Trust Me, I'm a Tax Advisor: Influencing Factors for Adopting Generative AI Assistants in Tax Law

Research Paper

Ben Möllmann[1], Leonardo Banh[1], Jan Laufer[1], and Gero Strobel[1]

[1] University of Duisburg-Essen, Rhine-Ruhr Institute of Information Systems, Essen, Germany
{leonardo.banh, jan.laufer, gero.strobel}@uni-due.de

Abstract. Generative AI is becoming increasingly important in areas such as tax law, where frequent legislative changes require up-to-date and accurate research. However, challenges such as bias, transparency, and hallucination can undermine user trust and hinder adoption. This paper explores trust as a critical factor for effective human-GenAI collaboration, focusing on a generative AI assistant for the domain of tax law. Using a mixed methods approach, we conduct quantitative questionnaires and qualitative expert interviews using two generative AI prototypes. The results show that factors such as transparency, anthropomorphism and compliance with social norms as well as moral standards positively influence trust, which in turn increases the intention to use generative AI assistants. By validating these trust determinants in a domain that requires rigorous accuracy, this study highlights the need to integrate trust-building measures when developing human-AI collaborative systems and provides valuable insights for designing more reliable and user-centered generative AI applications.

Keywords: Generative Artificial Intelligence, Human-GenAI Collaboration, Trust, GenAI Adoption

1 Introduction

The rapid advancement of artificial intelligence (AI), particularly generative AI (GenAI), fostered novel forms of human-AI collaboration across various fields, including finance, medicine and manufacturing (Banh & Strobel, 2023; Z.-H. Chen et al., 2021; Dellermann et al., 2019; Kunal et al., 2023; D. Wang et al., 2019). While GenAI can produce novel content, it also empowers conversational assistants by leveraging domain-specific knowledge and user-provided artifacts, thereby maintaining a controlled level of creativity (Banh & Strobel, 2023). In particular, large language models (LLMs) serve as powerful tools augmenting human capabilities, streamlining workflows, and reshaping human-performed tasks (Z. Chen & Chan, 2024; Eloundou et al., 2024). By blending human and AI capabilities, advantages emerge in efficiency, quality, creativity, safety, and overall human satisfaction (Boussioux et al., 2024; Schleiger

20th International Conference on Wirtschaftsinformatik,
September 2025, Münster, Germany

et al., 2024). However, challenges arise when applying GenAI in domains requiring precision, such as law, engineering, and medicine (Banh & Strobel, 2023; Feuerriegel et al., 2024). First, its outputs are often biased due to flaws in training data and algorithms. Second, its reasoning lacks transparency, functioning as a "black box", and its creative nature can lead to hallucinations, generating outputs based on nonexistent sources (Banh & Strobel, 2023; Feuerriegel et al., 2024). Thus, users' trust in GenAI capabilities and output is not guaranteed. Yet, a user's trust in novel technologies is essential for their adoption and effective use (Hasan et al., 2021; Yen & Chiang, 2021). For example, trusting AI leads to cognitive, affective, and behavioral changes (Lacity et al., 2024; Yang & Wibowo, 2022). Therefore, companies must understand how implementing AI affects this trust and identify the factors influencing users' confidence in AI systems (Berente et al., 2021; Gkinko & Elbanna, 2023).

Thus, to investigate user trust in a previously unexplored domain, this paper focuses on GenAI assistance in tax law. Tax law is characterized by dynamic developments, such as frequent legislative changes and court rulings (Haufe, 2024). Here, GenAI assistants offer improved knowledge sharing, despite concerns like hallucinations and generic responses, across various domains (Kernan Freire et al., 2023; Lewis et al., 2020). In the legal field, LLMs offer the possibility to enhance efficiency and accuracy in tasks such as legal research, provided they are supported by specialized data, tools to reduce biases, and methods to ensure transparency, and interpretability (Lai et al., 2024; Nay et al., 2024; Sun, 2023). Through human-GenAI collaboration, tax attorneys could save valuable time, enabling them to work more efficiently while dedicating greater focus to interpersonal engagement with their clients (Fagan, 2024). However, the domain-specific adoption of such tools relies heavily on user trust (Choi & Schwarcz, 2025). For instance, misplaced trust in a GenAI system in tax law can lead to severe financial penalties for clients, malpractice litigation, and career-ending professional sanctions for the attorney. Given these high-stakes outcomes, understanding the trust dynamics between legal experts and GenAI is critical.

Therefore, we investigate the role trust in generative AI plays in this human-GenAI collaboration by leveraging GenAI as the foundation for a digital tax advisor, focusing on delivering transparent, unbiased, and hallucination-free results to support tax attorneys in their daily work. We expect the accuracy and reliability of information to be crucial in relation to trust in AI (Abdallah et al., 2023; Bartels, 2023). Hence, the central research question of this paper is:

Which trust factors shape effective human-AI collaboration between tax attorneys and a generative AI-based tax advisor?

The intended outcome is to enhance the understanding of trust in human-AI collaborations and the use of generative AI. This involves exploring the potential adoption of AI technologies into tax attorneys' workflows. To achieve this, we first delve into the domain of tax law and how it can be assisted with GenAI-based information systems. Next, we outline our research methodology, i.e., the development of two software prototypes used in experiments and the evaluation in a mixed-method user study (a quantitative questionnaire followed by qualitative expert interviews) with tax attorneys. We present our findings and discuss their theoretical and practical implications. Finally, we conclude our work and provide an outlook on future research.

2 Computer-assistance for Tax Law

Tax law is a cornerstone of the economic framework, shaping financial flexibility through tax burdens while providing essential funding for public services. Tax law both influences and is influenced by other legal domains, such as civil and social law (Tipke et al., 2021). It is particularly advisable for businesses to involve a tax law expert, such as a tax advisor, in the preparation of tax declarations, as incorrect tax filings constitute a legal violation and can result in criminal prosecution in countries like Germany or the USA (Bundesfinanzministerium, 2024; Bundesministerium der Justiz, 2024; Internal Revenue Service, n.d.). To achieve optimal tax law assistance, the lifecycle of legal documents proposed by Pietrosanti and Graziadio (1999) serves as a guideline. Three key steps are essential: drafting legal documents, capturing supplementary information like cross-references and classification concepts, as well as navigating legal databases.

Information retrieval systems can support this process but, unlike generic question-answering systems, must ensure high quality and reliability, as their responses directly impact legal case outcomes (Abdallah et al., 2023; Sansone & Sperlí, 2022). Existing tools and services already in use include legal databases and legal expert systems (van Opijnen & Santos, 2017). Thus, Abdallah et al. (2023) emphasize the need of precise, legally substantiated answers with proper references, presented in a formal, structured language. These systems require expertise in laws, rulings, statutes, and regulations, along with frequent updates to address the complexity and dynamic nature of the legal landscape. Additionally, strict security measures are required to protect confidential data and comply with privacy laws.

GenAI's capabilities to analyze and understand natural language texts as well as to act as an assistant through chatbot interactions in various domains (Banh & Strobel, 2023; Feuerriegel et al., 2024) are promising for applications in the legal domain. This can involve either specially trained or general-purpose LLMs (Lai et al., 2024; Sun, 2023). The continuously improving reasoning abilities of LLMs, combined with data actuality ensured through Retrieval-Augmented Generation (RAG), can provide a robust foundation for legal expertise (Deroy et al., 2024; Nay et al., 2024; Savelka, 2023). Prompt engineering techniques enable formally structured and precise language, including proper citations (Yu et al., 2022, 2023). In combination with the aforementioned system requirements for legal IS, GenAI offers novel opportunities to support legal practitioners. Yet, an open question remains how a GenAI-based assistant must be designed to ensure adoption and usage by tax attorneys and what role factors such as *trust* play in this process.

3 Research Method

To identify the key factors influencing trust in human-GenAI collaboration and their impact in the context of tax law, we adopt a mixed-method approach following the six steps proposed by Venkatesh et al. (2013; 2016). The mixed-methods approach is highly appropriate as it collects a diverse set of quantitative and qualitative data to expand insights, confirm findings, and compensate the scarcity of available experts in the

tax law domain. The study's strategy aims to test the conceptual framework of Yang and Wibowo (2022) through confirmatory analysis. Adhering to the phases of conceptualization, investigation, and inference (Venkatesh et al., 2016), we employ a sequential mixed-methods design involving experiments with two prototypes to collect quantitative questionnaire data followed by qualitative interview data and subsequent analyses, valuing both data sources and benefiting from each research paradigm (Figure 1).

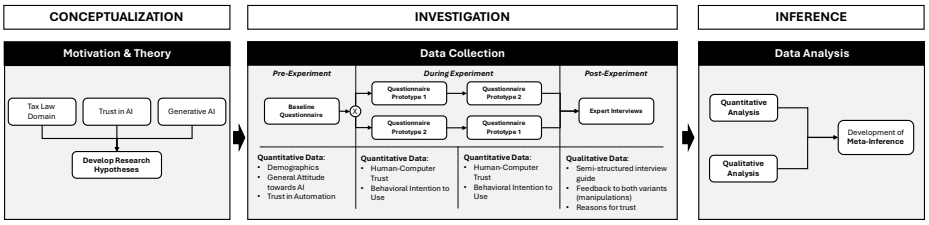

Figure 1. Research Design based on Venkatesh et al. (2013; 2016)

Next, we develop research hypotheses based on the AI trust framework by Yang and Wibowo (2022) (Table 1). The research model with its dependencies between the hypotheses is illustrated in Figure 2. H1, H2, and H3 propose *transparency*, *anthropomorphism*, and *compliance with social norms and moral standards* positively influence *trust in AI*, while H4 suggests *trust in AI* positively impacts the *intention to use AI/AI collaboration*. To test our propositions, we develop two AI-supported tax law advisor prototypes based on our hypotheses and a requirements analysis of relevant literature.

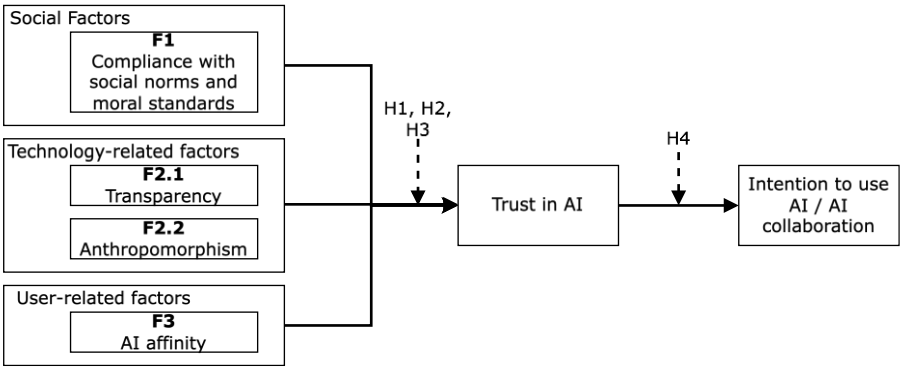

Figure 2. Research Model based on Yang and Wibowo (2022)

To collect and analyze study data, we employ a sequential mixed-methods sampling strategy, involving a within-subjects design with the same participants in both quantitative and qualitative phases (Venkatesh et al., 2016). The aim of the study is to manipulate the factors of our research hypotheses, so they are less pronounced in one prototype and more pronounced in the other. We collect quantitative data through a standardized questionnaire and gather qualitative data via semi-structured expert interviews.

All participants take part in experiments including the usage of both prototypes (in randomized order), completing identical questionnaire sections, and participating in follow-up interviews. The study focuses on tax advisors and legal practitioners experienced in legal research from adjacent legal domains, with at least as many tax law experts as non-tax law participants to maintain validity within the tax law context. To derive integrated conclusions and synthesize results (i.e., developing meta-inferences), the study uses a deductive approach focusing on evaluating hypotheses from mixed-methods results (Venkatesh et al., 2016). To assess the quality of our results and to address threats, quality criteria guide the presentation of results and conclusions, including a critical evaluation of limitations and a discussion of potential remedies.

Table 1. Hypotheses of Our Research Model

ID	Focus	Description	Theoretical Basis
H1	Transparency	Higher transparency of the AI assistant positively influences trust in AI.	Kawakami et al. (2022), Vössing et al. (2022), Yang and Wibowo (2022)
H2	Anthropomorphism	Higher anthropomorphism of the AI assistant positively influences trust in AI.	Yang and Wibowo (2022), Yen and Chiang (2021), J.-C. Lee and Chen (2022)
H3	Social and Ethical Compliance	Recognizable social and ethical compliance by the AI assistant positively influences trust in AI.	Yang and Wibowo (2022)
H4	Trust and Behavioral Intention	Higher trust in AI positively influences the intention for continued AI collaboration.	Hsiao and Chen (2022), Yang and Wibowo (2022)

3.1 Prototype Conceptualization

Before developing the two prototypes, we first identify the requirements for tax law research systems based on the foundations in Section 2. These are (*R1*) ensuring every response is supported by relevant sources, including laws, rulings, legal commentaries, statutes, and regulations; (*R2*) accessing specialized, up-to-date tax law data reflecting expertise in laws, jurisprudence, commentaries, statutes, and regulations; and (*R3*) providing precise, reliable, and contextually relevant responses to complex legal questions, adhering to the formal and structural conventions of legal language.

Next, the relevant manipulation factors for the experiments (Figure 2) are addressed, which will differentiate the prototypes. The previously mentioned requirements relate to the factor of *transparency*. Additionally, *anthropomorphism* and *social and ethical norm compliance* will also be manipulated. However, the factor *AI affinity* cannot be influenced through the prototype development. Table 2 provides an overview of how these factors are manipulated. Prototype Variant 1 (lower trust) excludes a greeting, displays only the current question and answer, identifies itself as an GenAI-powered tool, and omits source citations. In contrast, Prototype Variant 2 (higher trust) includes

a greeting with a brief self-introduction, shows a chat history of questions and answers, identifies itself as a GenAI-powered personal tax advisor, and provides context sections with source titles.

We develop our prototypes in Python, using RAG with the LangChain framework to provide source-supported answers. We use the embedding model *multi-lingual-e5-large* (L. Wang et al., 2024) for vector computation, with ChromaDB as the vector database. After testing and rejecting various LLMs due to incompatibility with the German language or poor linguistic quality, the focus shifted to the multilingual LLM *mixtral-8x7b-instruct-v0.1* by Mistral AI. This model outperforms leading LLMs like *Llama 2 70B* and *GPT-3.5* in most benchmarks at the time of its release (Jiang et al., 2024). To build a comprehensive dataset, we incorporate a well-regarded German trade tax manual as an authoritative source (Sternkiker, 2023). This reference is essential for equipping the generative AI tax advisor with reliable and accurate insights.

Table 2. Differentiation between Prototype Variants and Relating Trust Factor

Area of Difference	**Prototype**		**Factors**		
	Variant 1	**Variant 2**	**F1**	**F2.1**	**F2.2**
Greetings of the AI	No greeting	Greetings and introduction of abilities and guidelines	✓	✓	✓
Presentation of Dialog	Current question and answer	Chat history of current and past questions and answers		✓	✓
User Interface	Presented as an AI-powered tool	Presented as AI-powered personal tax advisor			✓
Citation of Sources	No citation	Citation of relevant context sections with source titles		✓	

3.2 Study Design

Questionnaire: The aim of the experiment study is to measure quantitatively in which prototype the participants have higher trust by asking them to complete job-related research tasks. Tasks include determining when a company becomes liable for trade tax, identifying companies exempt from trade tax, assessing the effects of conversions, mergers, or divisions on trade tax, and evaluating the trade tax implications of relocating a company's headquarters to another municipality. Our questionnaire collects demographic data, incorporates the *General Attitudes towards Artificial Intelligence Scale* (GAAIS) by Schepman and Rodway (2023), and uses Körber's (2019) *Trust in Automation* questionnaire in its German version, as AI is treated as a form of (work) automation in this study. After each experiment, participants assess their perceived technical competence and faith based on the *Human-Computer Trust* questionnaire (Madsen & Gregor, 2000). The questionnaire concludes with questions on *behavioral intentions to use* technology, adapted from Woen et al. (2018), that are tailored to the AI-supported tax law advisor prototypes. All scales are presented as 5-point Likert scales.

Expert Interview: The qualitative expert interview design follows Recker (2021) and employs explanatory interviews. This format is chosen to verify whether our presumed relationships and causal connections between trust in GenAI concepts are observed, perceived, or experienced by participants in our experiments (Schultze & Avital, 2011). Hence, this approach supports the study's goal of gathering data to test the research hypotheses. A semi-structured format is used, offering flexibility while maintaining sufficient structure for evaluating the hypotheses (Myers & Newman, 2007). Finally, the interview data is coded and analyzed after the principles of grounded theory (Gioia et al., 2013; Strauss & Corbin, 1996) to abductively derive in-depth insights regarding influencing trust factors in generative AI-based tax law advisors.

3.3 Procedure

The study was conducted virtually via Zoom in an online experiment (Fink, 2022). First, the study's framework and procedure were explained to the participants. They then received a link to the questionnaire hosted on LimeSurvey. Participants began by completing the first part of the questionnaire before proceeding with the first experiment. Task descriptions for both experiments were embedded in the questionnaire, with hyperlinks directing participants to the respective prototypes. For the second experiment, participants switched to the second prototype with guidance from the study instructor. Apart from a preview image included in the task description, participants encountered each prototype for the first time in-vivo during the experiments. After the first experiment, participants completed the second part of the questionnaire. The second experiment followed, using the second prototype. Upon completing the second experiment, participants proceeded to the third part of the questionnaire.

Finally, participants engaged in an expert interview, conducted by the study instructor and recorded via Zoom using its native recording function. This enabled transcription, analysis, and precise citation of participant responses for further examination.

4 Results

This section first presents the results of the quantitative experiment study, followed by the presentation of findings from the qualitative interview study. Then, the results will be discussed in light of our hypotheses. In total, six legal experts (n = 6) participated between May and June 2024 in our experiment study, including survey and interview. All participants have professional specializations in a legal field, with 50 % having a tax law background. The group comprised 66.67 % women. On average, the participants had four years of experience (SD = 2.608) in a (tax) legal field and 3.67 years of experience (SD = 2.066) in (tax) legal research.

4.1 Quantitative Analysis

For the upcoming data analysis, we consider mean values, standard deviations, and standard errors, evaluating the hypotheses based on paired t-tests with a focus on effect

sizes and statistical significance. The interpretation of the paired t-test results and calculation of effect sizes follow the approach of Field and Hole (2003), utilizing Cohen's (1992) benchmark scale: small effect ($r \geq 0.10$, explaining 1 % of variance), medium effect ($r \geq 0.30$, explaining 9 % of variance), and large effect ($r \geq 0.50$, explaining 25 % of variance). Additionally, we assess whether the *t*-value exceeds the critical *t*-value of 2.015 for 5 degrees of freedom at a 5 % significance level.

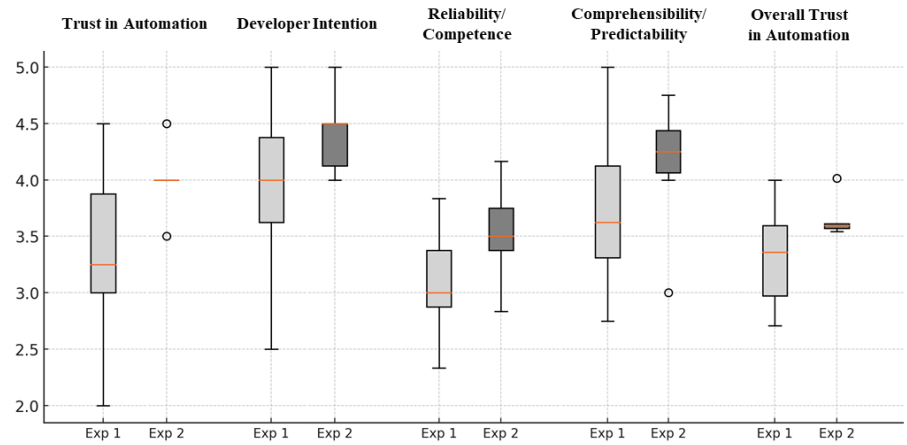

Figure 3. Trust Metrics Across Experiments

Trust in Automation: The first analysis focuses on participants' attitudes and trust in AI (Figure 3). In terms of the subscale *Trust in Automation*, the comparison between the two experiments revealed significant differences. Trust was notably higher after the second experiment (M = 4.0, SD = 0.3162) compared to the first (M = 3.333, SD = 0.8756), with a one-sided *p*-value of 0.041 and a large effect size (T(5) = 2.169, r = 0.696). On the subscale *Developer Intention*, the results also indicated a significant improvement following the second experiment (M = 4.4167, SD = 0.3764) compared to the first (M = 3.9167, SD = 0.8612). The one-sided *p*-value of 0.038 and a large effect size (T(5) = 2.236, r = 0.7071) demonstrated the statistical significance of this difference. Similarly, the subscale *Reliability/Competence* showed higher scores after the second experiment (M = 3.5278, SD = 0.4524) versus the first (M = 3.0833, SD = 0.5244), with a *p*-value of 0.049 and a large effect size (T(5) = 2.039, r = 0.6738). These findings underscore a significant increase in perceived reliability and competence. For the subscale *Comprehensibility/Predictability*, the second experiment (M = 4.125, SD = 0.6072) achieved higher scores than the first (M = 3.75, SD = 0.7906), but the difference was not statistically significant (p = 0.129, T(5) = 1.275, r = 0.4953). Overall, *trust in automation* was significantly higher after the second experiment (M = 3.6551, SD = 0.1777) compared to the first (M = 3.3241, SD = 0.4813). With a *p*-value of 0.038 and a large effect size (T(5) = 2.222, r = 0.705), this result confirms the observed increase in trust was unlikely due to chance. However, despite the significance of most sub-scale results, the 95 % confidence intervals for effect sizes included 0, indicating some uncertainty about the true effect sizes.

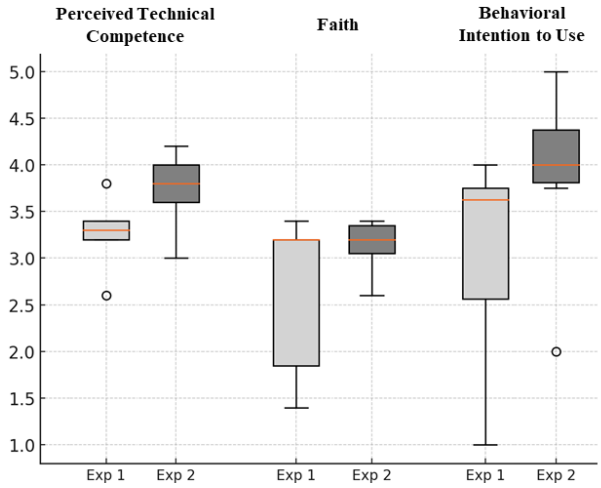

Figure 4. Boxplots for Perceived Technical Competence, Faith and Behavioral Intention to Use

Perceived Technical Competence and Faith: The *perceived technical competence* scale (Figure 4) showed a significant increase after the second experiment (M = 3.7333, SD = 0.432) compared to the first (M = 3.2667, SD = 0.3933), with a one-sided *p*-value of 0.049 and a large effect size (T(5) = 2.038, *r* = 0.6736). For the *Faith* scale (Figure 4), the average score was higher after the second experiment (M = 3.1333, SD = 0.3011) compared to the first (M = 2.6333, SD = 0.9584). However, this difference was not statistically significant ($p = 0.097$), despite a large effect size (T(5) = 1.499, $r = 0.5568$). The *t*-value below the threshold suggests the difference may be due to chance, and 95 % confidence intervals including 0 indicate effect size uncertainty.

Behavioral Intention to Use: The analysis of the *Behavioral Intention to Use* scale (Figure 4) revealed no statistically significant difference between the second experiment (M = 3.875, SD = 1.02164) and the first experiment (M = 3.0417, SD = 1.177), with a one-sided *p*-value of 0.072. Although a large effect size was observed (T(5) = 1.73, $r = 0.6119$), the *t*-value did not exceed the critical threshold, indicating the difference in means could be due to chance. Additionally, the 95 % confidence interval for the differences [-0.4047; 2.0713] includes 0, confirming no significant mean difference and reflecting uncertainty in the results.

Correlation: The analysis of the differences between experiments for the overall *Trust in Automation* scale and the *Behavioral Intentions to Use* scale revealed a Pearson correlation coefficient of $r = 0.680$, indicating a tendency for increased trust in AI to align with a greater intention for continued human-GenAI collaboration. However, this interpretation is constrained by the one-sided *p*-value of 0.069, which exceeds the study's significance threshold of $p < 0.05$ (Field & Hole, 2003). Thus, we find no significant correlation between *Trust in Automation* and *Behavioral Intention to Use*.

Summary: The manipulation of trust factors in the experiments revealed statistically significant large effects, as confirmed by t-tests, supporting the research hypotheses H1, H2, and H3. These hypotheses propose increased levels of *transparency*, *anthro-*

pomorphism, and *compliance with social norms and moral standards* by the AI assistant positively influence *trust in AI*. Additionally, the positive, though not statistically significant, Pearson correlation coefficient between *trust in automation* and *behavioral intentions to use* provides partial support for H4, which suggests higher trust in AI positively impacts the intention for continued AI collaboration.

4.2 Qualitative Analysis

The subsequent expert interviews explored individual preferences between the prototypes and provided insights into (generative) AI trust factors and their implications. Experts were also asked to suggest improvements and to assess the prototypes' suitability for tax law research. The coding and analysis process for the expert interviews follows the structure of Gioia et al. (2013) for grounded theory and combines both an inductive and a deductive approach (i.e., abductive). The deductive approach consists of defining the variables *transparency*, *anthropomorphism*, and *compliance with social norms and moral standards* as second-order themes, which are used to measure *trust in an AI tax law advisor*, representing the third-order dimension. First-order concepts are identified inductively through coding of the experts' statements.

Transparency: The experts collectively emphasize the *clear and helpful citation of sources* in the second prototype, supporting hypothesis H1. One expert linked this feature directly to trust, noting *"the second system is definitely more trustworthy, precisely because the sources have been specified"* (expert 2, translated from German). This positive feedback stemmed from factors such as *access to supplementary information*, *facilitation of further research*, and *time savings*. All in all, the importance of transparency and availability of legal sources for fostering trust in AI-generated content was emphasized. Participants also preferred the second prototype for its *ability to display conversation history*. Additionally, *clear communication of system capabilities and limitations* was noted as a trust-enhancing feature, further supporting hypothesis H1. Suggestions emphasized the *need for explicit section references* and *clear source type identification*, as their absence hindered trust.

Anthropomorphism: The experts enjoyed the *appealing design of human-like symbols* and the *personalized introduction or greeting by the AI tax advisor*. Furthermore, one expert noted the human-like communication increased their sympathy for the AI tax advisor: *"[...] through the different layout, such as in the form of a chat or a normal conversation, you felt more familiar and better supported"* (expert 3, translated from German). Another expert even attributed *increased trust* in the second prototype to its communication design, specifically highlighting the *question-and-answer format*.

Compliance with social norms and moral standards: The AI tax advisor's *greeting, which included references to its ethical guidelines, honesty, capabilities, and limitations*, was positively highlighted. The *disclosure of system limitations in handling legal questions* was also perceived favorably. Additionally, *warnings regarding the complexity of legal matters* were interpreted positively, particularly due to the potentially serious consequences of legal misinformation. However, dissatisfaction was expressed regarding the system's lack of flexibility when dealing with incomplete inputs. As one participant noted: "*And with the second [prototype], I once typed in bullet*

points. It said it couldn't find a result, and then I reformulated it as a sentence, and a result came out" (expert 5, translated from German). This suggests the positive impact of adhering to social norms and moral values may depend on how well the system meets user expectations regarding its capabilities.

Summary: The experts emphasized their willingness to use a GenAI tax law advisor depending on its perceived trustworthiness and they would not blindly trust AI. Their greater trust in the second prototype complemented by their preference for working with it supports H4. The respondents unanimously agreed that the second prototype was better suited for tax law research. The analysis also revealed usage intentions are influenced not only by trust but also by the convenience offered by the AI tax advisor.

5 Discussion

Both quantitative and qualitative analysis results support hypotheses H1–H4, subject to statistical significance discussed in Section 4.1. Yet, the evidence is reinforced by the qualitative results and their high degree of convergence, which enhances the plausibility of the quantitative findings. Particularly noteworthy are the complementary, independently derived statements from the interviewed experts, who suggested the manipulated features of the second prototype were responsible for their increase in trust.

In conclusion, we demonstrated that GenAI models can effectively serve as AI tax law advisors. Our study shows how (tax) law professionals favor the idea of collaborating with a GenAI assistant. Addressing our research question, we found the factors identified in Yang and Wibowo's (2022) conceptual trust framework are equally relevant in the tax law context. Specifically, *social*, *technology-related*, and *user-related* factors were shown to positively influence trust in AI, which in turn enhances the collaboration with an AI tax law advisor. Thus, our findings align with the conceptual framework proposed by Yang and Wibowo (2022) and extend the applicability for the paradigm of generative AI in the tax law and legal research domain.

Implications: Our work strengthens the external validity of Yang and Wibowo's (2022) framework by confirming the findings in the previously unexamined use case of tax law and legal research with the novel GenAI paradigm. Our work is also consistent with findings from other examined domains (Hsiao & Chen, 2022; J.-C. Lee & Chen, 2022; Vössing et al., 2022). However, regarding compliance with social norms and moral standards, our observations differ from those of Vössing et al. (2022), who found providing additional information about AI prediction uncertainty reduced experts' trust in the system and task outcomes. In contrast, our qualitative analysis revealed that experts generally viewed the disclosure of system limitations positively. It remains possible, however, that trust in specific AI-generated responses may have decreased, as this aspect was not explicitly measured. If differences from Vössing et al.'s (2022) findings do exist, they may be attributed to variations in use cases and domains. For instance, trust cues may be domain-sensitive and factors such as transparency could matter more in law than anthropomorphism, whereas the reverse may hold in retail or healthcare. This opens up opportunities for future research.

In terms of practical implications, the manipulations of trust factors should be considered as a basis for designing (generative) AI-based assistants, particularly the importance of providing diverse sources and transparent citations, including explicit references to text passages. For a GenAI-based tax law advisor, additional considerations include referencing *up-to-date sources*, specifying the *type of source*, providing *links to statutes and legal rulings*, and enabling *referrals to related legal topics*.

Limitations and Future Work: First, the qualitative analysis confirms the relevance of all three factors to trust in AI, yet it does not measure the degree of *perceived transparency*, *anthropomorphism*, or *social and ethical compliance*. As a result, it remains vague to what extent these factors differ between the two prototypes and how much each factor specifically contributed to the increase in trust. Future research could explore the contextual relevance of individual trust factors and develop a questionnaire specifically designed to measure trust in GenAI. There is also a risk of misinterpreting expert statements; however, this risk was mitigated by incorporating the perspective of all authors of this paper. Second, the quantitative results suggest a statistical possibility of chance in some observed effects. However, we employed robust statistical methods, clear thresholds, and reliable interpretation scales to ensure confidence in the results and conclusions (Venkatesh et al., 2016). Additionally, the expert interview evaluations in our mixed-method design further validate and enrich the quantitative findings. Last, the generalizability of the results is limited by the sample size, and the findings may not fully represent the target population. Scaling the study for further evaluations (e.g., in-situ work settings with larger participant pools) could provide deeper insights into adoption patterns and design knowledge. However, the convergence of results during the qualitative analysis supports the quality of the experts' opinions and strengthens the overall validity of the meta-inferences and the study as a whole (Venkatesh et al., 2016).

6 Conclusion

This study demonstrates the successful employment of a generative AI-based tax law advisor and its positive evaluation by experts. We hypothesized *trust in AI* as a key factor influencing the *intention to use AI*, which is, in turn, positively affected by the factors *compliance with social norms and moral standards*, *transparency*, *anthropomorphism*, and *AI affinity*. To test our assumptions, we designed and employed two GenAI tax law advisor prototypes, manipulated based on the identified factors. These prototypes were evaluated using experiments included in a mixed-methods approach. The evaluation confirmed our hypotheses, validating the identified factors.

Thus, our study contributes to the theoretical foundation of human-GenAI collaboration by demonstrating trust in AI plays a central role in the continued use of GenAI-based systems. This highlights the need to integrate trust as a key component into theoretical models of human-GenAI interaction and emphasizes the importance of trust-building measures as critical success factors. As GenAI becomes more prevalent in regulated professions such as medicine or law, its adoption will ultimately depend on how well its design fosters trust and aligns with professional standards.

References

Abdallah, A., Piryani, B., & Jatowt, A. (2023). Exploring the state of the art in legal QA systems. *Journal of Big Data*, **10**(1). https://doi.org/10.1186/s40537-023-00802-8

Banh, L., & Strobel, G. (2023). Generative artificial intelligence. *Electronic Markets*, **33**(1). https://doi.org/10.1007/s12525-023-00680-1

Bartels, S. (2023). *ChatGPT und Steuerberatung: Wie KI-Chatbots die Steuerbranche verändern können - Steuerberatung in Münster*. https://bartels-stb.de/chatgpt-und-steuerberatung/.

Berente, N., Gu, B., Recker, J., & Santhanam, R. (2021). Special Issue Editor's Comments: Managing Artificial Intelligence. *MIS Quarterly*, **45**(3), 1433–1450.

Boussioux, L., Lane, J. N., Zhang, M., Jacimovic, V., & Lakhani, K. R. (2024). The Crowdless Future? Generative AI and Creative Problem-Solving. *Organization Science*, **35**(5), 1589–1607. https://doi.org/10.1287/orsc.2023.18430

Bundesfinanzministerium (Ed.). (2024). *BMF-Monatsbericht Oktober 2024: Verfolgung von Steuerstraftaten und Steuerordnungswidrigkeiten im Jahr 2023*. https://www.bundesfinanzministerium.de/Monatsberichte/Ausgabe/2024/10/Inhalte/Kapitel-3-Analysen/3-1-verfolgung-von-steuerstraftaten-2023.html.

Bundesministerium der Justiz (Ed.). (2024). *Einkommensteuergesetz*. https://www.gesetze-im-internet.de/estg/.

Chen, Z., & Chan, J. (2024). Large Language Model in Creative Work: The Role of Collaboration Modality and User Expertise. *Management Science*, **70**(12), 9101–9117. https://doi.org/10.1287/mnsc.2023.03014

Chen, Z.-H., Lin, L., Wu, C.-F., Li, C.-F., Xu, R.-H., & Sun, Y. (2021). Artificial intelligence for assisting cancer diagnosis and treatment in the era of precision medicine. *Cancer Communications (London, England)*, **41**(11), 1100–1115. https://doi.org/10.1002/cac2.12215

Choi, J., & Schwarcz, D. (2025). AI Assistance in Legal Analysis: an Empirical Study. *Journal of Legal Education*, **73**(2). https://jle.aals.org/home/vol73/iss2/5.

Cohen, J. (1992). A power primer. *Psychological Bulletin*, **112**(1), 155–159. https://doi.org/10.1037/0033-2909.112.1.155

Dellermann, D., Ebel, P., Söllner, M., & Leimeister, J. M. (2019). Hybrid Intelligence. *Business & Information Systems Engineering*, **61**(5), 637–643. https://doi.org/10.1007/s12599-019-00595-2

Deroy, A., Ghosh, K., & Ghosh, S. (2024). Applicability of large language models and generative models for legal case judgement summarization. *Artificial Intelligence and Law.* https://doi.org/10.1007/s10506-024-09411-z

Eloundou, T., Manning, S., Mishkin, P., & Rock, D. (2024). Gpts are GPTs: Labor market impact potential of LLMs. *Science (New York, N.Y.)*, **384**(6702), 1306–1308. https://doi.org/10.1126/science.adj0998

Fagan, F. (2024). *A View of How Language Models Will Transform Law.* https://doi.org/10.48550/arXiv.2405.07826

Feuerriegel, S., Hartmann, J., Janiesch, C., & Zschech, P. (2024). Generative AI. *Business & Information Systems Engineering*, **66**(1), 111–126. https://doi.org/10.1007/s12599-023-00834-7

Field, A., & Hole, G. (2003). *How to design and report experiments*. Sage.

Fink, L. (2022). Why and How Online Experiments Can Benefit Information Systems Research. *Journal of the Association for Information Systems*, **23**(6), 1333–1346. https://doi.org/10.17705/1jais.00787

Gioia, D. A., Corley, K. G., & Hamilton, A. L. (2013). Seeking Qualitative Rigor in Inductive Research. *Organizational Research Methods*, **16**(1), 15–31. https://doi.org/10.1177/1094428112452151

Gkinko, L., & Elbanna, A. (2023). Designing trust: The formation of employees' trust in conversational AI in the digital workplace. *Journal of Business Research*, **158**, 113707. https://doi.org/10.1016/j.jbusres.2023.113707

Hasan, R., Shams, R., & Rahman, M. (2021). Consumer trust and perceived risk for voice-controlled artificial intelligence: The case of Siri. *Journal of Business Research*, **131**, 591–597. https://doi.org/10.1016/j.jbusres.2020.12.012

Haufe (Ed.). (2024). *Steueränderungen.* Haufe-Lexware GmbH & Co. KG. https://www.haufe.de/thema/steueraenderungen/.

Hsiao, K.-L., & Chen, C.-C. (2022). What drives continuance intention to use a food-ordering chatbot? An examination of trust and satisfaction. *Library Hi Tech*, **40**(4), 929–946. https://doi.org/10.1108/LHT-08-2021-0274

Internal Revenue Service. (n.d.). *Tax preparer penalties*. Accessed 24.01.2025, from https://www.irs.gov/payments/tax-preparer-penalties

Jiang, A. Q., Sablayrolles, A., Roux, A., Mensch, A., Savary, B., Bamford, C., Chaplot, D. S., Casas, D. d. l., Hanna, E. B., Bressand, F., Lengyel, G., Bour, G., Lample, G., Lavaud, L. R., Saulnier, L., Lachaux, M.-A., Stock, P., Subramanian, S., Yang, S., . . . Sayed, W. E. (2024). *Mixtral of Experts.* https://doi.org/10.48550/arXiv.2401.04088

Kawakami, A., Sivaraman, V., Cheng, H.-F., Stapleton, L., Cheng, Y., Qing, D., Perer, A., Wu, Z. S., Zhu, H., & Holstein, K. (2022). Improving Human-AI Partnerships in Child Welfare: Understanding Worker Practices, Challenges, and Desires for Algorithmic Decision Support. In S. Barbosa, C. Lampe, C. Appert, D. A. Shamma, S. Drucker, J. Williamson, & K. Yatani (Eds.), *CHI Conference on Human Factors in Computing Systems* (pp. 1–18). ACM. https://doi.org/10.1145/3491102.3517439

Kernan Freire, S., Foosherian, M., Wang, C., & Niforatos, E. (2023). Harnessing Large Language Models for Cognitive Assistants in Factories. In M. Lee (Ed.), *ACM Digital Library, Proceedings of the 5th International Conference on Conversational User Interfaces* (pp. 1–6). Association for Computing Machinery. https://doi.org/10.1145/3571884.3604313

Körber, M. (2019). Theoretical Considerations and Development of a Questionnaire to Measure Trust in Automation. In S. Bagnara (Ed.), *Advances in Intelligent Systems and Computing Ser: v.823. Proceedings of the 20th Congress of the International Ergonomics Association (IEA 2018): Volume VI: Transport Ergonomics and Human Factors (TEHF), Aerospace Human Factors and Ergonomics* (Vol. 823, pp. 13–30). Springer International Publishing AG. https://doi.org/10.1007/978-3-319-96074-6_2

Kunal, Rana, M., & Bansal, J. (2023). The Future of OpenAI Tools: Opportunities and Challenges for Human-AI Collaboration. In *2nd International Conference on Futuristic Technologies (INCOFT)* (pp. 1–6). IEEE. https://doi.org/10.1109/INCOFT60753.2023.10424990

Lacity, M. C., Schuetz, S. W., Le Kuai, & Steelman, Z. R. (2024). IT's a matter of trust: Literature reviews and analyses of human trust in information technology. *Journal of Information Technology.* https://doi.org/10.1177/02683962231226397

Lai, J., Gan, W., Wu, J., Qi, Z., & Yu, P. S. (2024). Large language models in law: A survey. *AI Open*, **5**, 181–196. https://doi.org/10.1016/j.aiopen.2024.09.002

Lee, J.-C., & Chen, X. (2022). Exploring users' adoption intentions in the evolution of artificial intelligence mobile banking applications: the intelligent and anthropomorphic perspectives. *International Journal of Bank Marketing*, **40**(4), 631–658. https://doi.org/10.1108/IJBM-08-2021-0394

Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., Küttler, H., Lewis, M., Yih, W., Rocktäschel, T., Riedel, S., & Kiela, D. (2020). Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks. In H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, & H. Lin (Eds.), *Advances in Neural Information Processing Systems* (Vol. 33, pp. 9459–9474). Curran Associates, Inc.

Madsen, M., & Gregor, S. (2000). Measuring human-computer trust. In *11th Australasian Conference on Information Systems (ACIS2000),* Brisbane, Australia.

Myers, M. D., & Newman, M. (2007). The qualitative interview in IS research: Examining the craft. *Information and Organization*, **17**(1), 2–26. https://doi.org/10.1016/j.infoandorg.2006.11.001

Nay, J. J., Karamardian, D., Lawsky, S. B., Tao, W., Bhat, M., Jain, R., Lee, A. T., Choi, J. H., & Kasai, J. (2024). Large language models as tax attorneys: A case study in legal capabilities emergence. *Philosophical Transactions. Series A, Mathematical, Physical, and Engineering Sciences*, **382**(2270), 20230159. https://doi.org/10.1098/rsta.2023.0159

Pietrosanti, E., & Graziadio, B. (1999). Advanced techniques for legal document processing and retrieval. *Artificial Intelligence and Law*, **7**(4), 341–361. https://doi.org/10.1023/A:1008304118095

Recker, J. (2021). *Scientific Research in Information Systems*. Springer International Publishing. https://doi.org/10.1007/978-3-030-85436-2

Sansone, C., & Sperlí, G. (2022). Legal Information Retrieval systems: State-of-the-art and open issues. *Information Systems*, **106**, 101967. https://doi.org/10.1016/j.is.2021.101967

Savelka, J. (2023). Unlocking Practical Applications in Legal Domain. In F. Andrade & M. Grabmair (Eds.), *Proceedings of the Nineteenth International Conference on Artificial Intelligence and Law* (pp. 447–451). ACM. https://doi.org/10.1145/3594536.3595161

Schepman, A., & Rodway, P. (2023). The General Attitudes towards Artificial Intelligence Scale (GAAIS): Confirmatory Validation and Associations with Personality, Corporate Distrust, and General Trust. *International Journal of Human–Computer Interaction*, **39**(13), 2724–2741. https://doi.org/10.1080/10447318.2022.2085400

Schleiger, E., Mason, C., Naughtin, C., Reeson, A., & Paris, C. (2024). Collaborative Intelligence: A Scoping Review Of Current Applications. *Applied Artificial Intelligence*, **38**(1), Article 2327890. https://doi.org/10.1080/08839514.2024.2327890

Schultze, U., & Avital, M. (2011). Designing interviews to generate rich data for information systems research. *Information and Organization*, **21**(1), 1–16. https://doi.org/10.1016/j.infoandorg.2010.11.001

Sternkiker, O. (Ed.). (2023). *Veranlagungshandbuch Gewerbesteuer 2022: Gewerbesteuergesetz in der für 2023 geltenden Fassung* (72. Auflage). IDW.
Strauss, A. L., & Corbin, J. M. (1996). *Grounded theory: Grundlagen qualitativer Sozialforschung*. Beltz.
Sun, Z. (2023). *A Short Survey of Viewing Large Language Models in Legal Aspect.* https://doi.org/10.48550/arXiv.2303.09136
Tipke, K., Seer, R., Hey, J., Englisch, J., & Hennrichs, J. (2021). *Steuerrecht* (24. neu bearbeitete Auflage). Otto Schmidt. https://doi.org/10.9785/9783504387150
van Opijnen, M., & Santos, C. (2017). On the concept of relevance in legal information retrieval. *Artificial Intelligence and Law*, **25**(1), 65–87. https://doi.org/10.1007/s10506-017-9195-8#Sec5
Venkatesh, V., Brown, S. A., & Bala, H. (2013). Bridging the Qualitative-Quantitative Divide: Guidelines for Conducting Mixed Methods Research in Information Systems. *MIS Quarterly*, **37**(1), 21–54. https://doi.org/10.25300/MISQ/2013/37.1.02
Venkatesh, V., Brown, S., & Sullivan, Y. (2016). Guidelines for Conducting Mixed-methods Research: An Extension and Illustration. *Journal of the Association for Information Systems*, **17**(7), 435–494. https://doi.org/10.17705/1jais.00433
Vössing, M., Kühl, N., Lind, M., & Satzger, G. (2022). Designing Transparency for Effective Human-AI Collaboration. *Information Systems Frontiers*, **24**(3), 877–895. https://doi.org/10.1007/s10796-022-10284-3
Wang, D., Weisz, J. D., Muller, M., Ram, P., Geyer, W., Dugan, C., Tausczik, Y., Samulowitz, H., & Gray, A. (2019). Human-AI Collaboration in Data Science. *Proceedings of the ACM on Human-Computer Interaction*, **3**, 1–24. https://doi.org/10.1145/3359313
Wang, L., Yang, N., Huang, X., Yang, L., Majumder, R., & Wei, F. (2024). *Multilingual E5 Text Embeddings: A Technical Report.* https://doi.org/10.48550/arXiv.2402.05672
Woen, A., Sylvia, C., Handoko, H., & Abdurachman, E. (2018). E-learning acceptance analysis using technology acceptance model (Tam). *Journal of Theoretical and Applied Information Technology*, **96**, 6292–6305.
Yang, R., & Wibowo, S. (2022). User trust in artificial intelligence: A comprehensive conceptual framework. *Electronic Markets*, **32**(4), 2053–2077. https://doi.org/10.1007/s12525-022-00592-6
Yen, C., & Chiang, M.-C. (2021). Trust me, if you can: a study on the factors that influence consumers' purchase intention triggered by chatbots based on brain image evidence and self-reported assessments. *Behaviour & Information Technology*, **40**(11), 1177–1194. https://doi.org/10.1080/0144929X.2020.1743362
Yu, F., Quartey, L., & Schilder, F. (2022). *Legal Prompting: Teaching a Language Model to Think Like a Lawyer.* https://doi.org/10.48550/arXiv.2212.01326
Yu, F., Quartey, L., & Schilder, F. (2023). Exploring the Effectiveness of Prompt Engineering for Legal Reasoning Tasks. In A. Rogers, J. Boyd-Graber, & N. Okazaki (Eds.), *Findings of the Association for Computational Linguistics: ACL 2023* (pp. 13582–13596). Association for Computational Linguistics. https://doi.org/10.18653/v1/2023.findings-acl.858