American Journal of Information Science and Technology
2026, Vol. 10, No. 2, pp. 66–74
https://doi.org/10.11648/j.ajist.20261002.11

Research Article

# Security and Privacy Implications of Microsoft 365 Copilot and GenAI Integration in Enterprise Environments

**Pullaiah Chowdary Vutla**[1, *], **Triveni Yenugu**[2]

[1]Department of Information Technology, University of the Cumberlands, Williamsburg, The United States
[2]Computer Science and Engineering, Jawaharlal Nehru Technological University, Hyderabad, India

## Abstract

The introduction of Generative Artificial Intelligence (GenAI) into the world of enterprise Software-as-a-Service (SaaS) is a crucial development in digital transformation. As one of the highest-performing applications of large language models (LLMs) to productivity software, Microsoft 365 Copilot is a product that promises productivity-level efficiency to increase but at the same time presents a difficult compound pose on the security and privacy issue. The paper has used a secondary qualitative study to review the threat climate of Copilot, privacy-leakage threats, and governance. The study uncovers several vectors of concern by combining technical documentation evidence, regulatory frameworks, and previous scholarly research that the focus of AI-based prompt injection, unintentional data synthesis, and accidental cross-tenant exposure. Findings indicate that Microsoft Purview, Data Loss Prevention (DLP), and Role-Based Access Control (RBAC) offer basic protection but are not very effective against moving GenAI workflows that constantly consume and recreate sensitive enterprise data. As can be seen, a comparison with Google Workspace Duet AI and Salesforce Einstein shows that the risks are not specific to any platform but are common to all SaaS environments that are augmented by AI. The article summarizes that AI-conscious and holistic governance strategy, consisting of technical, organizational, and regulatory controls, is the key to reducing the vulnerability of the enterprise level, as well as maintaining the productivity gains of GenAI integration.

## Keywords

Enterprise LLM Security, Microsoft 365 Copilot, AI Governance, Data Leakage Prevention, GenAI in SaaS, Prompt Injection in Cloud AI, Regulatory Compliance, Information Barriers

## 1. Introduction

The introduction of Generative Artificial Intelligence (GenAI) to the enterprise setting is an eventual change in the way the organizations approach knowledge management, communication, and decision-making [1]. As a flagship application of GenAI integration within Software-as-a-Service (SaaS) platforms, Microsoft 365 Copilot is a service that relies on large language models (LLMs) to support users in writing emails, creating documents, summarizing meeting notes, and automating workflows in programs like Word, Excel, Teams, Outlook, and its Security (Intrusion Detection with Secure Cloud Storage Enabled) [2]. Although such capabilities could have a positive impact on their productivity, new security and

*Correspondence: Pullaiah Chowdary Vutla (vchowdary256@gmail.com)

**Received:** 14 November 2025; **Accepted:** 8 June 2026; **Published:** 26 June 2026

privacy issues have become much more critical than before. Contrary to the old-fashioned enterprise applications, Copilot has deep access to sensitive company information, such as emails, internal files, chat logs, and calendar entries, which is a possible source of data leaks, unauthorized access, or insider attacks.

The fast-rate of Copilot and other AI-based assistants' adoption is predetermined by several factors. To begin with, businesses desire to enhance their productivity by decreasing the level of manual cognitive processes and automating content generation. Second, AI implementation into the current SaaS providers will allow adopting it without additional platforms or tools [3]. This convenience is however at the cost of more attack surfaces. The fact that AI can process, synthesize, and store the information about the organization databases provokes acute concerns regarding confidentiality, integrity, and adherence to ethics standards in such regulations as GDPR, HIPAA, and CCPA [4]. In comparison to traditional software applications, LLMs are constantly learning based on interaction and may accidentally reveal sensitive or regulated information unless well controlled.

The academic literature on the security posture and privacy impact of enterprise AI tools such as Copilot is very limited even though there is an increasing trend towards the adoption of AI in enterprise workflows. The literature on the topic is mostly dedicated to the overall potential of LLMs, and little is said about the real-world threats that may arise due to the deep integration into corporate ecosystem, and Cloud Security Framework Using Deep Feature Learning [5]. Companies that implement Copilot must grapple with traditional cybersecurity hazards, besides AI-related risks that comprise immediate injection assaults, model interference, as well as information leakage due to the outputs generated. A malicious user can generate inputs such that copilot recognizes sensitive data of a different team or tenant and circumvents standard access controls [6].

The purpose of this paper is to fill this research gap and explore the security and privacy of Microsoft 365 Copilot used by enterprises. In particular, the research performs threat modeling to determine the availability of attack vectors, risks of privacy leaks in the context of AI-assisted workflows, and the usefulness of such governance mechanisms as Microsoft Purview, data loss prevention (DLP) policies, and role-based access control (RBAC). In addition, the study can be assessed from a comparative perspective where the security framework of Copilot is compared to other well-regarded, AI-enabled SaaS systems (e.g. Google Workspace Duet AI and Salesforce Einstein).

A thorough review of these areas can provide organizations with potential practical recommendations for using GenAI in a way that is both secure and compliant with regulations. Ultimately, the use of GenAI with enterprise SaaS tools is a double-edged sword: it offers the possibility of increased productivity and new security risks associated with information security and privacy [7]. It is imperative that organizations that are thinking about using Microsoft 365 Copilot know about these ramifications and the means to ensure the safe and responsible use of GenAI. The current paper resides at the crossroads of AI innovation, enterprise security and privacy, and regulatory compliance by combining an analysis of the theoretical aspects of AI and the practical implications of how risk can be mitigated in AI enabled workflows.

## 2. Methodology

The study proposed uses secondary qualitative research to systematically study the security and privacy implications of Microsoft 365 Copilot in the enterprise context. Since Generative AI on the SaaS platform is relatively new and there is little primary empirical data or data collection, secondary qualitative research allows for a more thorough review of the literature, technical documents, security whitepapers, rules and regulations, and enterprise implementation documents. This will be a comprehensive method that gives the study depth as well as breadth and the worth of searching the field and verifying the pieces of information which are factual and reliable.

It was taken in the form of thorough threat modeling of how Copilot would integrate into Microsoft 365. Based on publicly available documentation, security warnings, and scholarly investigations of enterprise LLMs, possible attack vectors were outlined, sorted, as well as evaluated. Special consideration was given to cases in which Copilot has access to emails, files, Teams chat messages and calendar events, which is deep enough to be used. The threats of spoofing, tampering, repudiation, information disclosure, denial of service, and privilege escalation were examined with the help of the STRIDE framework.

$$\text{Threat Severity } (Rt) = i=1 \sum n \, (Li \times Ii) \quad (1)$$

Where:

1) $R_t$= total risk score
2) $L_i$= likelihood of threat *i*
3) $I_i$= impact of threat *i*

*Equation (1) formalizes the overall risk evaluation using the STRIDE framework, where each threat category's likelihood and impact are multiplied to derive a composite score.*

Secondary sources offered some information on recognized vulnerabilities, attack techniques peculiar to AI, including the immediate injection, and when infected by malicious prompts, sensitive company information can be impacted. The second stage was devoted to the privacy risk assessment that relied on the case studies, compliance reports, and published reviews of data loss prevention (DLP) policies and Microsoft Purview. The analysis factored the process of Copilot about personal identifiable information (PII), protected health information (PHI) and corporate intellectual property. Through qualitative synthesis, the scenarios were created to check the possibility of unintentional data leakage, the type of information that can be compromised, and the consequences that might arise in regulatory compliance in the framework of the GDPR, HIPAA,

or CCPA. The secondary qualitative evidence allowed defining the weaknesses on the systemic level of the enterprise AI workflows without the need to conduct any experiments or gain access to proprietary data.

The paper also explores access control and governance systems according to the published technical documentation, white papers, and analysts of Microsoft. The use of role-based access control (RBAC), information barriers, logging, and audit capabilities, were evaluated qualitatively to understand their effectiveness in ensuring the privilege of least privilege is upheld and in addressing the risks that may arise because of the data exposure caused by AI. Special emphasis was made on the multi-tenant setting, in which data leakage across organizations may take place, and it is difficult to establish governance policies of the complex enterprise deployment.

Lastly, there was a comparative qualitative analysis to place Microsoft 365 Copilot in the context of AI-based enterprise tools. Google Workspace Duet AI and Salesforce Einstein were examined with the help of publicly available security documentation and technical white papers and the corresponding academic research. The comparison was done by considering architecture, privacy controls, access restrictions, and AI-related weaknesses. The use of synthesized secondary qualitative data in the comparison results in the identification of common and platform-specific strengths to provide practical advice for organizations contemplating the adoption of GenAI [8].

All the findings were organized into thematic conclusions, gaining coherence. Visuals and summaries were used to convey the conclusions gained from secondary sources, all organized in conceptual diagrams, tables or flowcharts. The systematic secondary qualitative methodologies result in use of a logs of viable evidence based on the evaluation of Copilot's security and privacy environment for organizations to assess potential adoption of GenAI while protecting personal information and maintaining compliance.

## 3. Results

Synthesized secondary qualitative data was used to identify common trends and distinctive capabilities by platform to provide actionable recommendations to organizations thinking about implementation of GenAI.

Findings were also synthesized into coherent thematic conclusions accompanied by conceptual diagrams, tables, and flowcharts constructed using secondary sources. While using systematic secondary qualitative research, this study provides a useful and credible evaluative assessment of Copilot's security and privacy landscape to help organizations who wish to adopt GenAI, compliance with protection.

### 3.1. Threat Exposure and Attack Surface Expansion

From our analysis, we see that Microsoft 365 Copilot significantly expands the enterprise attack surface due to its significantly broad access rights and agent role across key productivity applications such as Outlook, Word, Teams, and SharePoint [9]. Traditional enterprise applications (or official business software applications) operate under widely understood boundaries on their data access; on the flip side, Copilot is powered by large language models (LLM) to pull data from multiple applications to generate enriched context-based outputs [10]. The Alt items, (that agent role) increase overall benefit for the user experience, but expose operations to new, multi-vector attacks.

$$Attack\ \ Surface = \alpha D + \beta C + \gamma P \quad (2)$$

Where:

1) $D$= number of data sources Copilot accesses
2) $C$= number of connected applications
3) $P$= number of users with privileged access
4) $\alpha, \beta, \gamma$= weight coefficients

*Equation (2) captures the expansion of the attack surface as a function of data sources, interconnections, and privileged users. As ASincreases, so does the likelihood of prompt injection and data leakage events.*

Secondary sources underscore prompt injection and contextual contamination as the foremost threats to consider. In a prompt injection event, an adversary could insert harmful text or hidden instructions within documents, chat messages, or emails that target Copilot's behavior [11]. Due to the probabilistic function of LLMs application of user inputs, it can accidentally commit harmful queries, or unauthorized queries that could access or summarize sensitive information. Contextual contamination, where prior information from one user's session affects another user's prompt and output, ultimately poses risks in shared or multi-tenant environments [12]. Assessments indicate that even with tenant isolation, hidden memory and caching mechanisms in the model may reveal data within future interactions, purposively or unintentionally.

Furthermore, although Copilot depends on user authentication through Azure Active Directory, it remains vulnerable to insider threats. Employees authenticated to a supported identity can use Copilot to extract and synthesize sensitive information from a variety of resources [13]. Qualitative evidence implies that these threats are greater in organizations that do not have enhanced monitoring and/or prompt-level logging in place to capture sensitive information disclosed with no logging in the output of interest. Taken together, these threats suggest that Copilot's intelligence, while valuable, could also function as a threat intelligence tool against the organization producing it.

### 3.2. Privacy Leakage (PL) and Regulatory Concerns

Another important theme that arose from the content analysis was privacy exposure and its influence on compliance

risk. Microsoft 365 Copilot engages with regulated data "categories" such as personally identifiable information (PII), protected health information (PHI), and proprietary corporate content [14]. While Microsoft has affirmed its commitment to data residency, data encryption, and data security, qualitative evidence from compliance documents and privacy case studies indicates that the utilization of GenAI has also shifted the landscape of complying with privacy regulatory obligations such as the General Data Protection Regulation (GDPR), the Health Insurance Portability and Accountability Act (HIPAA), and California Consumer Privacy Act (CCPA) [15].

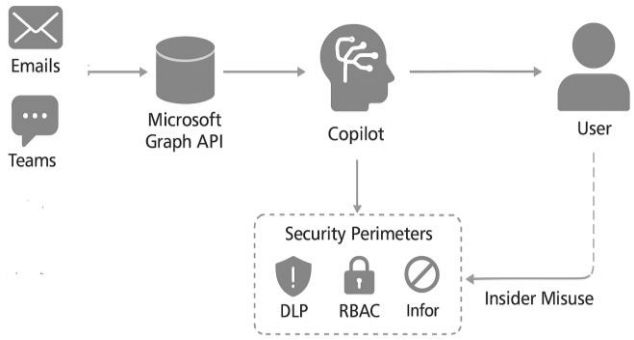


***Figure 1.*** *Threat model showing Copilot's data flow and potential exposure points.*

The research identifies different ways that privacy leaks can happen. The first, inadvertent synthesis, is when Copilot combines pieces of information from several documents together to create a response that seems benign, but that nonetheless contains sensitive data [16]. The second, contextual overreach, is when Copilot uses or cites the file that a user intended to provide, but the incorporation of other files is outside of what the user intended to provide and is possible because of insufficiently bounded permissions, or vague phrasing of the user prompt [17]. The last, shadow data persistence, is the temporary caching of the data for context enrichment that leads to ongoing privacy exposure risk if that data is not erased promptly [18].

$$\text{PL Probability} = 1 - e^{-(\lambda \times N)} \quad (3)$$

Where:

1) $\lambda$= leakage rate per interaction

2) $N$= number of Copilot user prompts or sessions

*As shown in Equation (3), the probability of privacy leakage increases exponentially with the number of AI interactions, emphasizing the need for prompt-level monitoring and governance.*

Microsoft Purview and Data Loss Prevention (DLP) policies alleviate portions of risk by implementing data classification and retention controls; however, qualitative observations suggest Purview operate at a structural level and Copilot operates in context [19]. Therefore, there can be offense gaps where Purview classifies data statically and Copilot's generative capabilities can dynamically reinterpret and rephrase content [20]. The discrepancies highlight a key tension. Privacy controls designed for deterministic systems don't transfer easily to probabilistic AI systems. This requires firms to broaden their DLP policies to cover both monitoring AI behavioral and validating the output.

## 3.3. Governance Mechanisms and Organizational Readiness

One of the main conclusions of this research is that the effectiveness of governance is dependent on the own organizational maturity concerning AI risk governance. The secondary qualitative data informed us that Microsoft 365 Copilot does contain some governance features, such as Role-Based Access Control (RBAC), Information Barriers, and compliance auditing. However, the efficiency of these features is dependent on enterprise configuration and enterprise diligence [21]. MisconFigured RBAC policies are one of the frequent problems cited in enterprise adoption, as these policies tend to provide Copilot with too much access across repositories [22]. This research suggests that if such a misconfiguration occurs, it is inadvertently violating the principal of least privilege, and Copilot can pull information that may be compartmentalized.

$$\text{Access Control Effectiveness (ACE)} = \frac{A_c}{A_t} \quad (4)$$

Where:

1) $A_c$= correctly conFigured access roles

2) $A_t$= total access roles

*Equation (4) quantifies governance accuracy, where lower ACEvalues indicate higher misconfiguration risk and potential AI data exposure.*

Furthermore, although Microsoft offers logging and audit features, other sources indicate that logging of LLM-specific events has not progressed as much. Traditional logs provide information about who accessed which files, but not about how Copilot's generative model worked with the underlying data [23]. Without clear tracking of prompts and outputs, accountability becomes difficult in the case of data leakage. Organizations also encounter scrutiny from AI observability tools that can elucidate model behaviors, recognize when things of interest are 'off' or 'out of the ordinary' from the model and even alert or flag incidents in real-time per violations of policy [24]. These requirements demonstrate the necessity of AI-attuned governance frameworks - systems of policy that identify and govern how those model's function.

Human factors are essential for the overall security posture of Copilot, and from the organization viewpoint qualitative studies emphasize employee understanding and training in the ethics of AI Variational Stacked LSTM Network for High-Dimensional Threat Detection. Uninformed users might unknowingly submit sensitive materials as a prompt or may not understand that the output they generated or are reviewing

might contain sensitive information [25]. According to the secondary case analysis, organizations that couple technical controls with regularly ongoing AI security awareness training have much greater compliance and subsequently decreased risk level.

### 3.4. Comparative Platform Resilience

To evaluate Copilot's security and privacy position, we performed a qualitative comparative analysis using publicly available information about Google Workspace Duet AI and Salesforce Einstein. The results showed similarities and differences between the three platforms. Each also has the possibility of overexposing data through AI generated content. As a security and privacy position, each also counts on user/administrative permissions and tenant isolation. Microsoft 365 Copilot is distinct from Google Workspace Duet AI and Salesforce Einstein in the level of ecosystem integration and depth of access to data, which means the consequences of a security incident may be more severe than for the other platforms [26-28].

Google Workspace Duet AI points to an emphasis on integration with Google's data governance suite and claims it effectively uses real-time security scanning on its prompts related to content creation. While this adds some filtering ability at the prompt level, there is only limited visibility into the model reasoning and therefore limited forensic awareness post incident. In the case of Salesforce Einstein, there is more elaborate auditing and explainability aligned with the CRM problem space, but it similarly suffers from a multi-tenant vulnerability. In summary, comparisons indicate that Copilot's security architecture is more mature than that of compliance, but it adds complexity as it hybridizes local processing with in-cloud processing, [29]. So, these factors do not only suggest that Copilot has established itself as a leader in enterprise AI in the broader ecosystem but conversely place it as a brand-new focal point of cybersecurity concerns.

### 3.5. Synthesis of Findings

When we draw together the complete domains, we draw a few central conclusions. First, Microsoft 365 Copilot is a new example for enterprise threat modeling, reflecting a shift in the AI assistant's role from productivity tool to potential method of data exposure [30]. Second, existing privacy and governance tools may be useful, but they remain decisively reactive-not proactive-and need to be supplemented by tools and policies already in place, and/or other known technologies, with AI [31]. Third, this comparison shines a light on the systematic aspects of GenAI-related risks across SaaS systems and consequently raises the issue of cross-industrial standards for meta governance of AI [32]. Fourth, qualitative data lends support for the thesis that firms that assemble technology protections with access to cultural awareness-consumer user training, continual auditing styles, and AI styles including transparency-could take much stronger security relief [33]. In summary, we can advocate for the partition that Microsoft 365 Copilot possesses some promise, but it has also added new and/or measurable risks. It is clear these risks can only be mitigated through a layered security architecture lacking not only organizational control but effective AI governance. The following section presents the implications of the findings for enterprise security and governance strategies, regulatory compliance, and the development of the AI governance model in SaaS ecosystems.

## 4. Discussion

The results from the second qualitative analysis provide important perspectives on the implications of security and privacy using Microsoft 365 Copilot in an enterprise context. The use of Generative AI in, for example, productivity workflows is a new technological paradigm in order to optimize business processes, provide automation and assist decisions through intelligent synthesis of organizational knowledge [34]. However, the benefits of using Copilot include its access to emails, documents, chat messages and calendars that increase risk to information security, and introduces unique risks created in a more complex context [35]. Copilot is continuously processing information from multiple context sources resulting in an output of dynamically contextualized information, i.e., utilizing folder-based data that can be sensitive by default [36]. The dynamism introduces risk of data leakage that has to be dealt with by an encompassing AI-aware security policy.

The outcomes of threat modeling suggest Copilot has a significant effect on an enterprise’s attack surface. A malevolent insider, a compromised account, or a third-party taking advantage of AI-mediated vulnerabilities such as prompt injections, could steal data that they would not have had access to according to traditional access controls [37]. Further qualitative evidence indicates that the issue with these attacks is not entirely theoretical because case studies and technical advisories describe cases in which the LLMs in the enterprise setting produced outputs that contain sensitive data by mistake. These situations demonstrate that the LLM architecture of Copilot could be misused to bypass existing security measures, and create a second type of data leak, and that enterprises must address these AI-generated threat vectors almost as an extension of their overall security posture [38]. Furthermore, this challenge, or risk was made worse by the existence of multi-tenant environments, where there is a potential for data leakage across organizations if proper controls do not exist for AI products to either manage or separate enterprise groups.

Privacy issues are also of major importance. The handling of PII, PHI, and proprietary corporate data of Copilot creates regulatory and compliance implications that go beyond simply cybersecurity concerns. The secondary qualitative analysis indicates that, even with traditional DLP policies and monitoring via Microsoft Purview, AI-generated outputs can still be

put into context by outputs that contain confidential data. Copilot, for instance, could aggregate confidential documents and any associated data sources by creating a logical and relevant response to the user's request and disclosure in some manner (39). Regulatory schemes inclusive of GDPR, HIPAA, and CCPA all have strong obligations related to the protection of personal and sensitive data, centered around ideas of data minimization, auditability, and data interface in the event of a data breach (40). Thus, enterprises employing Copilot should consider additional oversight / controls on AI-driven workflows, including real-time tracking of outputs as they are generating, modeling user prompts and inquiry in a way that confirms accurate outputs, and notifying the enterprise when abnormalities point to unauthorized access of data [39, 40].

Governance and access control are also essential enablers for deploying Copilot tools securely. RBAC policies, information barriers, logging capabilities, and auditing capabilities were assessed and are appropriate access controls on their own but are not adequate [41]. Outside sources indicate that configuration errors or permissions that are overly permissive can effectively increase the access scope of the Copilot integration to unauthorized users or to sensitive data. Thus, organizations need to add governance approaches that would allow strict role-based policies to govern human action but with AI-aware monitoring and anomaly detection [42]. Outside sources indicated the use of AI observability tools that will track model inputs, outputs, and decision pathways in the name of accountability and compliance [43]. With audit trails in addition to proactive governance, enterprise organizations can address risks and productivity advancements arising from AI integration.

When compared to the security posture of Copilot, Google Workspace Duet AI, and Salesforce Einstein adds more context around Copilot. DLP integration is more robust in Duet AI, while Einstein has a more robust focus on auditing and access controls. As with Copilot, these applications can also have the same types of assumptions that could compromise them, even with AI being introduced to content generation and prompt manipulation [44]. Therefore, the purpose of comparison really reinforces the notion that security concerns in LLM integration would be concerns in all AI-enabled enterprise applications and not just one application. While organizations develop appropriate platform-specific policies, they must also be cognizant of a broader AI governance framework, for instance monitoring model behavior, validating prompts and isolating tenants.

The implications of these findings are going to be of a practical nature. First, to whatever extent it means to manage risk of an enterprise, the risk management framework related to enterprise risk would need to explicitly consider threat vectors and privacy issues surrounding AI. In addition, security policies will need to evolve from traditional and shift functions about the use context of risk, where access control and encryption were prioritized, and shift to address specifically the LLM enabling processes with their risk vectors related to prompt injection, and unintended data generation. Second, employees will need education surrounding the capabilities and risks of Copilot, and some type of awareness training to minimize the possibility of accidental disclosure or exploitation. Third, compliance within regulatory may need to be (re)built into the AI workflow proactively; as well as comprehensive auditing and anomalous detection processes around a process for ongoing incident response to any AI enabled process around GenAI flows. Finally, and related to first as a process, there will be continuous measures process for measurement in checking against broader AI enables platforms to more conveniently, those you cannot monitor otherwise via standard benchmarks. The measure process will allow discovery and means to be informed on practices identification either through accidental discovery or ongoing threats or changes to measures. Leveling newly in AI, will create institutional culture in building continuous improvement in a QC process around the organization governance in AI.

To summarize, the discussion highlights that Microsoft 365 Copilot presents both an opportunity and a challenge for organizations because, while there is potential for workflow optimization and enhanced human productivity, emotionally linking these to a complicated set of security and privacy challenges, needs to be finessed within the governance framework, which is, in and of itself, AI-aware (and therefore, can create further risks). Secondary qualitative analysis shows that to mitigate risk, threat modelling must become explicit, privacy assessments and governance controls must be in place, and employees must be trained in how to use the platform, and how to use the data derived from that use in workflows; even understanding platform benchmarking would be of value. An Integrative perspective provides organizations with a safe way to harness the potential of Copilot while still being compliant within established frameworks, protecting sensitive data, and mitigating the consequences of AI-based security breaches. The insights here will offer a current foundation not only for providing organizations with pragmatic guidance, but for future research on how to implement LLMs not only securely, but also authentically in a privacy conscious way, within the scope of organized contexts.

***Table 1.*** *Comparative analysis of GenAI SaaS security dimensions.*

| Parameter | Microsoft 365 Copilot | Google Duet AI | Salesforce Einstein | Observations |
|---|---|---|---|---|
| Data Access Depth | High (Graph API, M365 data) | Moderate | Moderate | Copilot highest exposure |

| Parameter | Microsoft 365 Copilot | Google Duet AI | Salesforce Einstein | Observations |
|---|---|---|---|---|
| Explainability | Limited | Low | Medium | Need for interpretability |
| DLP Integration | Native Purview | Partial | Strong | Microsoft leads |
| Tenant Isolation | Strong | Strong | Variable | Similar across vendors |

## 5. Conclusion

This paper shows that Microsoft 365 Copilot, although it is a revolutionary change in productivity in the workplace, essentially changes the paradigm of enterprise security. The secondary qualitative analysis of available literature and security reports, as well as compliance frameworks, demonstrates that GenAI-based assistants can be viewed as dual-identity phenomena; they enable innovation and, at the same time, represent potential vectors of data exposure. The extent of integration of Copilot into Microsoft 365 will provide the company with first-time access to organizational information resources, such as emails, documents, chat messages, and calendars. This wide scope presents multidimensional risks including breach of unauthorized disclosure, non-compliance with the regulations, and the intensification of insider threats.

The results highlight that traditional security measures cannot be used to deal with AI-mediated interactions. Role-based access control, encryption, and DLP are also important underpinnings that need to be improved to respond to the LLM-specific vulnerabilities identified like prompt injection and context manipulation. The partial blackness of Microsoft Purview and other governance tools is also mentioned in the analysis; although these tools allow increasing visibility and compliance, additional monitoring is needed (AI-conscious logging, model-output verified, and real-time anomaly detection). The organizations therefore need to shift the state of immobile enforcement of policies to responsive intelligence-driven administration that constantly examines the manner through which GenAI systems process and produce information.

By comparison, the risk profile of Copilot can be compared to that of Google Workspace Duet AI and Salesforce Einstein, and it can be concluded that GenAI-related security issues are endemic to SaaS ecosystems. This supports the necessity to develop industry-wide standards and collaboration between vendors in the field of AI safety, transparency, and interoperability. Both regulators and enterprises need to come up with frameworks that combine the notion of privacy-by-design, explainable-AI, and automated compliance auditing to ensure that people trust AI-driven processes.

## Abbreviations

| | |
|---|---|
| AI | Artificial Intelligence |
| ACE | Access Control Effectiveness |
| CCPA | California Consumer Privacy Act |
| DLP | Data Loss Prevention |
| GDPR | General Data Protection Regulation |
| GenAI | Generative Artificial Intelligence |
| HIPAA | Health Insurance Portability and Accountability Act |
| LLM | Large Language Model |
| M365 | Microsoft 365 |
| PHI | Protected Health Information |
| PII | Personally Identifiable Information |
| PL | Privacy Leakage |
| RBAC | Role-Based Access Control |
| SaaS | Software-as-a-Service |

## Author Contributions

**Pullaiah Chowdary Vutla:** Conceptualization, Data curation, Formal Analysis, Investigation, Methodology, Software, Writing – original draft, Writing – review & editing

**Triveni Yenugu:** Conceptualization, Data curation, Formal Analysis, Methodology, Software, Validation, Visualization, Writing – original draft, Writing – review & editing

## Conflicts of Interest

The authors state that they have no financial or personal connections that could have affected this work. The study received no funding or support from any commercial or external organization and was carried out independently. There are no competing interests related to the methods, results, or conclusions of this study.

## References

[1] Q. Zhang, J. Zuo, and S. Yang, "Research on the impact of generative artificial intelligence (GenAI) on enterprise innovation performance: a knowledge management perspective," *Journal of Knowledge Management*, Apr. 2025, https://doi.org/10.1108/jkm-10-2024-1198

[2] P. C. Vutla and T. Yenugu, "Intelligent Intrusion Detection with Secure Cloud Storage Enabled by a Hybrid DNN-FNN Framework," *2026 Fourth International Conference on Augmented Intelligence and Sustainable Systems (ICAISS)*, Trichy, India, 2026, pp. 401-406, https://doi.org/10.1109/ICAISS68683.2026.11526496

[3] H. M. S. S. Herath, H. M. K. K. M. B. Herath, B. G. D. A. Madhusanka, and L. G. P. K. Guruge, "Data Protection Challenges in the Processing of Sensitive Data," *Data Protection*, pp. 155–179, 2024, https://doi.org/10.1007/978-3-031-76473-8_8

[4] F. Forsén, "Large Language Models and business applications in an R&D environment," *Theseus.fi*, 2024, https://www.theseus.fi/handle/10024/863103

[5] P. C. Vutla and T. Yenugu, "AI-Driven Cloud Security Framework Using Deep Feature Learning and Hybrid Threat Classification," *2026 International Conference on Visual Analytics and Data Visualization (ICVADV)*, Tirunelveli, India, 2026, pp. 163-167, https://doi.org/10.1109/ICVADV67766.2026.11469856

[6] Narasimha Rao Oruganti, "Integrating Azure AI Copilot into Enterprise Applications: A Conceptual Framework for Adoption and Impact," *Journal Of Engineering And Computer Sciences*, vol. 4, no. 9, pp. 433–444, 2025.

[7] "Generative AI Security," *SpringerLink*, 2024, https://doi.org/10.1007/978-3-031-54252-7

[8] M. Zbořil, "Security risks associated with deployment of AI solutions into organizations," *IDIMT-2024: Changes to ICT, Management, and Business Processes through AI*, 2024.

[9] P. P. Ray, "A Survey on Model Context Protocol: Architecture, State-of-the-art, Challenges and Future Directions," Apr. 2025, https://doi.org/10.36227/techrxiv.174495492.22752319

[10] Ç. Aksoy and E. Söğüt, "PROMPT ATTACKS ON LARGE LANGUAGE MODELS CASE STUDY: A CYBER SECURITY SCENARIO," *Gazi.edu.tr*, 2025, https://doi.org/02a0f8ae-0d02-4a7f-82db-d028a70c810a

[11] W. Hashim and N. A.-H. K. Hussein, "Securing Cloud Computing Environments: An Analysis of Multi-Tenancy Vulnerabilities and Countermeasures," *SHIFRA*, vol. 2024, pp. 8–16, Feb. 2024, https://doi.org/10.70470/shifra/2024/002

[12] W. Z. Khan, Hareem Kibriya, A. Siddiqa, and M. K. khan, "Privacy Issues in Large Language Models: A Survey," Jan. 2024, https://doi.org/10.2139/ssrn.4871294

[13] A. Dang, "Derivative Data: Rethinking Market Definitions in the Age of Generative AI," *UC Law SF Scholarship Repository*, 2025.

[14] A. Tomassi, "Data Security and Privacy Concerns for Generative AI Platforms - Webthesis," *Polito.it*, Oct. 2024.

[15] C. J. Chong, Z. Yao, and I. Neamtiu, "Artificial-Intelligence Generated Code Considered Harmful: A Road Map for Secure and High-Quality Code Generation," *arXiv.org*, 2024.

[16] D. Atkinson, "Open Shouldn't Mean Exempt: Open-Source Exceptionalism in Generative AI," 2025, https://doi.org/10.2139/ssrn.5355736

[17] S. Banerjee, "Customizing security policies for data-centric program domains with multi-tenant threat models," *Utexas.edu*, Dec. 2024. Available: https://repositories.lib.utexas.edu/items/1d0dcf67-2101-4494-92dc-a92d39753c87 [Accessed: Oct. 15, 2025].

[18] R. Muppalaneni, Anil Chowdary Inaganti, and N. Ravichandran, "AI-Enhanced Data Loss Prevention (DLP) Strategies for Multi-Cloud Environments," *Journal of Computing Innovations and Applications*, vol. 2, no. 2, pp. 1–13, 2024, Available: https://ciajournal.com/index.php/jcia/article/view/9 [Accessed: Oct. 15, 2025].

[19] J. Koskula, "Generative artificial intelligence in support of analytics: Copilot 365," *Lutpub.lut.fi*, 2024.

[20] S. Ranjan, Divya Chembachere, and L. Lobo, "Agentic AI in Enterprise," *SpringerLink*, 2025, https://doi.org/10.1007-979-8-8688-1542-3

[21] G. A. Gani, "Securing AI Agents: Implementing Role-Based Access Control for Industrial Applications," *arXiv.org*, 2025. Available: https://arxiv.org/abs/2509.11431 [Accessed: Oct. 15, 2025].

[22] H. Khuat, "Leveraging Generative AI for Data Engineering Workflows," *Journal of Computer Science and Technology Studies*, vol. 7, no. 3, pp. 120–140, May 2025, https://doi.org/10.32996/jcsts.2025.7.3.14

[23] N. D. Annam, "AI-Powered Data Observability & Governance Agent for Cloud Analytics: Transforming Enterprise Data Management," *Journal of Computer Science and Technology Studies*, vol. 7, no. 3, pp. 804–811, May 2025, https://doi.org/10.32996/jcsts.2025.7.3.88

[24] C. Chen, X. Gong, Z. Liu, W. Jiang, S. Q. Goh, and K.-Y. Lam, "Trustworthy, Responsible, and Safe AI: A Comprehensive Architectural Framework for AI Safety with Challenges and Mitigations," *arXiv.org*, 2024. Available: https://arxiv.org/abs/2408.12935 [Accessed: Oct. 15, 2025].

[25] P. C. Vutla and T. Yenugu, "Quantum-Enhanced Variational Stacked LSTM Network for High-Dimensional Threat Detection in Cloud Computing Environments," *2026 5th International Conference on Sentiment Analysis and Deep Learning (ICSADL)*, Birendranagar, Nepal, 2026, pp. 152-157, https://doi.org/10.1109/ICSADL67539.2026.11451816

[26] N. K. Sehgal, M. Saxena, and D. N. Shah, "AI on the Edge with Security," *SpringerLink*, 2025, https://doi.org/10.1007-978-3-031-78272-5

[27] D. Parmar, "Enhancing Customer Relationship Management with Salesforce Einstein GPT," *Theseus.fi*, 2023, https://www.theseus.fi/handle/10024/812434

[28] R. Vadisetty, A. Polamarasetti, Dr. S. kumar Rongali, S. Prajapati, and J. B. Butani, "Leveraging Generative AI for Automated Code Generation and Security Compliance in Cloud-Based DevOps Pipelines: A Review," *SSRN Electronic Journal*, 2025, https://doi.org/10.2139/ssrn.5218298

[29] M. Uddin *et al.*, "Generative AI revolution in cybersecurity: a comprehensive review of threat intelligence and operations," *Artificial Intelligence Review*, vol. 58, no. 8, May 2025, https://doi.org/10.1007/s10462-025-11219-5

[30] Y. Devi, "The Future of Cybersecurity: Predicting Trends and Preparing for Emerging Threats," *International Journal of Emerging Research in Engineering and Technology*, pp. 263–275, 2025, https://doi.org/10.63282/3050-922X.ICRCEDA25-128

[31] H. Vannesluoma, "The use of generative artificial intelligence from an innovation process perspective," *Lutpub.lut.fi*, 2024, https://lutpub.lut.fi/handle/10024/168538

[32] "Artificial intelligence in financial auditing: redefining accuracy and transparency in assurance services," *EDPACS*, 2025, https://doi.org/10.1080//07366981.2025.2459490

[33] J. T. Liang, C. Yang, and B. A. Myers, "A Large-Scale Survey on the Usability of AI Programming Assistants: Successes and Challenges," *arXiv (Cornell University)*, Jan. 2023, https://doi.org/10.48550/arxiv.2303.17125

[34] A. Sarkar, Xiaotong, Xu, N. Toronto, I. Drosos, and C. Poelitz, "When Copilot Becomes Autopilot: Generative AI's Critical Risk to Knowledge Work and a Critical Solution," *arXiv.org*, 2024.

[35] A. Mohsin, H. Janicke, A. Wood, I. H. Sarker, L. Maglaras, and N. Janjua, "Can We Trust Large Language Models Generated Code? A Framework for In-Context Learning, Security Patterns, and Code Evaluations Across Diverse LLMs," *arXiv.org*, 2024.

[36] P. P. Ray, "A Comprehensive Introspection on AI Risks: Taxonomy, Challenges and Future Directions," Jul. 2025, https://doi.org/10.36227/techrxiv.175339321.17050891/v1

[37] Y. Potter *et al.*, "Frontier AI's Impact on the Cybersecurity Landscape," *arXiv.org*, 2025.

[38] A. Wodi, "Artificial Intelligence (AI) Governance: An Overview," *SSRN Electronic Journal*, Jan. 2024, https://doi.org/10.2139/ssrn.4840769

[39] R. Nayak, Umashankar Ghugar, P. Gupta, S. Dash, and N. Gupta, "Data Privacy and Compliance in Information Security," pp. 17–33, Feb. 2025, https://doi.org/10.1002/9781394268917.ch2

[40] A. Abusini, "Enhancing Smart Home Security Through Risk-Based Access Control (RBAC): 'Closing the Gap,'" *Beadle Scholar*, 2024.

[41] L. Karadsheh, J. E. Barnard, M. Arafah, and A. F. Shubita, "Generative AI and the Evolving Threat Landscape," *Advances in computational intelligence and robotics book series*, pp. 71–98, May 2025, https://doi.org/10.4018/979-8-3373-0832-6.ch004

[42] C. Novelli, M. Taddeo, and L. Floridi, "Accountability in artificial intelligence: what it is and how it works," *AI & SOCIETY*, vol. 39, no. 4, pp. 1871–1882, Feb. 2023, https://doi.org/10.1007/s00146-023-01635-y

[43] None Surendra Vitla, "The Future of Identity and Access Management: Leveraging AI for Enhanced Security and Efficiency," *Journal of Computer Science and Technology Studies*, vol. 6, no. 3, pp. 136–154, Aug. 2024, https://doi.org/10.32996/jcsts.2024.6.3.12

[44] M. A. Ferrag, N. Tihanyi, D. Hamouda, L. Maglaras, and M. Debbah, "From Prompt Injections to Protocol Exploits: Threats in LLM-Powered AI Agents Workflows," *arXiv.org*, 2025.