practices and high standards of quality and safety.

Impact of information asymmetry and power asymmetry on trust can be influenced by other variables such as user experience, perceived utility, and the system's transparency. For example, a highly transparent AI system might mitigate the negative effects of information asymmetry by making it easier for users with limited technical knowledge to understand and trust the AI's decisions. Hence the following hypothesis:

H7: *Information and power asymmetries have significantly negative impact on user trust in GAI systems.*

2.2.4. Party 4: The Task Type

GAI encompasses different task types such as information seeking and learning, advice, and content creation due to the diverse needs and expectations of users across various applications. Each task type presents unique challenges and opportunities for GAI systems to build trust. For instance, mistakes made by AI in the medical field can lead to far more serious consequences (Omrani et al., 2022). Understanding which factors are more crucial in each task type can provide nuanced insights into the trust dynamics of GAI systems. We seek to investigate the GAI trust model in the following task categories.

Information Seeking and Learning: These tasks, including tutorials, coding assistance, and skill development, require GAI systems to prioritize factors such as accuracy, comprehensiveness, and educational effectiveness. Users trust these systems based on their ability to provide reliable information and guidance consistently. The trust model in these tasks hinges on the system's competence in delivering clear explanations and practical examples, thereby enhancing user learning and understanding. Investigating these factors can offer insights into how GAI systems can optimize educational content and foster greater trust through their knowledge and instructional capabilities.

Advice: In tasks involving advice, such as medical, legal, and relationship counseling, the trust model for GAI systems revolves around factors like expertise, credibility, and the ability to offer personalized recommendations. Users trust these systems to provide accurate and confidential advice that meets their specific needs. The effectiveness of GAI in these tasks depends on its capacity to handle sensitive information ethically and deliver trustworthy advice. Investigating these factors can provide deeper insights into how GAI systems can build and maintain trust by navigating complex scenarios and offering reliable insights that users can depend on.

Content creation: Content creation through GAI encompasses a wide range of activities, including writing, graphic design, multimedia production, and interactive storytelling. In both informational and entertainment contexts, these GAI systems are tasked with generating original, high-quality content that is creative, contextually appropriate, and aligns with user or business expectations. Trust in this domain is built on the GAI's ability to consistently deliver outputs that not only adhere to specific creative standards but also engage and captivate the audience. Whether the goal is to inform, promote, or entertain, users expect these systems to understand complex inputs and produce innovative content that enhances user engagement and satisfaction. Effective content creation GAI must therefore excel in interpreting user preferences and cultural nuances, ensuring that each piece of content is both relevant and enjoyable, thereby fostering a reliable and enriching user experience.

In our model, the task type serves as a control variable, influencing the impact of all other variables on trust in GAI systems. Task types such as the ones mentioned above introduce distinct challenges and requirements that significantly shape the trust dynamics of GAI systems. Figure 1 represents our conceptual model of trust in GAI. We plan to test this model in three different scenarios pertaining to the specific task types.

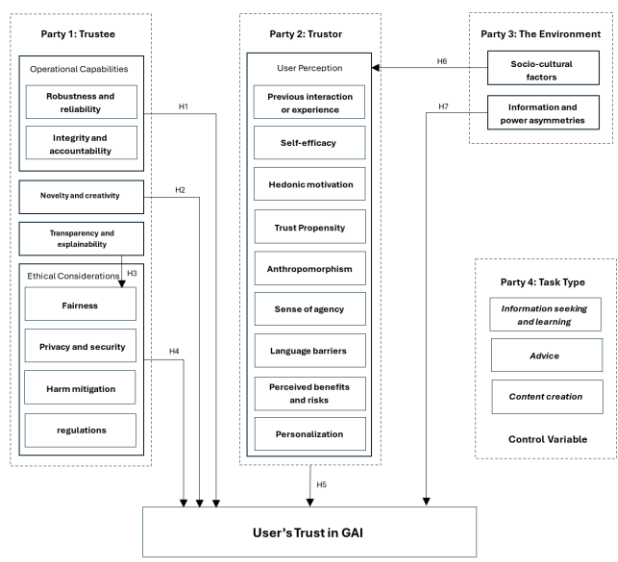

Figure 1. The Conceptual Model of Trust in GAI Systems.

3. Discussion and Conclusion

Our study examines trust in GAI by introducing a novel four-party framework. This framework encompasses: 1) System (Trustee) Factors: Including operational capabilities, ethical considerations, and fairness. 2) User (Trustor) Factors: Such as previous AI interactions, self-efficacy, hedonic motivation, trust propensity, anthropomorphism, sense of control, language barriers, perceived benefits, perceived risks, and personalization. 3) Environmental (Contextual) Factors: Covering socio-cultural factors and information and power asymmetries. 4) Task Type: Highlighting the unique challenges and opportunities for GAI systems across different task types like information seeking, advice, and content creation.

This framework underscores the critical role of task type in shaping user trust in GAI systems. By focusing on the distinct requirements and trust dynamics of various tasks, our study provides a comprehensive model that distinguishes the significant factors in GAI compared to traditional AI. Moreover, our contribution lays a foundational framework for measuring trust in GAI, which can be utilized in future research exploring GAI adoption. This comprehensive model addresses the multifaceted nature of trust in GAI, offering valuable insights for developers, researchers, and practitioners in the AI field.

To advance our research, we plan to empirically measure the impacts of the identified factors on trust in GAI. Central to our methodology is the utilization of questionnaires, leveraging self-reporting as a reliable tool to gauge user intentions and attitudes. Recognizing the intrinsic complexity of trust in AI, we acknowledge that trust cannot be fully understood through observable behavior alone. Trust is a deeply subjective experience, influenced by individual expectations, past experiences, and personal thresholds for risk and uncertainty. Therefore, our methodology places a strong emphasis on capturing users' internal states—those psychological dimensions that motivate and mediate their engagement with AI technologies. This emphasis is crucial because not all intentions lead to action. In this study, we focus primarily on the dimension of trust as an intention or attitude rather than observable actions such as the actual adoption of AI technologies. We are interested in understanding trust in AI per se, without necessarily linking it to subsequent behaviors like technology adoption.

Our methodology will involve survey and experimental studies to gather quantitative data on user interactions with GAI systems across different task types. These studies will help us understand how users, system, and environmental factors influence trust. We will also perform qualitative research to conduct In-depth interviews and focus groups to capture nuanced perspectives on trust in GAI. This qualitative approach will provide deeper insights into the contextual and socio-cultural dimensions affecting trust. Lastly, we will conduct a longitudinal study to examine how trust in AI evolves over time. This will help us understand the stability and changes in trust dynamics as users gain more experience with GAI systems.

By integrating these methods, we aim to develop a robust and comprehensive understanding of trust in GAI. This ongoing research will not only validate our theoretical framework but also provide practical guidelines for designing trustworthy GAI systems.

4. References

Abbasian, M., Khatibi, E., Azimi, I., Oniani, D., Shakeri Hossein Abad, Z., Thieme, A., Sriram, R., & others. (2024). Foundation Metrics for Evaluating Effectiveness of Healthcare Conversations Powered by Generative AI. *Npj Digital Medicine*, *7*(1), 82.

Alonso, V. (2018). System Transparency in Shared Autonomy: A Mini Review. *Frontiers in Neurorobotics*, *12*.

Amoozadeh, M., Daniels, D., Nam, D., Kumar, A., Chen, S., Hilton, M., Srinivasa Ragavan, S., & Alipour, M. A. (2024). Trust in Generative AI among Students: An exploratory study. *Proceedings of the 55th ACM Technical Symposium on Computer Science Education V. 1*, 67–73.

Bach, T. A., Khan, A., Hallock, H., Beltrão, G., & Sousa, S. (2024). A Systematic Literature Review of User Trust in AI-Enabled Systems: An HCI Perspective. *International Journal of Human–Computer Interaction*, *40*(5), 1251–1266.

Bandura, A., Freeman, W. H., & Lightsey, R. (1999). *Self-efficacy: The exercise of control*. Macmillan.

Benk, M., Tolmeijer, S., von Wangenheim, F., & Ferrario, A. (2022). *The Value of Measuring Trust in AI - A Socio-Technical System Perspective* (arXiv:2204.13480).

Cheng, R., Wang, R., Zimmermann, T., & Ford, D. (2023). 'It Would Work for Me Too': How Online Communities Shape Software Developers' Trust in AI-Powered Code Generation Tools. *arXiv*. http://arxiv.org/abs/2212.03491

Choudhury, A., & Shamszare, H. (2023). Investigating the impact of user trust on the adoption and use of ChatGPT: Survey analysis. *Journal of Medical Internet Research*, *25*, e47184.

Glikson, E., & Woolley, A. W. (2020). Human Trust in Artificial Intelligence: Review of Empirical Research. *Academy of Management Annals*, *14*(2), 627–660.

Hadi, M. U., Al Tashi, Q., Qureshi, R., Shah, A., Muneer, A., Irfan, M., Zafar, A., & others. (2023). *A Survey on Large Language Models: Applications, Challenges, Limitations, and Practical Usage*.

Hancock, P. A., Billings, D. R., Schaefer, K. E., Chen, J. Y., De Visser, E. J., & Parasuraman, R. (2011). A meta-analysis of factors affecting trust in human-robot interaction. *Human Factors*, *53*(5), 517–527.

Hendrycks, D., & Dietterich, T. (2019). Benchmarking neural network robustness to common corruptions and perturbations. *arXiv Preprint arXiv:1903.12261*.

Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence*, *1*(9), 389–399.

Kaplan, A. D., Kessler, T. T., Brill, J. C., & Hancock, P. A. (2023). Trust in artificial intelligence: Meta-analytic findings. *Human Factors*, *65*(2), 337–359.

Kim, J., Giroux, M., & Lee, J. C. (2021). When do you trust AI? The effect of number presentation detail on consumer trust and acceptance of AI recommendations. *Psychology & Marketing*, *38*(7), 1140–1155.

Klein, U., Depping, J., Wohlfahrt, L., & Fassbender, P. (2023). Application of artificial intelligence: Risk perception and trust in the work context with different impact levels and task types. *AI & SOCIETY*, 1–12.

Langer, M., Oster, D., Speith, T., Hermanns, H., Kästner, L., Schmidt, E., Sesing, A., & Baum, K. (2021). What do we want from Explainable Artificial Intelligence (XAI)?–A stakeholder perspective on XAI and a conceptual model guiding interdisciplinary XAI research. *Artificial Intelligence*, *296*, 103473.

Li, Y., Wu, B., Huang, Y., & Luan, S. (2024). Developing Trustworthy Artificial Intelligence: Insights from Research on Interpersonal, Human-Automation, and Human-AI Trust. *Frontiers in Psychology*, *15*, 1382693.

Mcknight, D. H., Carter, M., Thatcher, J. B., & Clay, P. F. (2011). Trust in a specific technology: An investigation of its components and measures. *ACM Transactions on Management Information Systems (TMIS)*, *2*(2), 1–25.

Morgan, R. M., & Hunt, S. D. (1994). The commitment-trust theory of relationship marketing. *Journal of Marketing*, *58*(3), 20–38.

Omrani, N., Rivieccio, G., Fiore, U., Schiavone, F., & Agreda, S. G. (2022). To Trust or Not to Trust? An Assessment of Trust in AI-Based Systems: Concerns, Ethics and Contexts. *Technological Forecasting and Social Change*, *181*, 121763.

Pal, A., Chua, A. Y., & Banerjee, S. (2023, March). Examining Trust and Willingness to Accept AI Recommendation Systems. *Proceedings of the Association for Information Science and Technology Mid-Year Conference: Expanding Horizons of Information Science and Technology and Beyond*.

Saßmannshausen, T., Burggräf, P., Wagner, J., Hassenzahl, M., Heupel, T., & Steinberg, F. (2021). Trust in artificial intelligence within production management–an exploration of antecedents. *Ergonomics*, *64*(10), 1333–1350.

Schlicker, N., & Langer, M. (2021). Towards warranted trust: A model on the relation between actual and perceived system trustworthiness. In *Proceedings of Mensch und Computer 2021* (pp. 325–329).

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.

Tamò-Larrieux, A., Guitton, C., Mayer, S., & Lutz, C. (2023). Regulating for Trust: Can Law Establish Trust in Artificial Intelligence? *Regulation & Governance*.

Troshani, I., Rao Hill, S., Sherman, C., & Arthur, D. (2021). Do we trust in AI? Role of anthropomorphism and intelligence. *Journal of Computer Information Systems*, *61*(5), 481–491.

Van Brummelen, J., Tian, M. C., Kelleher, M., & Nguyen, N. H. (2023). Learning Affects Trust: Design Recommendations and Concepts for Teaching Children—And Nearly Anyone—About Conversational Agents. *Proceedings of the AAAI Conference on Artificial Intelligence*, *37*(13), 15860–15868.

Venkatesh, V., Thong, J. Y. L., & Xu, X. (2012). Consumer Acceptance and Use of Information Technology: Extending the Unified Theory of Acceptance and Use of Technology. *MIS Quarterly*, *36*(1), 157–178.

Vereschak, O., Bailly, G., & Caramiaux, B. (2021). How to Evaluate Trust in AI-Assisted Decision Making? A Survey of Empirical Methodologies. *Proceedings of the ACM on Human-Computer Interaction*, *5*(CSCW2), 1–39.

Wu, W., Huang, Y., & Qian, L. (2024). Social Trust and Algorithmic Equity: The Societal Perspectives of Users' Intention to Interact with Algorithm Recommendation Systems. *Decision Support Systems*, *178*, 114115.

Yang, F., Huang, Z., Scholtz, J., & Arendt, D. L. (2020). How do visual explanations foster end users' appropriate trust in machine learning? *Proceedings of the 25th International Conference on Intelligent User Interfaces*, 189–201.

Zhang, P. (2008). Motivational Affordances: Reasons for ICT Design and Use. *Communications of the ACM*, *51*(11), 145–147.