Salud, Ciencia y Tecnología. 2025; 5:2468
doi: 10.56294/saludcyt20252468

ORIGINAL

The Impact of AI-induced Ethical Conflict and the Moderating Role of Psychological Safety on the Mental Health of Accountants in Financial Technology Enterprises in Vietnam

Impacto del conflicto ético inducido por la Inteligencia Artificial y el rol moderador de la seguridad psicológica en la salud mental de los contadores en empresas de tecnología financiera en Vietnam

Hong Van Tran[1]

[1]University of Finance – Marketing. Ho Chi Minh City, Vietnam.

Cite as: Van Tran H. The Impact of AI-induced Ethical Conflict and the Moderating Role of Psychological Safety on the Mental Health of Accountants in Financial Technology Enterprises in Vietnam. Salud, Ciencia y Tecnología. 2025; 5:2468. https://doi.org/10.56294/saludcyt20252468

Submitted: 15-07-2025 **Revised:** 25-09-2025 **Accepted:** 06-11-2025 **Published:** 07-11-2025

Editor: Prof. Dr. William Castillo-González

Corresponding Author: Hong Van Tran

ABSTRACT

Amid the rapid integration of Artificial Intelligence (AI) within Vietnam's dynamic Fintech sector, this study investigates an emerging psychological challenge: AI-induced ethical conflict (AIEC). The study aims to test the impact of AIEC on the mental health (MH) of accountants, while also examining the moderating roles of psychological safety (PS) and work experience within the context of financial technology (Fintech) enterprises in Vietnam. To achieve this objective, a cross-sectional survey was conducted on a sample of 416 accountants. The data was analyzed using Structural Equation Modeling (SEM), including Confirmatory Factor Analysis (CFA) to validate the measurement scales and multi-group analysis to compare differences between experience groups. The analysis results show that AIEC has a strong, negative, and statistically significant impact on the mental health of accountants (β = 0,38). Notably, the study found that psychological safety, a traditional protective mechanism, did not demonstrate a moderating role in weakening this relationship. However, work experience emerged as a significant moderating factor, with the impact of AIEC on mental health being considerably stronger in the group of accountants with low experience compared to the high-experience group. These findings conclude that traditional psychological protection mechanisms, which focus on interpersonal interactions, may not be effective enough to cope with stressors that are systemic and algorithmic in nature. The study emphasizes the need to develop new intervention strategies, such as "algorithmic safety," and suggests that support programs should pay special attention to young employees, who are the most vulnerable.

Keywords: Artificial Intelligence; Ethical Conflict; Mental Health; Psychological Safety; Accountants; Fintech; Vietnam.

RESUMEN

En el contexto de la rápida integración de la Inteligencia Artificial (IA) en el dinámico sector Fintech de Vietnam, este estudio investiga un desafío psicológico emergente: el conflicto ético inducido por la IA (AIEC). El objetivo del estudio es examinar el impacto del AIEC en la salud mental (SM) de los contadores, así como analizar el rol moderador de la seguridad psicológica (SP) y la experiencia laboral en el contexto de las empresas de tecnología financiera (Fintech) en Vietnam. Para lograr este objetivo, se realizó una encuesta transversal con una muestra de 416 contadores. Los datos se analizaron utilizando el Modelado de Ecuaciones Estructurales (SEM), que incluyó un Análisis Factorial Confirmatorio (AFC) para validar las escalas de medición y un análisis multigrupo para comparar las diferencias entre los grupos de experiencia. Los resultados del análisis muestran que el AIEC tiene un impacto negativo, fuerte y estadísticamente significativo en la salud

mental de los contadores (β = 0,38). Notablemente, el estudio encontró que la seguridad psicológica, un mecanismo de protección tradicional, no demostró un rol moderador en la atenuación de esta relación. Sin embargo, la experiencia laboral sí fue un factor moderador importante, siendo el impacto del AIEC en la salud mental significativamente más fuerte en el grupo de contadores con menos experiencia en comparación con el grupo de mayor experiencia. Estos hallazgos concluyen que los mecanismos de protección psicológica tradicionales, enfocados en la interacción interpersonal, pueden no ser suficientemente eficaces para hacer frente a los estresores de naturaleza sistémica y algorítmica. El estudio subraya la necesidad de desarrollar nuevas estrategias de intervención, tales como la "seguridad algorítmica", y sugiere que los programas de apoyo deben prestar especial atención a los empleados jóvenes, quienes son los más vulnerables.

Palabras clave: Inteligencia Artificial; Conflicto Ético; Salud Mental; Seguridad Psicológica; Contadores; Fintech; Vietnam.

INTRODUCTION

The digital transformation is reshaping the global financial services industry, and Vietnam is emerging as a dynamic hub for financial technology innovation. With an impressive compound annual growth rate, Fintech enterprises in Vietnam are actively applying Artificial Intelligence (AI) to create a competitive advantage, from automating processes to making complex decisions in credit assessment, fraud detection, and asset management.[1,2,3] The deep penetration of AI not only changes business models but also profoundly reshapes the role of accounting professionals. They are no longer mere recorders and processors of data but have transitioned into supervisors, analysts, and strategic collaborators with AI systems.[4,5] This human-machine symbiosis, while promising superior efficiency, gives rise to unprecedented psychological and ethical challenges.

At the heart of these challenges is a new phenomenon we term "AI-induced ethical conflict" (AIEC). Building on the foundation of Role Conflict Theory,[6] accountants today must navigate the incompatibility between two job demands: on one hand, the obligation to adhere to core professional ethical principles such as integrity and objectivity, and on the other, the pressure to accept processes and decisions made by the organization's AI systems. This tension is exacerbated when viewed through the lens of Moral Distress Theory.[7] When AI, acting as a "system," provides recommendations based on algorithms with complex and non-transparent internal processing,[8] or potentially containing hidden biases,[9] accountants may recognize the right course of action but are prevented from implementing it by technological or organizational barriers.[10] This state of powerlessness can lead to profound psychological suffering, as highlighted in previous studies showing that employees often feel reduced engagement and trust when faced with non-transparent AI systems.[10,11,12] Therefore, we define AIEC as the psychological stress experienced by accountants when they perceive that decisions, recommendations, or processes generated by AI conflict with their professional judgment, professional ethical principles, or standards of fairness and transparency.

Previous studies have consistently shown that prolonged exposure to role conflict and moral distress are significant predictors of mental health problems, including symptoms of anxiety, depression, and stress.[13,14,15] Therefore, a fundamental and urgent research goal is set:

(i) To examine the impact of AI-induced ethical conflict (AIEC) on the mental health of accountants in Vietnamese Fintech enterprises

To cope with workplace stressors, organizations often seek to cultivate a psychologically safe environment. Defined as a shared belief that an individual will not be punished or humiliated for speaking up about mistakes or risks,[16,17] psychological safety is considered a powerful protective mechanism. Theoretically, in a safe environment, accountants would feel empowered to discuss and question problematic AI decisions, thereby helping to alleviate stress and find appropriate solutions. This leads to the second research objective:

(ii) To analyze the moderating role of psychological safety in weakening the negative relationship between AI-induced ethical conflict and the mental health of accountants

However, applying existing knowledge to this context raises a challenging question. Previous research has convincingly argued that psychological safety acts as an effective "buffer," mitigating the impact of interpersonal stressors; but does this protective mechanism, which focuses on human interaction, retain its effectiveness when faced with a completely new type of conflict that is s..ystemic and algorithmic in nature? AI-induced ethical conflict is not pressure from a colleague or a superior; it is a tension arising from the interaction with an algorithmic entity whose decision-making process is too complex to be fully explained, operating on logic and objectives that may be alien to human ethical judgment.

The necessity of this study stems from the potential incompatibility between these two concepts. We

hypothesize that solutions focused on human-to-human interaction, which are the foundation of psychological safety, may be insufficient to effectively address a type of stress that is systemic and technological in nature. By testing this model in the context of Vietnamese Fintech - a unique environment for Human-AI interaction - this study not only seeks to fill a critical knowledge gap but, more importantly, re-evaluates the foundational assumptions about psychological protection mechanisms in the digital age. To further explore the nuances of this phenomenon, we will also examine potential differences based on work experience, this leads to the third research objective:

(iii) To test for differences in the proposed research model (regarding the impact of AI-induced ethical conflict on mental health and the moderating role of psychological safety) among groups of accountants with different years of work experience

By addressing the aforementioned objectives, this study aims to initiate a dialogue on whether new theoretical models and intervention strategies are necessary to safeguard the mental well-being of the workforce from the psychological ramifications of Artificial Intelligence.

METHOD

Research design and sample

Research Design: This study employs an observational study approach, utilizing a cross-sectional survey design. This design is considered highly appropriate as it allows for the collection of data on key variables (ethical conflict, psychological safety, and mental health) from a large sample of individuals at a single point in time. This approach is not only efficient in terms of time and resources but is also well-suited for examining the proposed relationships and moderating roles within our theoretical model.

Data Collection Period: Primary data for this research were collected over the period from April 12, 2025, to August 16, 2025.

Research Setting and Participants: the research was conducted in the major economic and technological hubs of Vietnam, specifically Hanoi, Ho Chi Minh City, and Da Nang. These cities were selected due to their high concentration of financial technology (Fintech) companies. The target population consists of accountants, financial specialists, and individuals in equivalent roles working within Fintech enterprises that have implemented and utilize Artificial Intelligence (AI) systems in their professional operations.

Due to the lack of a formal and comprehensive sampling frame for this specialized population, we adopted a combined non-probability sampling strategy, including convenience sampling and snowball sampling. To ensure the suitability of the sample, a clear set of screening criteria was established. An individual was considered eligible to participate in the study if they met all the following conditions simultaneously: (i) Professional Role: Has worked as an accountant, internal auditor, or financial specialist for 1 year or more; (ii) Work Environment: Is currently employed at a company identified as a financial technology firm in Vietnam; and (iii) Interaction with Technology: has frequent direct or indirect interaction with AI-based systems/tools in their daily professional work. The recruitment process was carried out by contacting existing professional connections within the Fintech industry. After an individual was identified as meeting the criteria and agreed to participate, they were encouraged to refer the survey to other colleagues they knew who also met the same conditions.

Regarding sample size, after the data collection and cleaning process, the final sample size used for analysis was 416 valid responses. We used an a priori power analysis conducted with G*Power 3,1 software. The results indicated that to detect a small effect size ($f^2 = 0,02$) with a desired statistical power of 0,80 and a significance level of $\alpha = 0,05$ in a regression model with 3 predictors (including the interaction term), a minimum sample size of 395 is required. Therefore, the final sample size of 416 not only meets but exceeds the minimum requirement, ensuring the study has sufficient statistical power to detect significant relationships and minimize the risk of a Type II error.

Data collection procedure

Data was collected through an online questionnaire designed on the Google Forms platform. A formal invitation letter was sent to potential participants. This letter clearly stated the academic purpose of the study, emphasized that participation was completely voluntary, and guaranteed the absolute confidentiality of personal information and the anonymity of responses.

Ethical Considerations

This study strictly adhered to the ethical principles for research involving human subjects. Prior to participation, all individuals were provided with detailed information regarding the study's purpose and content, and were assured that their involvement was entirely voluntary. Informed consent was obtained from each participant. Anonymity and the confidentiality of personal information were absolutely guaranteed; all collected data were coded and used solely for research purposes. Recognizing the sensitive nature of topics

related to mental health, the questionnaire was designed to minimize any potential distress, and participants had the full right to withdraw from the study at any time without explanation.

Measures

To ensure content validity, all scales for the latent variables were adapted or developed based on foundational theories and reputable prior empirical research. The items were measured on a 5-point Likert scale, ranging from "1 = Strongly disagree" to "5 = Strongly agree". Table 1 below provides a list of all measurement items used in the survey, along with their coding, academic source, and detailed content.

Table 1. Measurement Scales in the Study

Variable	Code	Items
AI-induced Ethical Conflict (Developed for this study)	AIEC1	I often feel a conflict between following the AI system's recommendations and my own professional judgment.
	AIEC2	The AI system sometimes provides financially optimal solutions that conflict with the ethical standards of the accounting profession.
	AIEC3	I feel pressured when I have to take responsibility for decisions based on suggestions from an AI system whose operational logic I do not fully understand.
	AIEC4	Relying on AI to handle complex situations makes me concerned about the erosion of my own ethical judgment skills.
	AIEC5	There are times when I have to choose between following an AI-proposed procedure and doing what I believe is ethically right.
	AIEC6	I feel stressed when I have to justify a decision suggested by AI that goes against the interests of stakeholders (e.g., clients, employees).
	AIEC7	The lack of transparency in how the AI reaches its conclusions creates an ethical gray area in my work.
	AIEC8	I worry that excessive automation with AI may lead to overlooking important human factors in financial decisions.
	AIEC9	I find it difficult to balance the efficiency brought by AI with the ethical obligations of an accountant.
Psychological Safety (Edmondson[16])	PS1	If I make a mistake on this team, it is not held against me.
	PS2	Members of my team are able to bring up problems and tough issues.
	PS3	People on this team sometimes reject others for being "different". (R)
	PS4	It is safe to take a risk on this team.
	PS5	It is difficult to ask other members of this team for help. (R)
	PS6	No one on this team would deliberately act to undermine my efforts.
	PS7	Working with members of this team, my unique skills and talents are valued and utilized.
Mental Health (Stress) Adapted from Lovibond et al.[18]	MH1	I found it hard to wind down.
	MH2	I tended to over-react to situations.
	MH3	I felt that I was very irritable.
	MH4	I felt I was using a lot of nervous energy.
	MH5	I was worried about situations in which I might panic and make a fool of myself.
	MH6	I felt I was close to panic.
	MH7	I was aware of dryness of my mouth.
	MH8	I experienced trembling (e.g., in the hands).
	MH9	I couldn't seem to experience any positive feeling at all.
	MH10	I found it difficult to work up the initiative to do things.
	MH11	I felt that I had nothing to look forward to.
	MH12	I felt that life was meaningless.

Note: (R) indicates a reverse-coded item.
Source: adapted from Edmondson[16] and Lovibond[18]

AI-induced Ethical Conflict (AIEC): Due to the novelty of this concept, a new 9-item scale was developed

specifically for this study. The scale development process followed these steps: (i) Item Generation: This step was not only based on exploratory interviews with 8 Fintech accountants to capture the practical context but was also built on a solid multidisciplinary theoretical foundation. Specifically, ideas for the items were synthesized and adapted from three theoretical pillars: Moral Distress Theory (e.g., Hamric et al.[11]), to capture the psychological suffering when an individual knows the right thing to do but is hindered by system barriers (in this case, AI); Role Conflict Theory Rizzo et al.[6], to frame the conflict between the role of adhering to professional ethical standards and the role of complying with AI-proposed operational procedures; and studies in the field of Human-AI Interaction, which highlight issues of transparency and accountability when working with algorithms, a direct source of conflict; (ii) Content Validity Assessment: A panel of 4 experts (2 scholars in AI ethics, 2 senior financial managers in the Fintech industry) evaluated the clarity and relevance of each item. Items with a Content Validity Ratio (CVR) below 0,8 were eliminated or revised. (iii) Pilot test: The draft scale was tested on a small sample (n=40) to check for clarity of wording and to assess preliminary reliability. (iv) Finalization: Based on the analysis from the pilot test, the final 9-item scale was finalized for use in the main study.

Psychological Safety (PS): we used the widely recognized 7-item scale by.[16] To ensure conceptual and semantic equivalence, a rigorous back-translation procedure following Brislin[19] was applied. Specifically, the original English scale was translated into Vietnamese by a bilingual expert. Then, another bilingual expert, working independently and unaware of the original version, translated the Vietnamese version back into English. The two English versions were then compared to identify and correct any discrepancies, ensuring that the final Vietnamese version accurately conveyed the meaning of the original concept.

Mental Health (MH): participants' mental health was measured using a validated Vietnamese version of the Depression, Anxiety, and Stress Scale.[18] However, to minimize respondent burden and optimize the survey completion rate among a group of professionals with limited time, a shortened 12-item version was used. The selection of the 4 items with the highest factor loadings from each subscale (Depression, Anxiety, Stress) was based on reputable psychometric analyses of the scale's structure, particularly in the Vietnamese context and studies on shortened versions.[20] We argue that this approach helps maintain the core content validity of the scale while ensuring conciseness, thereby enhancing the quality of the collected data.

To ensure the accuracy of the model and isolate the main relationships, the study controlled for several demographic and work-related variables. These variables included Age (continuous), Gender (dummy variable: 1=Male, 2=Female), Work Experience (number of years, continuous), Job Position (ordinal variable: 1=Staff, 2=Specialist, 3=Manager), and AI Interaction Level (5-point Likert scale). The selection of these variables was based on previous research showing their potential influence on mental health, thus allowing for a more accurate estimation of the main variables' impact.

Data analysis method

The study used SPSS 26 and AMOS 26 for data analysis, applying the two-step procedure of Structural Equation Modeling (SEM). The first stage focused on assessing the measurement model. After data cleaning and screening, a Confirmatory Factor Analysis (CFA) was conducted to examine the quality of the scales. Scale reliability was confirmed through Cronbach's Alpha and Composite Reliability (CR > 0,7). Convergent validity was assessed using factor loadings (> 0,5) and Average Variance Extracted (AVE > 0,5). To ensure the concepts were clearly distinct, discriminant validity was thoroughly examined using both the Fornell-Larcker criterion and the HTMT ratio (< 0,85).

After the measurement model was validated, the second stage involved testing the structural model to answer the research questions. The overall fit of the model was evaluated using indices such as CFI, TLI (>0,9), RMSEA, and SRMR (<0,08). To achieve the first research objective, we examined the statistical significance of the path coefficient from AI-induced ethical conflict to mental health. The second objective was addressed by testing the coefficient of the standardized interaction term. Finally, to address the third objective, a multi-group analysis was performed after confirming measurement invariance to compare differences in the model between low and high experience groups.

Furthermore, we recognized that since the study's data was collected from a single source at a single point in time, Common Method Bias (CMB) was a potential issue to consider.[21] To mitigate this risk, we implemented preventive measures during the survey design, such as ensuring anonymity and randomizing the order of items. To more rigorously assess the impact of CMB, we conducted a full collinearity assessment as recommended by [22]. According to this method, we built a regression model in which all latent variables (both independent and dependent) predicted a random variable, and the Variance Inflation Factor (VIF) indices of these latent variables were examined. Kock[22] suggests that if all VIF values are below 3,3, the model can be considered free from common method bias. Our analysis showed that all VIF values of the latent variables ranged from 1,28 to 2,52, significantly lower than the 3,3 threshold. This result provides strong evidence that CMB is not a significant concern and does not distort the study's findings.

RESULTS

Descriptive statistics

Table 2 summarizes the demographic characteristics of the 416 participants. The sample shows a significant gender disparity, with females constituting the majority (70,9 %). The average age of participants was 31,5 years, and the average work experience was 8,2 years. Regarding job positions, the majority of participants were staff (54,1 %), followed by specialists (32,5 %) and managers (13,4 %). Notably, the mean score for the level of interaction with AI was 3,92 on a 5-point scale, indicating that AI systems were relatively deeply integrated into the daily work of the accountants in the sample.

Table 2. Demographic Characteristics of the Sample (N=416)

Characteristic	Category	Frequency (n)	Percentage (%)	Mean (M)	Std. Dev. (SD)	Min	Max
Gender	Male	121	29,1				
	Female	295	70,9				
Age				31,50	5,83	23	52
Work Experience				8,21	6,14	1	29
Job Position	Staff	225	54,1				
	Specialist	135	32,5				
	Manager	56	13,4				
AI Interaction Level				3,92	0,88	2	5

Table 3 presents the descriptive statistics, preliminary reliability (Cronbach's Alpha), and correlation matrix among the main latent variables of the study. The mean score for AI-induced Ethical Conflict (AIEC) was 3,45 (Standard Deviation = 0,95), which is above the scale's midpoint, suggesting that this is a prevalent issue in the sample. Conversely, the mean score for Psychological Safety was 3,61 (Standard Deviation = 0,85). Mental Health (MH), measured on a negative scale (higher scores indicate poorer mental health), had a mean value of 2,88 (Standard Deviation = 0,91). The Cronbach's Alpha coefficients for all scales exceeded 0,85, indicating very good internal consistency.

The correlation matrix reveals initial relationships consistent with the research hypotheses. Specifically, AIEC was positively and significantly correlated with MH ($r = 0,48$, $p < 0,001$), suggesting that as ethical conflict increases, mental health tends to worsen. Conversely, PS was negatively and significantly correlated with both AIEC ($r = -0,35$, $p < 0,001$) and MH ($r = -0,52$, $p < 0,001$). This indicates that a psychologically safe environment is associated with reduced ethical conflict and improved mental health. These correlations provide a solid foundation for testing the structural model in the subsequent steps.

Table 3. Descriptive Statistics, Reliability, and Correlation Matrix

Variable	M	SD	α	1	2	3	4	5	6
1. Age	31,50	5,83	-	1					
2. Gender	-	-	-	0,04	1				
3. Experience	8,21	6,14	-	0,82***	0,02	1			
4. AIEC	3,45	0,95	0,89	0,09	0,06	0,12*	1		
5. PS	3,61	0,85	0,91	-0,05	-0,08	-0,07	-0,35***	1	
6. MH	2,88	0,91	0,93	-0,15**	0,11*	-0,18**	0,48***	-0,52***	1

Note: N = 416. α = Cronbach's Alpha. M = Mean. SD = Standard Deviation. Gender: 1=Male, 2=Female. AIEC = AI-induced Ethical Conflict; PS = Psychological Safety; MH = Mental Health. ***, **, and * denote statistical significance at the 1 %, 5 %, and 10 % levels, respectively

Measurement Model Assessment

To assess the validity and reliability of the measurement scales, a Confirmatory Factor Analysis (CFA) was conducted on the three-factor measurement model (AIEC, PS, and MH) using AMOS 26 software. The results showed that the measurement model fit the data well, with all fit indices meeting the recommended thresholds ($x^2/df = 2,45$, CFI = 0,95, TLI = 0,94, RMSEA = 0,06, SRMR = 0,05).

Table 4 presents the detailed results of the reliability and convergent validity assessment. The results show that all standardized factor loadings were highly statistically significant ($p < 0,001$) and greater than the 0,5

threshold, ranging from 0,68 to 0,89. The Composite Reliability (CR) values all far exceeded the 0,7 threshold (from 0,91 to 0,94), and the Average Variance Extracted (AVE) values were all higher than the 0,5 level (from 0,58 to 0,65). These results collectively provide strong evidence that the scales used in the study meet the requirements for reliability and convergent validity.

Table 4. Results of Reliability and Convergent Validity Assessment (N=416)

Construct & Items	Standardized Factor Loading	CR	AVE
AI-induced Ethical Conflict (AIEC)		0,91	0,58
AIEC1	0,75		
AIEC2	0,79		
AIEC3	0,82		
AIEC4	0,71		
AIEC5	0,80		
AIEC6	0,77		
AIEC7	0,83		
AIEC8	0,68		
AIEC9	0,72		
Psychological Safety (PS)		0,92	0,61
PS1	0,74		
PS2	0,81		
PS3 (R)	0,70		
PS4	0,85		
PS5 (R)	0,73		
PS6	0,82		
PS7	0,79		
Mental Health (MH)		0,94	0,65
MH1	0,78		
MH2	0,75		
MH3	0,80		
MH4	0,84		
MH5	0,82		
MH6	0,81		
MH7	0,76		
MH8	0,79		
MH9	0,89		
MH10	0,86		
MH11	0,88		
MH12	0,87		

Note: CR = Composite Reliability; AVE = Average Variance Extracted. (R) = reverse-coded item. All factor loadings are significant at $p < 0,001$.

Discriminant validity was examined using two rigorous criteria and is presented in table 5.

Table 5. Results of Discriminant Validity Assessment (N=416)

	AIEC	PS	MH
AIEC	0,76	0,42	0,57
PS	-0,38	0,78	0,61
MH	0,51	-0,55	0,81

Values on the main diagonal (in bold) are the square roots of the Average Variance Extracted (AVE). Values in the lower triangle are the correlation coefficients between constructs. Values in the upper triangle are the HTMT ratios. Discriminant validity is established when the diagonal values are greater than the off-diagonal correlations in the corresponding rows and columns, and the HTMT values are < 0,85. Values in the lower triangle are the correlation coefficients between latent constructs estimated from the CFA model.

According to the Fornell-Larcker criterion, the values on the main diagonal (square roots of AVE, in bold) are all greater than their correlation coefficients with all other concepts (the values in the lower triangle). Additionally, to further reinforce this, the Heterotrait-Monotrait ratio of correlations (HTMT) are presented in the upper triangle of the table. The results show that all HTMT ratios are significantly lower than the strict threshold of 0,85. Meeting both of these criteria convincingly demonstrates that the three research concepts - AI-induced Ethical Conflict, Psychological Safety, and Mental Health - are empirically distinct.

Test the structural model

Structural Model Fit

After validating the measurement model, we proceeded to test the overall structural model to assess its fit with the empirical data. The analysis results indicate that the model has very good fit indices: $x^2/df = 2,58$; GFI = 0,92; CFI = 0,96; TLI = 0,95; RMSEA = 0,058 (90 % Confidence Interval: 0,049 – 0,067), and SRMR = 0,052. All these indices meet or exceed the widely accepted standard thresholds in research (e.g., $x^2/df < 3$; GFI, CFI, TLI > 0,90; RMSEA, SRMR < 0,08). This confirms that the proposed theoretical model is highly compatible with the collected data, providing a solid basis for testing the research hypotheses in the next step.

Results of the Path Analysis

The results of the path analysis are presented in detail in table 6 and figure 1. The model explains 39 % of the variance in Mental Health ($R^2 = 0,39$).

Table 6. Path Analysis Results for the Structural Model (N=416)

Path	Standardized Beta (ß)	Standard Error (S.E.)	p-value	Result
Main Effects				
AIEC -> MH	0,38	0,05	< 0,001	Supported
PS -> MH	-0,48	0,04	< 0,001	Supported
Moderation Effect				
AIEC x PS -> MH	0,04	0,04	0,352	Not Supported
Control Variables				
Age -> MH	-0,07	0,03	0,041	Supported
Gender -> MH	0,05	0,04	0,189	Not Supported
Experience -> MH	-0,09	0,03	0,015	Supported
AI Interaction Level -> MH	0,11	0,04	0,008	Supported

Note: AIEC = AI-induced Ethical Conflict; PS = Psychological Safety; MH = Mental Health. Standardized Beta coefficients are reported

To provide a comprehensive and visual overview of the tested relationships, the final structural model with standardized path coefficients is presented in figure 1.

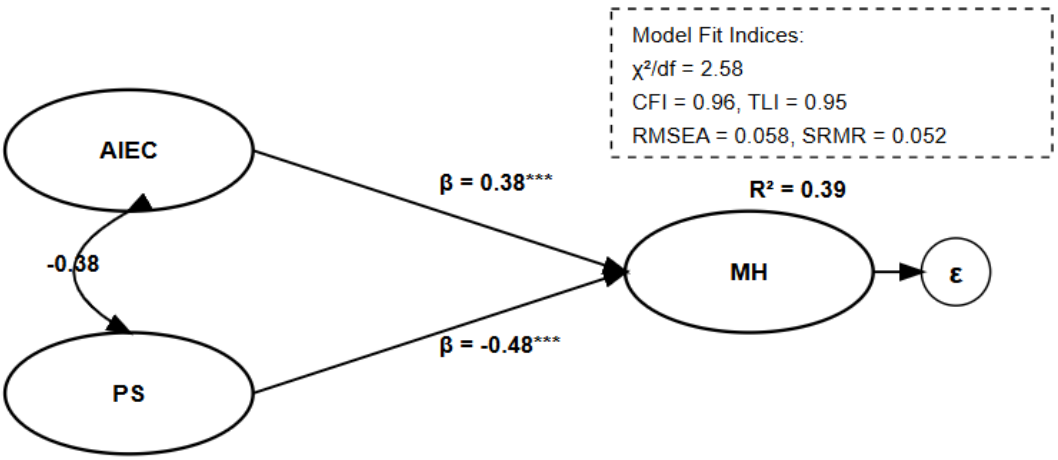

Figure 1. Final Structural Model Results

Regarding the first research objective, we examined the path from AI-induced Ethical Conflict to Mental Health. The results show that AIEC has a positive and highly statistically significant impact on MH (β = 0,38, p < 0,001). Since the MH scale is coded in a negative direction (higher scores indicate poorer mental health), this result confirms that as the level of AI-induced ethical conflict increases, the mental health of accountants significantly deteriorates.

Regarding the second research objective, we tested the moderating role of Psychological Safety by examining the effect of the interaction term (AIEC x PS) on MH. The analysis results show that the coefficient of the interaction term is not statistically significant (β = 0,04, p = 0,352).

This result, though contrary to initial expectations, reveals a concerning reality. It shows that while AI-induced ethical conflict is a strong negative factor affecting accountants' mental health, the widely recognized protective mechanism of psychological safety does not serve to mitigate this impact. This implies that even when empowered to speak up in a safe environment, accountants still suffer psychological harm stemming from the systemic, algorithmic, and non-transparent nature of this specific type of conflict.

Figure 2 visually illustrates this lack of a moderating effect. The two lines representing the relationship between AIEC and MH at high and low levels of psychological safety are nearly parallel, confirming that the negative impact of AIEC on mental health does not diminish even in an environment considered to be safe. This result leads to the rejection regarding the moderating role of psychological safety and poses a major challenge to existing theories of psychological health protection.

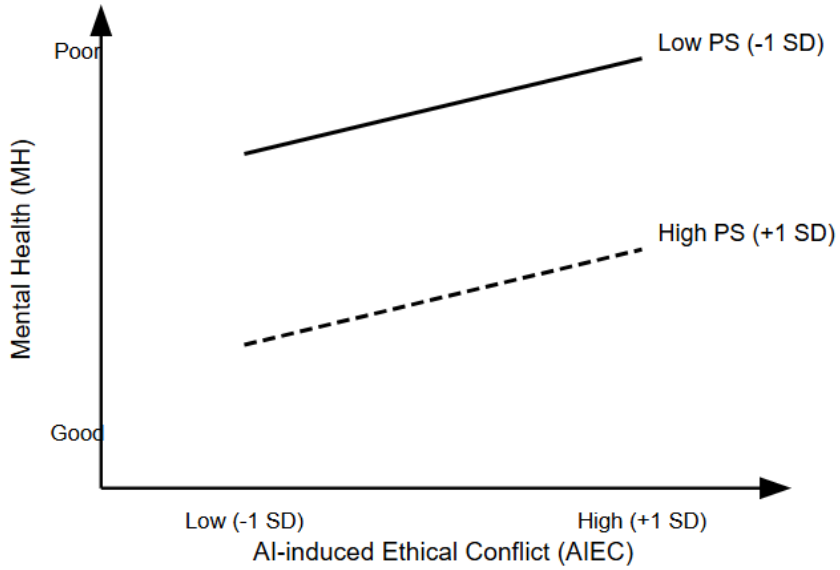

Figure 2. Moderating Effect of Psychological Safety

Multi-group Analysis by Work Experience

To address the third research objective—exploring whether the impact model differs between groups of accountants with different work experience—we conducted a multi-group analysis. Based on previous studies on career development and stress-coping abilities, we divided the sample into two groups: the "Low Experience" group (≤ 3 years of work, n = 145) and the "High Experience" group (> 3 years of work, n = 271). The choice of the 3-year mark is considered reasonable, as this is often the stage when an employee transitions from the early phase of their career to a phase of stability and accumulated professional expertise.

Measurement Invariance Test

Before comparing the path coefficients between the two groups, a prerequisite step is to test the measurement invariance of the scales. This is to ensure that the concepts (AIEC, PS, MH) are understood and interpreted similarly by both groups, thereby making the comparison meaningful. We conducted a sequential test across three levels: configural invariance, metric invariance, and scalar invariance. The results are presented in table 7.

Table 7. Measurement Invariance Test Results Across Experience Groups

Model	Invariance Level	x^2	df	x^2/df	CFI	RMSEA	Comparison Model	ΔCFI	Conclusion
Model 1	Configural	1285,4	544	2,36	0,942	0,058	-	-	Baseline fit established
Model 2	Metric	1301,9	569	2,29	0,938	0,056	Model 1	-0,004	Invariance supported
Model 3	Scalar	1388,7	594	2,34	0,925	0,057	Model 2	-0,013	Invariance not supported

Note: A ΔCFI ≤ -0,01 indicates a significant decrease in model fit; therefore, invariance is not supported at that level.

The results in table 7 show:

The configural model (Model 1), the baseline model with no constraints between groups, has good fit indices (CFI = 0,942, RMSEA = 0,058), indicating that the factor structure is similar in both groups.

When comparing the metric model (Model 2) with Model 1, the change in the CFI index (ΔCFI = -0,004) is smaller than the 0,01 threshold. This confirms that metric invariance is established. This means that the factor loadings of the scales are equivalent between the low and high experience groups.

However, when testing the scalar model (Model 3), the ΔCFI value is -0,013, which exceeds the allowable threshold. This indicates that scalar invariance is not established.

Although scalar invariance was not achieved, the successful establishment of metric invariance is a sufficient condition to allow us to reliably compare the path coefficients (beta coefficients) in the structural model between the two groups.

Structural Model Comparison between Groups

After confirming metric invariance, we proceeded to compare the path coefficients of the main relationships between the two experience groups. The study used a Chi-square difference test to determine whether the differences in these coefficients were statistically significant. The results are summarized in table 8.

Table 8. Multi-group Analysis Results and Path Coefficient Difference Test

Path	Low Experience Group (≤ 3 years)	High Experience Group (> 3 years)	Difference Test
	ß (p-value)	ß (p-value)	Δx^2 (df=1)
AIEC -> MH	0,51 (< 0,001)	0,24 (0,012)	5,18
AIEC * PS -> MH	0,05 (0,581)	0,03 (0,715)	0,15

Note: ß represents the standardized beta coefficient. Δx^2 is the Chi-square difference value from constraining the corresponding path coefficient to be equal across the two groups

The results of the multi-group analysis provide profound and meaningful findings for the third research objective.

The analysis shows that the negative impact of AI-induced Ethical Conflict (AIEC) on Mental Health (MH) is statistically significantly stronger in the group of accountants with low experience (ß = 0,51; p < 0,001) compared to the group with high experience (ß = 0,24; p = 0,012). The Chi-square difference test confirmed this difference to be significant ($\Delta x^2(1) = 5,18$; p = 0,023).

This indicates that accountants who are new to the profession or have fewer years of experience are a

significantly more vulnerable group to the psychological stress arising from AI-induced ethical conflict. It can be argued that this group has not yet accumulated sufficient confidence in their professional judgment, has not developed effective coping mechanisms, and may feel greater pressure to comply with the procedures proposed by the organization's "intelligent" systems. In contrast, more experienced accountants, with a solid professional foundation and higher self-confidence, are better able to "resist" this type of conflict, thus mitigating the negative impact on their mental health.

At the same time, the results also show that the interaction effect (AIEC x PS) remains statistically insignificant in both groups, further reinforcing the conclusion from RQ2 that psychological safety is not an effective moderator for this type of conflict, regardless of work experience.

To visualize this important result, figure 3 illustrates the difference in the slope of the relationship between AIEC and MH in the two experience groups.

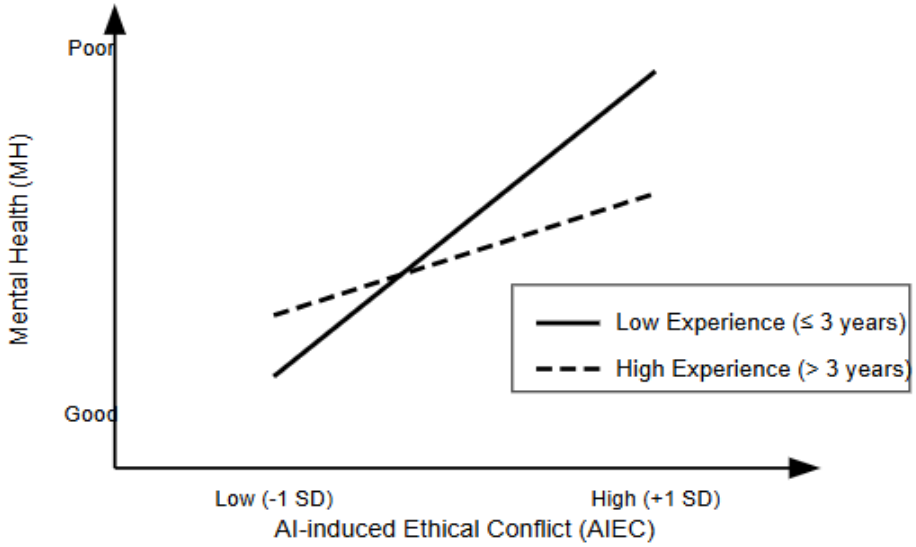

Figure 3. The Moderating Effect of Work Experience on the AIEC -> MH Relationship

DISCUSSION

Our study has yielded three core findings: (i) AI-induced ethical conflict (AIEC) has a strong and negative impact on the mental health of accountants; (ii) Most notably, psychological safety (PS) does not exhibit a moderating role, meaning it fails to weaken this negative relationship; (iii) The harmful impact of AIEC on mental health is significantly more pronounced and stronger in the group of accountants with low experience compared to the high-experience group.

The first finding (AIEC -> MH, β = 0,38) provides strong empirical evidence supporting the study's theoretical foundation. This result is perfectly consistent with Role Conflict Theory[6] and Moral Distress Theory.[7] It shows that when accountants have to deal with the conflict between their duty to adhere to professional judgment and ethics and the pressure to accept decisions from an AI system whose internal logic is not fully interpretable, they experience significant psychological stress. This finding aligns with previous research on the negative impact of role conflict and moral distress on mental health in other professions,[13,14,15] but it also extends these theories into a completely new context: the interaction between humans and algorithms. In the context of Vietnamese Fintech companies, where the pace of AI adoption is rapid and competitive pressure is extremely high, AI-proposed processes and decisions are often prioritized for efficiency. This places accountants, who bear the responsibility of being ethical "gatekeepers", in a state of powerlessness when they notice irregularities but cannot intervene, leading to a decline in their mental health. The implication of this result is clear: AIEC is not a trivial issue, but a serious occupational hazard in the modern workplace.

Our subsequent finding on the moderating role of work experience further clarifies the nature of the above relationship. The stronger impact of AIEC on the low-experience group (β = 0,51) compared to the high-experience group (β = 0,24) is a result consistent with theoretical expectations regarding professional development and stress-coping abilities. Accountants new to the profession (≤ 3 years) often lack sufficient confidence in their professional judgment, have not accumulated effective coping mechanisms, and may perceive a greater sense of authority from the organization's imposed technology systems. They are more likely to doubt themselves when their judgment contradicts the output of an "intelligent" system. Conversely, more seasoned accountants, with their well-honed professional knowledge and confidence, are better able to "resist". They may recognize the conflict, but their confidence in their own abilities helps them reduce the level of psychological stress. This

result implies that mental health support programs need to pay special attention to young employees, who are at the forefront of interaction with AI and are also the most vulnerable.

One of the most significant findings of the study is that psychological safety does not play a moderating role in reducing the negative impact of AIEC on mental health (β of the interaction term = 0,04; p > 0,05). This result not only contradicts the study's initial expectations but also suggests the need to reconsider a foundational assumption in the literature on management and organizational psychology: that psychological safety is a universal "buffer" for workplace stressors.[16,17]

The reason this protective mechanism has no significant effect against AIEC can be explained by the inherent incompatibility between the mechanism and the source of the threat. Specifically, psychological safety is designed to address interpersonal risk. It creates a safe space where an individual can speak up, admit a mistake, or question a decision in front of other people (colleagues, superiors) without fear of being punished or humiliated. However, AIEC is not a person-to-person conflict. It is a human-system conflict, which is algorithmic, impersonal, and unemotional in nature.

In the Vietnamese Fintech context, an accountant may feel safe enough to tell their manager: "I think the AI system's decision to reject this loan is unfair and potentially discriminatory". The manager, while very supportive and not disciplining the employee, might only be able to reply: "I understand your concern, but that's the system's process, and we can't change it". In this case, psychological safety facilitates the expression of concern, but it cannot resolve the source of the stress - an algorithm with an opaque, rigid, and non-negotiable operating mechanism. This inability to effect real change leaves the moral distress intact, or even intensifies it, as the employee realizes that even in an open environment, they lack the ability to influence the system.

These findings suggest that traditional mental health protection models, designed for a world of human-to-human interaction, may be becoming obsolete in the age of AI. Simply building an open and trusting work environment is not enough to protect employees from new types of technology-induced stress. Organizations need to go further, moving towards building "algorithmic safety" - an environment where employees are not only allowed to speak up but also have clear, effective channels to question, audit, and demand explanations from the algorithmic systems themselves. This requires the development of Explainable AI (XAI) systems, transparent AI governance processes, and grievance mechanisms specifically for AI-driven decisions. Otherwise, organizations will only be creating a superficial form of safety, while their employees silently endure psychological harm from the technological systems they are forced to collaborate with.

Limitations and Future Research Directions

Although the study's objectives have been met, this study has several limitations that should be honestly acknowledged, which in turn open up new avenues for future research.

First, the reliance on self-reported data is inherently susceptible to biases from participants' subjective perceptions. Furthermore, the cross-sectional design captures relationships at a single point in time, which limits the ability to draw causal inferences. To better understand the psychological processes involved, longitudinal studies are needed to track changes in accountants' mental health over time as they interact with and adapt to AI systems.

Second, the study focuses on a very specific context: accountants within Vietnam's Fintech sector. While this provides deep insights, it also raises questions about the generalizability of the findings to other industries, other professional roles (e.g., auditors, financial analysts), or different cultural and regulatory contexts. Therefore, comparative studies across industries and countries, as well as investigations into the impact of specific types of AI algorithms (e.g., explainable AI vs. black-box AI), would be an invaluable direction. Such efforts would help build a more comprehensive theory of psychological protection mechanisms in the digital age.

CONCLUSIONS

This study was conducted to address a pressing psychological challenge emerging in the digital era: the impact of AI-induced ethical conflict (AIEC) on the mental health of accountants. The study's focus was to test this relationship while also re-evaluating the effectiveness of traditional protective mechanisms, such as psychological safety, and the role of work experience within the specific context of Vietnam's Fintech enterprises.

Based on the empirical analysis of 416 accountants, three core conclusions were drawn. First, AI-induced ethical conflict has been proven to be a real and serious occupational risk factor, exerting a statistically significant negative impact on the mental health of accountants. Second, and most notably, psychological safety—a widely recognized protective mechanism—proved ineffective in weakening this negative relationship. Third, work experience emerged as a significant moderating factor, confirming that less experienced accountants are a distinctly more vulnerable group to this new form of stress.

These conclusions carry significant implications, both theoretically and practically.

Theoretically, the study's most significant finding challenges the universality of traditional psychological

protection theories. The results suggest that mechanisms designed for interpersonal risks, such as psychological safety, may no longer be suitable or robust enough to cope with stressors that are algorithmic, systemic, and impersonal in nature. This opens an urgent academic dialogue on the necessity of developing new theoretical frameworks, such as the concept of "algorithmic safety," to protect the workforce in human-machine interaction environments.

Practically, this study sends a strong message to managers and policymakers in the Fintech industry. Merely fostering an open and trusting work environment is insufficient. Organizations must intervene directly at the root of the problem by: (i) Prioritizing the development and implementation of Explainable AI (XAI) systems to reduce the opacity in decision-making processes; (ii) Establishing transparent AI governance processes and effective grievance channels specifically for algorithm-driven decisions; and (iii) Designing targeted mental health support programs, with a special focus on young, less-experienced employees, who are at the forefront of this interaction and are also the most vulnerable.

In summary, this study not only identifies a new type of conflict in the modern workplace but also underscores that safeguarding the mental health of the workforce in the digital age requires a fundamental paradigm shift: from focusing solely on interpersonal safety mechanisms to forging new frameworks for safety in human-algorithm interaction.

ACKNOWLEDGMENTS

This research is funded by University of Finance – Marketing.

BIBLIOGRAPHIC REFERENCES

1. Soyombo OT. Reviewing the role of AI in fraud detection and prevention in financial services. International Journal of Science and Research Archive. 2024;11(1):2101-10. https://doi.org/10.30574/ijsra.2024.11.1.0279

2. Jain N, Patil S. Artificial intelligence models for fraud detection: advancements, challenges, and future prospects. 2024. https://doi.org/10.21428/e90189c8.6d8ab5f6

3. Pham HH, Nguyen TH, Tran MT. Audit quality of financial statements of commercial banks, whether or not there is a difference in audit quality provided by Big4 and Non-Big4 audit firms. International Journal of Economics and Financial Issues. 2025;15(1):159. https://doi.org/10.32479/ijefi.17541

4. Amato S, Broccardo L, Tenucci A. Family firms, management control and digitalization effect. Management Decision. 2024;62(5):1645-67. https://doi.org/10.1108/MD-03-2023-0347

5. Cardinaels E, Künneke J, Sandu I, Széles M. Empowering accounting with artificial intelligence. Maandblad Voor Accountancy en Bedrijfseconomie. 2024;98(3):75-8. https://doi.org/10.5117/mab.98.125065

6. Rizzo JR, House RJ, Lirtzman SI. Role conflict and ambiguity in complex organizations. Administrative Science Quarterly. 1970;15(2):150–63. https://doi.org/10.2307/2391486

7. Jameton A. Nursing practice: The ethical issues. Englewood Cliffs, NJ: Prentice-Hall; 1984.

8. Adadi A, Berrada M. Peeking inside the black-box: a survey on explainable artificial intelligence (XAI). IEEE Access. 2018;6:52138–60. https://doi.org/10.1109/ACCESS.2018.2870052

9. O'Neil C. Weapons of math destruction: how big data increases inequality and threatens democracy. New York: Crown; 2017.

10. Braganza A, Chen W, Canhoto A, Sap S. Productive employment and decent work: the impact of ai adoption on psychological contracts, job engagement and employee trust. Journal of Business Research. 2021;131:485-94. https://doi.org/10.1016/j.jbusres.2020.08.018

11. Hamric AB, Borchers CT, Epstein EG. Development and testing of an instrument to measure moral distress in healthcare professionals. AJOB Primary Research. 2012;3(2):1–9. https://doi.org/10.1080/21507716.2011.652337

12. Yang Q, Lee Y. Ethical ai in financial inclusion: the role of algorithmic fairness on user satisfaction and recommendation. 2024. https://doi.org/10.20944/preprints202407.1655.v1

13. Kalisch BJ, Lee H, Rochman M. Nursing staff teamwork and job satisfaction. Journal of nursing management. 2010;18(8):938-47. https://doi.org/10.1111/j.1365-2834.2010.01153.x

14. Ashall V. Reducing moral stress in veterinary teams? evaluating the use of ethical discussion groups in charity veterinary hospitals. Animals. 2023;13(10):1662. https://doi.org/10.3390/ani13101662

15. Küçükkaya B, Süt H. The relationship between turkish women's self-efficacy for managing work-family conflict and depression, anxiety and stress during the covid-19 pandemic: a web-based cross-sectional study. Work. 2022;73(4):1117-24. https://doi.org/10.3233/wor-220190

16. Edmondson A. Psychological safety and learning behavior in work teams. Administrative Science Quarterly. 1999;44(2):350–83. https://doi.org/10.2307/2666999

17. Zhou W, Zhu Z, Vredenburgh D. Emotional intelligence, psychological safety, and team decision making. Team Performance Management. 2020;26(1/2):123-41. https://doi.org/10.1108/tpm-10-2019-0105

18. Lovibond SH, Lovibond PF. Manual for the Depression Anxiety Stress Scales. 2nd ed. Sydney: Psychology Foundation of Australia; 1995.

19. Brislin RW. Back-translation for cross-cultural research. Journal of Cross-Cultural Psychology. 1970;1(3):185–216. https://doi.org/10.1177/135910457000100301

20. Henry JD, Crawford JR. The short-form version of the Depression Anxiety Stress Scales (DASS-21): construct validity and normative data in a large non-clinical sample. British Journal of Clinical Psychology. 2005;44(2):227–39. https://doi.org/10.1348/014466505X29657

21. Podsakoff PM, MacKenzie SB, Lee J-Y, Podsakoff NP. Common method biases in behavioral research: a critical review of the literature and recommended remedies. Journal of Applied Psychology. 2003;88(5):879–903. https://doi.org/10.1037/0021-9010.88.5.879

22. Kock N. Common method bias in PLS-SEM: a full collinearity assessment approach. International Journal of e-Collaboration. 2015;11(4):1–10. https://doi.org/10.4018/ijec.2015100101

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest.

AUTHORSHIP CONTRIBUTION

Conceptualization: Hong Van Tran.
Data curation: Hong Van Tran.
Formal analysis: Hong Van Tran.
Research: Hong Van Tran.
Methodology: Hong Van Tran.
Project management: Hong Van Tran.
Resources: Hong Van Tran.
Software: Hong Van Tran.
Supervision: Hong Van Tran.
Validation: Hong Van Tran.
Display: Hong Van Tran.
Drafting - original draft: Hong Van Tran.
Writing - proofreading and editing: Hong Van Tran.