**International Journal of Emerging Trends in Computer Science and Information Technology**
ISSN: 3050-9246 | https://doi.org/10.63282/3050-9246/ICRTCSIT-116
Eureka Vision Publication | ICRTCSIT'25-Conference Proceeding

*Original Article*

# The AI-Augmented Data Engineer: How LLMs and Copilots are Redefining the Engineering Workflow

Sougandhika Tera
Data Engineer, Filterona, USA.

*Abstract - This paper examines how Large Language Models (LLMs) and AI copilots such as GitHub Copilot and ChatGPT are transforming the role of the modern data engineer. By automating routine tasks like code generation, debugging, documentation, and query optimization, these tools allow engineers to focus on higher-level architectural decisions, innovation, and collaboration. The paper explores current impacts, potential risks, and future opportunities of adopting Aico pilots within enterprise data engineering workflows, supported by research insights, productivity studies, and real-world use cases.*

*Keywords - Artificial Intelligence, Data Engineering, Large Language Models, Copilot, Automation, Productivity*.

## 1. Introduction

The role of data engineers is evolving rapidly as artificial intelligence reshapes how software is built and maintained. Traditionally, engineers have spent considerable time on repetitive coding, debugging, and documentation tasks. How- ever, with the emergence of AI copilots, these lower-level tasks are increasingly automated. Instead of replacing engineers, copilots augment their capabilities, improving productivity and reducing cognitive load. This shift opens the door for data engineers to dedicate more time to strategic problem-solving, advanced architecture, and business-driven analytics solutions.

## 2. AI Copilots and the Data Engineering

### *2.1. Workflow*

Copilots are embedded directly into Integrated Development Environments (IDEs) such as Visual Studio Code and Jet- Brains, allowing engineers to receive contextual code recommendations without breaking their workflow. For data engineers, this means auto-generating SQL queries, ETL scripts, or Spark transformations with minimal keystrokes. A GitHub (2023) study showed that developers using Copilot completed tasks 55% faster than their peers, underscoring the measurable productivity benefits.AI copilots like GitHub Copilot and ChatGPT provide real-time code suggestions, automate SQL query generation, and assist in debugging errors. They also support automated documentation, ensuring codebases re- main maintainable overtime. According to GitHub (2023), developers using Copilot complete tasks 55%faster, highlighting the potential time savings for data engineering teams. The integration of copilots into IDEs allows smoother workflow execution, reducing context switching and helping engineers stay focused on solving complex business problems rather than boilerplate coding.

## 3. Benefits of AI-Augmented Engineering

AI copilots deliver tangible benefits across multiple dimensions. Efficiency gains come from eliminating boilerplate work such as schema definitions or routine pipeline code. Code quality improves as copilots enforce standardized design pat- terns. Onboarding new engineers becomes faster with copilots acting as real-time mentors. Teams also report stronger collab- oration, since copilots create shared conventions for handling tasks like error logging or incremental loads. These factors collectively shift engineering focus away from repetitive coding and towards architectural innovation. The primary benefit of copilots lies in efficiency. By generating boilerplate ETL code, writing unit tests, or suggesting schema definitions, engineers save significant time. Improved code quality is another advantage, as copilots often recommend patterns consistent with best practices. AI copilots also facilitate onboarding by providing real-time learning aids for junior engineers. Finally, copilots improve collaboration between teams by standardizing approaches to common data engineering tasks such as incremental loads, error handling, and monitoring scripts.

## 4. Data and Figures

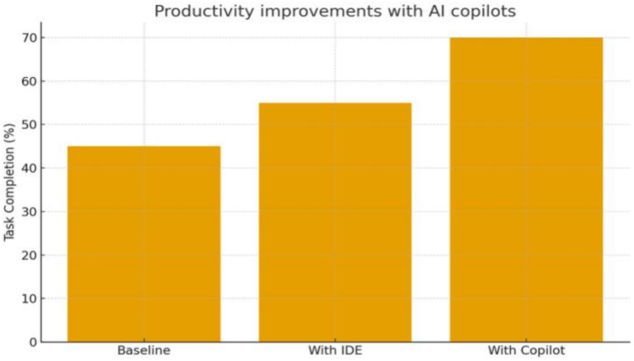

**Figure 1. Productivity improvements with AI copilots in development tasks**

**Table 1. Comparison of Data Engineering Tasks before and After AI Copilot Adoption**

| Task | Before Copilot | After Copilot |
|---|---|---|
| Code Generation | Manual boilerplate coding | Automated snippets and SQL generation |
| Debugging | Time-consuming error tracing | AI-suggested fixes in real-time |
| Documentation | Often neglected or outdated | Generated alongside code |
| Onboarding | Steep learning curve | Guided by real-time AI suggestions |
| Focus Areas | Routine coding tasks | Higher-level architecture and optimization |

**Table 2. Types of Data Engineering Tasks Automated By AI Copilots**

| Task Area | Copilot Contribution |
|---|---|
| SQL Querying | Generated optimized queries, joins, and aggregation logic |
| ETL Scripts | Produces boilerplate Python/PySpark transformations |
| Debugging | Suggests fixes for runtime and syntax errors |
| Documentation | Auto-generated inline comments and docstrings |
| Testing | Creates unit tests for data validation and edge cases |

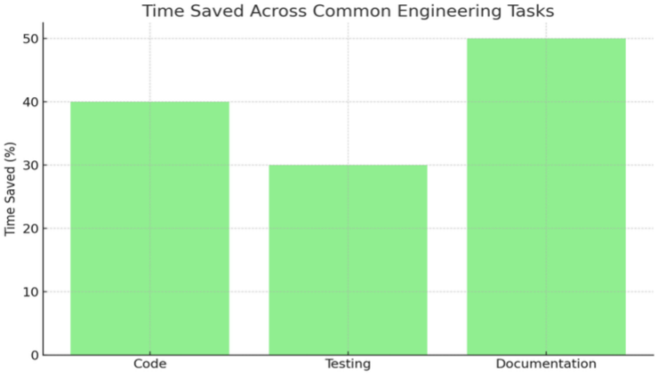

**Figure 2. Time Saved across Common Tasks When Copilots Areused**

**Table 3. Common Risks of AI Copilots and Mitigation Strategies**

| Risk | Mitigation Strategy |
|---|---|
| Incorrect code generation | Mandatory peer reviews and automated testing pipelines |
| Security/compliance breaches | Restrict copilots from accessing sensitive datasets |
| Engineer dependency | Training programs to balance AI use with problem-solving skills |
| Bias in generated outputs | Diverse training datasets and human oversight |

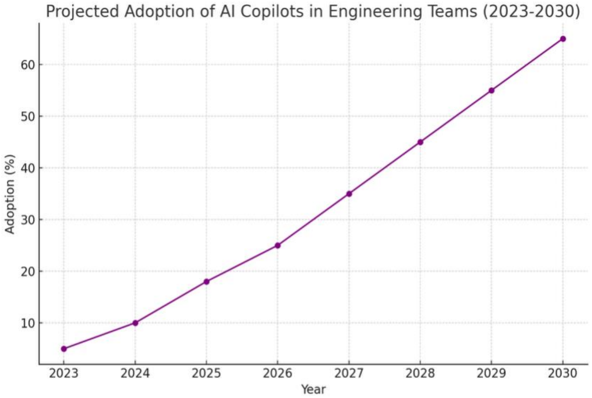

**Figure 4. Projected adoption of AI copilots among engineering teams (20232030)**

## 4. Risks and Challenges

Despite these clear advantages, copilots also introduce risks that cannot be ignored. LLMs occasionally produce incorrect queries or inefficient transformations that may go unnoticed if engineers are overly reliant on automation. Compliance risks arise if sensitive datasets or proprietary code snippets are shared with external AI services. Moreover, dependency on copilots could erode critical thinking and problem-solving skills over time. Organizations must therefore enforce governance policies and encourage human review to mitigate these risks, ensuring copilots remain assistants rather than authorities. Despite these advantages, risks remain. Over-reliance on copilots can reduce critical thinking and introduce hidden vulnerabilities. LLMs sometimes generate incorrect or nonoptimized queries, requiring careful human review. Security and compliance risks also arise if sensitive data is shared with AI assistants. Organizations must therefore balance automation with robust governance practices, ensuring AI copilots act as supportive tools rather than unquestioned authorities.

## 5. Future Outlook

The trajectory of copilot’s points toward proactive and autonomous engineering support. Beyond assisting with queries or code, future copilots may automatically optimize pipelines, detect anomalies in data streams, and design architectures visually. Industry analysts predict widespread adoption of copilots across software and data engineering teams by 2030, with copilots evolving into multi-modal systems capable of generating not just code but also dashboards, architecture diagrams, and governance policies. This transition suggests that the role of the data engineer will continue to shift toward higher-order responsibilities like data strategy, compliance, and innovation leadership. The future of data engineering will be deeply augmented by AI copilots. As models continue to improve, copilots will expand from code generation to proactive optimization, anomaly detection, and automated data pipeline monitoring. Visual copilots will also emerge, helping engineers design architectures and visualize dependencies. Far from replacing hu- man engineers, copilots will redefine workflows to emphasize creativity, strategic oversight, and advanced problem-solving in enterprise data ecosystems.

## 6. Conclusion

AI copilots mark a turning point in theevolution of data engineering. By automating repetitive tasks and providingin- telligent assistance, they increase productivity, elevate code quality, andfree engineers to focus on high-level architectural challenges. While risksexist, with proper governance and over- sight, the integration of copilots intoengineering workflows has the potential to usher in a new era of innovation,efficiency, and business value in the field of data engineering.

## References

[1] J. Kaddour, J. Harris, M. Mozes, H. Bradley, R. Raileanu, and R. Mchardy, Challenges and applications of large language models, *ArXiv*. Jul. 2023.

[2] P. Vaithilingam, T. Zhang, and E. Glassman, Expectation vs. Experience: Evaluating the usability of code generation tools powered by LLMs, *CHI conference*. 2022.

[3] Teja Thallam, N. S. (2025). AI-Powered Monitoring and Predictive Maintenance for Cloud Infrastructure: Leveraging AWS Cloud Watch and ML. *International Journal of Artificial Intelligence, Data Science, and Machine Learning*, *6*(1), 55-61. https://doi.org/10.63282/3050-9262.IJAIDSML-V6I1P107

[4] Y.Gao, Research: Quantifying GitHub copilots impact in the enterprise with accenture, *The GitHub blog*. May 2024.

[5] K. R. Kotte, L. Thammareddi, D. Kodi, V. R. Anumolu, A. K. K and S. Joshi, "Integration of Process Optimization and Automation: A Way to AI Powered Digital Transformation," *2025 First International Conference on Advances in Computer Science, Electrical, Electronics, and Communication Technologies (CE2CT)*, Bhimtal, Nainital, India, 2025, pp. 1133-1138, doi: 10.1109/CE2CT64011.2025.10939966.

[6] Reddy, R. R. P. (2024). Enhancing Endpoint Security through Collaborative Zero-Trust Integration: A Multi-Agent Approach. *International Journal of Computer Trends and Technology*, *72*(8), 86-90.

[7] B. C. C. Marella, G. C. Vegineni, S. Addanki, E. Ellahi, A. K. K and R. Mandal, "A Comparative Analysis of Artificial Intelligence and Business Intelligence Using Big Data Analytics," *2025 First International Conference on Advances in Computer Science, Electrical, Electronics, and Communication Technologies (CE2CT)*, Bhimtal, Nainital, India, 2025, pp. 1139-1144, doi: 10.1109/CE2CT64011.2025.10939850.

[8] Thirunagalingam, A. (2024). Transforming real-time data processing: the impact of AutoML on dynamic data pipelines. Available at SSRN 5047601.

[9] Sai Krishna Gunda (2024). Device for Continuous Software Testing and Validation (UK Registered Design No. 6400738). Registered with the UK Intellectual Property Office, Class 14-02, granted in November 2024.

[10] Maroju, P. K. (2024). Advancing synergy of computing and artificial intelligence with innovations challenges and future prospects. FMDB Transactions on Sustainable Intelligent Networks, 1(1), 1-14.

[11] Sandeep Rangineni Latha Thamma reddi Sudheer Kumar Kothuru , Venkata Surendra Kumar, Anil Kumar Vadlamudi. Analysis on Data Engineering: Solving Data preparation tasks with ChatGPT to finish Data Preparation. Journal of Emerging Technologies and Innovative Research. 2023/12. (10)12, PP 11, https://www.jetir.org/view?paper=JETIR2312580

[12] Sehrawat, S. K. (2023). The role of artificial intelligence in ERP automation: state-of-the-art and future directions. *Trans Latest Trends Artif Intell*, *4*(4).

[13] Sudheer Panyaram, (2025). Optimizing Processes and Insights: The Role of AI Architecture in Corporate Data Management. IEEE.

[14] Garg, A., Pandey, M., & Pathak, A. R. (2024). A Multi-Layered AI-IoT Framework for Adaptive Financial Services. *International Journal of Emerging Trends in Computer Science and Information Technology*, *5*(3), 47-57. https://doi.org/10.63282/3050-9246.IJETCSIT-V5I3P105

[15] Vijay Kumar Kasuba, (2025). Investigating the Issues and Challenges of Remote Working on Project Management: Case Studies from India. International Journal of Computer Trends and Technology(IJCTT), Volume 73 Issue 5, 64-69, May 2025

[16] Thallam, N. S. T. (2024). The Rise of Generative AI: Transforming Industries with Large Language Models and Deep Learning. *IJSAT-International Journal on Science and Technology*, *15*(4).

[17] Rajender Pell Reddy, "Cybersecurity for Critical Infrastructure: Protecting National Assets in the Digital Age," *International Journal of Computer Trends and Technology (IJCTT)*, vol. 73, no. 2, pp. 7-17, 2025. *Crossref*, https://doi.org/10.14445/22312803/ IJCTT-V73I2P102

[18] Pugazhenthi, V. J., Pandy, G., Jeyarajan, B., & Murugan, A. (2025, March). AI-Driven Voice Inputs for Speech Engine Testing in Conversational Systems. In *SoutheastCon 2025* (pp. 700-706). IEEE.

[19] Sehrawat, S. (2025). HealthTech Innovations: Revolutionizing Healthcare Access and Quality. In *Cutting-Edge Solutions for Advancing Sustainable Development: Exploring Technological Horizons for Sustainability-Part 2* (pp. 20-39). Bentham Science Publishers.

[20] Gopi Chand Vegineni. 2024/12/3. Exploring Anomalies in Dark Web Activities for Automated Threat Identification, FMDB Transactions on Sustainable Computing Systems. 2(4), PP - 189-200.