An Empirical Study of Microsoft Copilot Adoption in Financial Technology: Productivity, Performance, and Trust Perspectives

DOI
https://doi.org/10.18690/um.epf.7.2026.16

ISBN
978-961-299-166-1

Volkan Batmaca, Ceren Cubukcu Cerasi
Gebze Technical University, Faculty of Business Administration, Kocaeli/Gebze, Turkey
v.batmaca2024@gtu.edu.tr, cerencubukcu@gtu.edu.tr

This chapter presents an empirical study examining the determinants of Microsoft Copilot usage among software development professionals in the FinTech sector. Drawing on technology acceptance and socio-technical perspectives, the study investigates how perceived productivity, performance enhancement, learning and development, trust, and sectoral impact influence actual usage behavior. Data were collected from 154 developers through a structured survey and analyzed using non-parametric statistical methods. The findings indicate that usage frequency and application domain significantly shape general usage and learning perceptions, while gender differences emerge in the trust dimension; education level and professional experience show no statistically significant effects. Overall, the results highlight that contextual engagement, and practical performance benefits play a more decisive role in AI-assisted tool adoption than demographic characteristics, offering implications for structured and domain-specific implementation strategies in regulated software environments.

Keywords:
copilot,
sectoral impact,
productivity,
performance,
learning and development,
trust and criticism.

University of Maribor Press

1 Introduction

The rapid development of artificial intelligence technologies has significantly influenced the field of software engineering. AI-supported coding assistants have become increasingly integrated into development environments, offering real-time suggestions, automated code generation, debugging support, and optimization recommendations. Among these tools, Microsoft Copilot has emerged as one of the most widely adopted AI-based assistants designed to enhance software development processes.

Unlike traditional development tools, AI-powered assistants do not merely provide static functionalities; instead, they actively interact with developers, support decision-making processes, and contribute to productivity and code quality improvements. As organizations strive to accelerate development cycles and improve efficiency, AI-supported tools are increasingly positioned as strategic assets within software teams. However, the actual usage of such tools is not determined solely by their technical capabilities. Developers' perceptions regarding productivity, performance improvement, learning opportunities, trust, and broader sectoral impact play a crucial role in shaping adoption behavior.

While existing research largely focuses on the technical performance of AI-based systems, fewer studies examine the behavioral and perceptual factors that influence developers' actual usage of AI coding assistants. Understanding these determinants is essential for both academic research and practical implementation strategies, as effective adoption depends not only on availability but also on user acceptance and perceived value.

In this context, the present study aims to identify the key factors influencing Microsoft Copilot usage among software development professionals. Specifically, the study examines the effects of perceived productivity, performance enhancement, learning and development, trust and criticism, and sectoral impact on actual Copilot usage behavior. By empirically analyzing these relationships, this research seeks to contribute to the growing literature on AI-assisted software development and to provide insights into the behavioral dynamics underlying AI tool adoption.

2 Literature Review

AI-assisted software development tools, particularly GitHub Copilot, have significantly transformed contemporary software engineering practices. These systems are built upon large-scale transformer architectures (Vaswani et al., 2017) and large language models capable of few-shot learning (Brown et al., 2020). Leveraging OpenAI Codex (OpenAI, 2021; Chen et al., 2021), Copilot generates context-aware code suggestions based on natural language prompts, representing a shift from rule-based automation to probabilistic generative assistance embedded within development environments.

From a theoretical perspective, the adoption of such tools can be explained through the Technology Acceptance Model (Davis, 1989) and the Unified Theory of Acceptance and Use of Technology (Venkatesh et al., 2003), where perceived usefulness and ease of use are key determinants of behavioral intention. In high-pressure and regulated domains such as FinTech, these factors become even more critical.

Empirical studies consistently demonstrate that Copilot enhances developer productivity. Peng et al. (2023) report that developers complete tasks significantly faster when using AI assistance, while industry findings (GitHub Blog, 2024) confirm improvements in development speed and workflow efficiency. These findings align with diffusion theory (Rogers, 2003), suggesting that repeated exposure reinforces sustained adoption.

However, productivity gains are accompanied by important concerns. Research shows that AI-generated code may contain security vulnerabilities (Pearce et al., 2022; Fu et al., 2023), which is particularly critical in security-sensitive domains such as financial technologies. Trust in automation therefore becomes a central issue. Lee and See (2004) emphasize that inappropriate reliance may lead to overtrust, a risk also observed in AI coding tools (Vaithilingam et al., 2022).

Beyond productivity and trust, AI-assisted tools influence learning processes. While they can accelerate onboarding and exposure to new technologies (Chen et al., 2021), excessive reliance may reduce deep understanding and critical thinking (Vaithilingam et al., 2022). This creates a dual effect, highlighting the need for balanced human–AI interaction.

In the context of FinTech, the adoption of AI tools is further shaped by regulatory and ethical considerations. Emerging frameworks such as the EU AI Act emphasize transparency, accountability, and risk management (Veale & Zuiderveen Borgesius, 2021). Given the high sensitivity of financial systems, ensuring trustworthy and secure AI integration is essential.

Despite growing research on productivity and security, there is a lack of survey-based empirical studies examining multidimensional user perceptions–including productivity, performance, learning, and trust–within FinTech contexts. This study addresses this gap by providing an integrated empirical analysis of Copilot usage, contributing to both technology acceptance literature and AI governance discussions.

3 Method

This study employed quantitative research design using a survey method to examine the determinants of Microsoft Copilot usage among software development professionals. Data were collected between November 2025 and February 2026 through online questionnaires distributed via digital platforms. The survey was administered using Google Forms and shared with professionals actively involved in software development, particularly those familiar with AI-assisted coding tools.

A total of 154 responses were obtained and deemed suitable for statistical analysis. Participation was voluntary, and respondents were informed about the purpose of the study. The data collected were analyzed using IBM SPSS Statistics software.

3.1 Research Design and Hypotheses

Based on literature and theoretical background, this study proposes that perceived benefits and trust-related factors influence Copilot usage behavior.

The research model assumes that Copilot Usage is the dependent variable, while general usage habit, productivity, performance, learning and development, trust, and sectoral impact are independent variables.

In the measurement tool, 5-point Likert statements are scored as "I strongly disagree (1)", "I disagree (2)", "I am undecided (3)", "I agree (4)", "I strongly agree (5)". The scale does not contain any reverse-scored statements. In our study, the Cronbach Alpha internal consistency coefficient was calculated as 0.962.

Accordingly, the following hypotheses were formulated:

H1: Gender does not have a statistically significant effect on Copilot-supported benefit usage dimensions in the FinTech sector.

H2: The frequency of Copilot usage has a statistically significant and positive effect on Copilot-supported benefit usage across multiple dimensions in the FinTech sector.

H3: The area of Copilot usage (profession) has a statistically significant effect on selected dimensions (general usage, sectoral impact, productivity, and performance) of benefit usage in the FinTech sector.

H4: Education level has a statistically significant effect on selected dimensions (general usage and performance) of Copilot-supported benefit usage in the FinTech sector.

H5: Experience has a limited and dimension-specific effect on Copilot-supported benefit usage, primarily influencing the general usage dimension in the FinTech sector.

4 Results and Discussion

Data were analyzed using SPSS 31.0.2.0 with a 95% confidence interval and a significance level of $p < 0.05$. Normality was assessed using Shapiro-Wilk and Kolmogorov-Smirnov tests, and non-parametric methods were applied.

The Mann–Whitney U test results indicate that there are no statistically significant differences between male and female participants across any of the sub-dimensions ($p > 0.05$ for all variables). This suggests that gender does not play a determining role in the usage, perception, or evaluation of AI-assisted development tools.

Although female participants exhibit slightly higher mean ranks in most dimensions–particularly in general usage, sectoral impact, and performance–these differences are not statistically significant. Therefore, they should be interpreted as marginal tendencies rather than meaningful disparities.

The findings imply that AI-assisted coding tools provide a relatively gender-neutral experience, where both male and female users benefit similarly in terms of productivity, learning, and performance. This is an important result in the context of technology adoption, as it suggests that such tools do not reinforce gender-based differences in software development environments.

Overall, the results support the conclusion that gender is not a significant factor influencing either the adoption or perceived effectiveness of AI-supported development tools, highlighting their inclusive and broadly accessible nature.

Table 1: Evaluation of Technology Acceptance and Its Subdimensions According to Gender

Sub-dimension	Gender	N	Mean Rank	U	Z	P*
General Usage	Male	96	74.57	2503.000	-1.054	0.292
	Female	58	82.34			
Sectoral Impact	Male	96	74.12	2459.500	-1.218	0.223
	Female	58	83.09			
Productivity & Efficiency	Male	96	77.25	2760.000	-0.091	0.928
	Female	58	77.91			
Performance	Male	96	73.64	2413.000	-1.399	0.162
	Female	58	83.90			
Learning & Development	Male	96	75.40	2582.500	-0.764	0.445
	Female	58	80.97			
Trust & Critical Perspective	Male	96	76.01	2641.000	-0.538	0.590
	Female	58	79.97			

*Mann–Whitney U Test. *p <.05 indicates statistical significance.

The findings presented in Table 2 reveal that usage frequency has a statistically significant effect on multiple dimensions of Copilot-supported benefit utilization. The Mann–Whitney U test results indicate significant differences between participants who use the system every day and those who have never used it in the dimensions of general usage (U = 98.000, p = .022), productivity and efficiency (U = 101.500, p = .025), performance (U = 109.500, p = .037), learning and development (U = 103.500, p = .027), and trust and critical approach (U = 101.500, p = .026).

In all these dimensions, participants who reported daily usage demonstrated substantially higher mean rank scores compared to those who reported never using the system, indicating more positive evaluations regarding system effectiveness, usability, and contribution to professional tasks. This consistent pattern suggests that frequent interaction with Copilot enhances users' perceptions of its overall value and functional benefits.

Table 2: Evaluation of Technology Acceptance and Its Subdimensions According to General Usage

Sub-dimension	Frequency Of Use	N	Mean Rank	U	Z	P*
General Usage	Every Day	74	42.18	98.000	-2.295	0.022
	Never Used	6	19.83			
Sectoral Impact	Every Day	74	41.92	117.000	-1.935	0.053
	Never Used	6	23.00			
Productivity & Efficiency	Every Day	74	42.13	101.500	-2.243	0.025
	Never Used	6	20.42			
Performance	Every Day	74	42.02	109.500	-2.088	0.037
	Never Used	6	21.75			
Learning & Development	Every Day	74	42.10	103.500	-2.213	0.027
	Never Used	6	20.75			
Trust & Critical Perspective	Every Day	74	42.13	101.500	-2.231	0.026
	Never Used	6	20.42			

*Mann–Whitney U Test. *p <.05 indicates statistical significance.

Overall, these findings demonstrate that usage frequency is a critical determinant of perceived benefit and system evaluation. Regular use appears to strengthen user experience, improve perceived productivity and performance outcomes, and enhance learning and trust-related perceptions. These results support the argument that continuous exposure and active engagement with AI-assisted tools play a key role in maximizing their perceived effectiveness and adoption in the FinTech context.

The results presented in Table 3 indicate that profession plays a statistically significant role in shaping several dimensions of technology adoption. Specifically, statistically significant differences were observed in general usage ($p < .001$), sectoral impact ($p = .038$), productivity and efficiency ($p = .005$), and performance ($p = .045$) between software developers/engineers and participants from other professional backgrounds.

Across these dimensions, software developers consistently demonstrated higher mean rank scores compared to other professionals. This finding suggests that individuals with a technical background tend to perceive AI-assisted tools such as Copilot as more beneficial, effective, and performance-enhancing. Their familiarity with software development processes, coding environments, and digital tools likely contributes to stronger perceived usefulness and more efficient system utilization.

In contrast, no statistically significant differences were found in the learning and development ($p = .119$) and trust and critical perspective ($p = .066$) dimensions. Although developers again showed higher mean ranks, these differences did not reach statistical significance. This suggests that perceptions related to learning outcomes and trust in AI systems may be influenced by broader cognitive or organizational factors rather than profession alone.

Overall, the findings indicate that professional background, particularly being a software developer, constitutes an important determinant of perceived system utility, productivity gains, and performance outcomes. However, its influence appears to be more limited in shaping trust-related perceptions and learning experiences. These results highlight the importance of domain-specific familiarity and technical competence in maximizing the perceived benefits of AI-assisted development tools.

The findings presented in Table 4 indicate that educational level has a statistically significant effect on several dimensions of technology adoption. Specifically, general usage, sectoral impact, productivity and efficiency, and performance show statistically significant differences across education groups ($p < .001$ for all dimensions). These results suggest that education level is an important determinant of how individuals perceive and utilize AI-assisted technologies

Table 3: Evaluation of Technology Acceptance and Its Subdimensions According to Occupational Group

Sub-dimension	Profession	N	Mean Rank	U	Z	P*
General Usage	Developer / Engineer	74	44.62	190.000	-3.491	< .001
	Other	6	21.07			
Sectoral Impact	Developer / Engineer	74	42.96	299.500	-2.071	.038
	Other	6	28.89			
Productivity & Efficiency	Developer / Engineer	74	43.80	244.500	-2.792	.005
	Other	6	24.96			
Performance	Developer / Engineer	74	42.86	306.000	-2.007	.045
	Other	6	29.36			
Learning & Development	Developer / Engineer	74	42.34	340.500	-1.558	.119
	Other	6	31.82			
Trust & Critical Perspective	Developer / Engineer	74	42.67	318.500	-1.839	.066
	Other	6	30.25			

*Kruskal–Wallis H Test. *p < .05 indicates statistical significance.

A clear pattern emerges in which university graduates consistently demonstrate the highest mean rank scores, followed by master's degree holders, while high school and PhD participants report comparatively lower scores in most dimensions. This trend indicates that individuals with a university-level education tend to perceive greater benefits, higher efficiency, and stronger performance outcomes when using AI-based tools such as Copilot.

Interestingly, the results do not reveal statistically significant differences in the learning and development ($p = .052$) and trust and critical perspective ($p = .096$) dimensions, although both are close to the significance threshold. This suggests that while education level strongly influences practical and performance-related perceptions, its impact on learning outcomes and trust-related evaluations is more limited.

Furthermore, the Jonckheere–Terpstra test results ($p < .05$) indicate a significant ordered trend across education levels for most dimensions, reinforcing the presence of a systematic relationship between education and technology perception. However, the relatively lower scores observed among PhD participants may reflect differences in expectations, critical evaluation tendencies, or lower reliance on such tools for routine tasks.

Overall, these findings suggest that educational level plays a crucial role in shaping perceptions of usefulness, efficiency, and performance, while its effect on trust and learning remains comparatively weaker. This highlights the importance of considering both cognitive and contextual factors when evaluating technology adoption across different educational backgrounds.

Table 4: Evaluation of Technology Acceptance and Its Subdimensions According to Educational Level

Sub-dimension	Education	N	Mean Rank	χ^2	P*
General Usage	University	91	91.47	24.755	< .001
	Master's Degree	52	61.50		
	High School	7	38.79		
	PhD	4	35.38		
Sectoral Impact	University	91	89.79	19.470	< .001
	Master's Degree	52	62.79		
	High School	7	55.43		
	PhD	4	27.88		
Productivity & Efficiency	University	91	88.08	16.756	< .001
	Master's Degree	52	67.05		
	High School	7	45.14		
	PhD	4	29.25		
Performance	University	91	89.77	21.250	< .001
	Master's Degree	52	64.88		
	High School	7	37.64		
	PhD	4	32.13		
Learning & Development	University	91	84.72	7.747	.052
	Master's Degree	52	68.79		
	High School	7	70.21		
	PhD	4	39.25		
Trust & Critical Perspective	University	91	84.78	6.339	.096
	Master's Degree	52	67.40		
	High School	7	59.71		
	PhD	4	74.25		

*p <0.005 indicates statistical significance. Kruskal–Wallis Test

Table 5 presents the results of H5. The Kruskal–Wallis test results indicate that professional experience has a statistically significant effect only on the "General Usage" dimension ($\chi^2 = 10.163$, $p = 0.038$). Specifically, individuals with 2–5 years of experience exhibit the highest mean rank, suggesting that mid-level professionals tend to use AI-assisted tools more actively compared to both less experienced and highly experienced individuals.

Table 5: Evaluation of Technology Adoption and Its Sub-Dimensions by Groups Based on Professional Experience

Sub-dimension	Experience	N	Mean Rank	χ^2	P*
General Usage	0–1 years	7	63.43	10.163	0.038
	2–5 years	49	88.68		
	6–10 years	39	78.97		
	11–15 years	41	76.63		
	16+ years	18	51.31		
Sectoral Impact	0–1 years	7	71.21	2.634	0.621
	2–5 years	49	81.74		
	6–10 years	39	75.96		
	11–15 years	41	81.01		
	16+ years	18	63.72		
Productivity & Efficiency	0–1 years	7	59.57	6.019	0.198
	2–5 years	49	88.98		
	6–10 years	39	76.92		
	11–15 years	41	71.06		
	16+ years	18	69.14		
Performance	0–1 years	7	64.29	7.336	0.119
	2–5 years	49	86.66		
	6–10 years	39	82.76		
	11–15 years	41	72.30		
	16+ years	18	58.14		
Learning & Development	0–1 years	7	61.36	3.849	0.427
	2–5 years	49	81.60		
	6–10 years	39	81.99		
	11–15 years	41	77.61		
	16+ years	18	62.64		
Trust & Critical Perspective	0–1 years	7	74.93	1.734	0.784
	2–5 years	49	79.67		
	6–10 years	39	75.22		
	11–15 years	41	82.16		
	16+ years	18	66.92		

*p <0. 005 indicates statistical significance. Kruskal–Wallis Test.

For all other dimensions, including sectoral impact, productivity, performance, learning, and trust–no statistically significant differences were observed ($p > 0.05$). This suggests that perceptions regarding the broader impacts of AI-assisted development tools are relatively consistent across different experience levels.

Interestingly, the results reveal a non-linear pattern: while early-career professionals (0–1 years) and highly experienced individuals (16+ years) demonstrate lower mean ranks, mid-career professionals (2–10 years) show higher engagement and perceived benefits. This may indicate that mid-level developers are more adaptable to new technologies and more actively integrate AI tools into their workflows.

Overall, the findings suggest that experience level influences actual usage behavior more than perception, highlighting the importance of targeting different experience groups with tailored adoption strategies in organizations.

5 Conclusion

This study aimed to investigate the impact of demographic and behavioral factors on Copilot-supported benefit usage in the FinTech sector. Based on the empirical findings, revised hypotheses were developed to better align with the observed statistical evidence.

The results indicate that **gender does not have a statistically significant effect** on any of the examined dimensions. This finding supports H1 and suggests that perceptions of Copilot-supported benefits are consistent across genders, indicating a gender-neutral adoption and evaluation pattern.

In contrast, **frequency of Copilot usage emerges as the most influential factor**. Statistically significant and positive differences were observed across multiple dimensions, including general usage, productivity, performance, learning and development, and trust and critical approach. These findings strongly support H2 and highlight that frequent interaction with AI tools enhances perceived benefits and effectiveness.

The analysis of **profession (area of usage)** reveals statistically significant differences in general usage, sectoral impact, productivity, and performance, while no significant differences were found in learning and development or trust-related dimensions. Therefore, H3 is supported, demonstrating that professional roles shape how users perceive the practical and operational benefits of Copilot.

Regarding **education level**, statistically significant differences were identified in general usage and performance dimensions, whereas other dimensions did not show significant variation. Thus, H4 is supported, indicating that education influences certain functional aspects of Copilot usage but does not uniformly affect all dimensions.

Finally, **experience has a limited and selective impact**. A statistically significant effect was observed only in the general usage dimension, with no meaningful differences across other dimensions. These findings support H5 and suggest that experience alone is not a comprehensive predictor of Copilot-related benefit perception.

Overall, the study demonstrates that **behavioral engagement (usage frequency)** is a more critical determinant of perceived benefits than **demographic characteristics (gender, education, experience)**. This underscores the importance of promoting active and sustained use of AI-supported tools in the FinTech sector to maximize their value and impact.

References

Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., ... & Amodei, D. (2020). Language models are few-shot learners. Advances in Neural Information Processing Systems, 33, 1877–1901.

Chen, M., Tworek, J., Jun, H., Yuan, Q., de Oliveira Pinto, H. P., Kaplan, J., ... & Zaremba, W. (2021). Evaluating large language models trained on code https://doi.org/10.48550/arXiv.2107.03374

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Quarterly, 13(3), 319–340

Deloitte. (2023). Generative AI in financial services. Deloitte Insights.

Fu, Y., Liang, P., Tahir, A., Li, Z., Shahin, M., Yu, J., & Chen, J. (2023). Security Weaknesses of Copilot-Generated Code in GitHub Projects: An Empirical Study. https://arxiv.org/abs/2310.02059

GitHub Blog. (2024). Does GitHub Copilot improve code quality? Here's what the data says. https://github.blog/news-insights/research/does-github-copilot-improve-code-quality-heres-what-the-data-says/

Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. Human Factors, 46(1), 50–80.

OpenAI. (2021). Introducing OpenAI Codex. https://openai.com/blog/openai-codex/

Peng, S., Kalliamvakou, E., Cihon, P., & Demirer, M. (2023). The impact of AI on developer productivity. arXiv preprint arXiv:2302.06590.

Rogers, E. M. (2003). Diffusion of innovations (5th ed.). Free Press.

Tom B. Brown et al. (2020). Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., ... & Amodei, D. (2020). Language models are few-shot learners. Advances in Neural Information Processing Systems, 33, 1877–1901.

Vaithilingam, P., Zhang, T., & Glassman, E. L. (2022). Expectation vs. experience: Evaluating the usability of code generation tools. CHI Conference on Human Factors in Computing Systems. https://doi.org/10.1145/3491102.3517506

Vaithilingam, P., Zhang, T., & Guo, P. J. (2022).Is GitHub Copilot a substitute for human pair programming? Proceedings of the 44th International Conference on Software Engineering (ICSE). https://doi.org/10.1145/3510003.3510180

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., & Polosukhin, I. (2017). Attention is all you need. Advances in Neural Information Processing Systems, 30, 5998–6008.

Veale, M., & Zuiderveen Borgesius, F. (2021). Demystifying the draft EU Artificial Intelligence Act. Computer Law Review International, 22(4), 97–112.

Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. MIS Quarterly, 27(3), 425–478.