# Linux Foundation's LF AI & Data Foundation welcomes three new IBM open source projects



The Linux Foundation’s LF AI & Data Foundation has announced the addition of three new open source projects contributed by IBM, broadening its portfolio in key AI and data technology areas. The projects—Docling, Data Prep Kit, and BeeAI—address critical needs in semantic document understanding, enterprise-grade data preparation, and privacy-preserving federated learning, respectively.

LF AI & Data serves as an umbrella organisation under the Linux Foundation that supports open source innovation in artificial intelligence and data technologies. The three IBM-contributed projects were officially inducted by the LF AI & Data Technical Advisory Committee, a step that integrates them into the foundation’s broader governance and technical ecosystem.

BeeAI is presented as the first open-source platform supporting agent-to-agent communication, enabling developers to build, discover, run, and compose autonomous AI agents into multi-agent workflows. It leverages the open Agent Communication Protocol (ACP) to facilitate interoperability across different frameworks and technology stacks, supporting languages including JavaScript and Python.

Docling offers a comprehensive open-source ecosystem of Python tools designed for document conversion, generation, and manipulation. It focuses on extracting structured information from complex documents, which is critical for enhanced semantic understanding. The project has gained significant traction, with over 27,000 stars on GitHub, pointing to its growing status as a de facto standard in the space.

Data Prep Kit is designed as a modular suite to clean, transform, and trace unstructured data, specifically tailored for large language models (LLMs). It emphasises data quality, transparency, and scalability while supporting both batch and streaming data scenarios. This toolkit integrates seamlessly with modern AI workflows, aiming to enhance the rigour of enterprise data processes.

Todd Moore, Senior Vice President of Community Operations at the Linux Foundation and interim Executive Director of LF AI & Data, expressed enthusiasm for the new projects. Speaking to AiThority, he said, “We are excited to welcome Docling, Data Prep Kit, and BeeAI into the LF AI & Data family. These contributions from IBM reflect a strong commitment to open collaboration and responsible AI. I love BeeAI’s commitment to both Javascript and Python for aggregated learning.”

Brad Topol, Distinguished Engineer and Director of Open Source at IBM, highlighted the projects’ significance for accelerating innovation in generative AI. In remarks to AiThority, he noted, “Docling, Data Prep Kit, and BeeAI were born from a need to fill critical gaps in AI development tooling and accelerate innovation in the Generative AI space. We’re proud to see them as a catalyst enabling the broader open-source community to build AI applications and agentic workflows. We’re excited to collaborate with the open-source community to evolve these technologies and solve real-world challenges together.”

The governance of these projects will follow LF AI & Data’s model, benefiting from structured community-driven technical steering committees and ecosystem engagement. The projects are now publicly accessible, inviting participation from developers, data scientists, and researchers to contribute towards the future development of these impactful tools.

This expansion marks a notable enhancement in the tools available for AI workflows, with emphasis on open collaboration and scalable solutions for enterprise and developer communities.

Source: [Noah Wire Services](https://www.noahwire.com)

## Bibliography

1. <https://www.ibm.com/new/announcements/ibm-adds-open-source-projects-docling-beeaI-and-data-prep-kit-added-to-the-linux-foundation> - IBM's official announcement detailing the contribution of Docling, Data Prep Kit, and BeeAI to the Linux Foundation, highlighting their roles in semantic document understanding, data preparation, and federated learning.
2. <https://fossforce.com/2025/03/at-ato-ai-ibm-announces-its-handing-three-ai-projects-to-linux-foundation/> - Coverage of IBM's announcement at the All Things Open AI conference, discussing the donation of three AI projects to the Linux Foundation and their significance in the AI development stack.
3. <https://www.automationinside.com/article/ibm-contributes-key-open-source-projects-to-linux-foundation-to-advance-ai-community-participation> - An article highlighting IBM's contribution of Docling, Data Prep Kit, and BeeAI to the Linux Foundation, emphasizing their impact on AI community participation and open-source innovation.
4. <https://ossaidevjapan24.sched.com/event/1jKBm/data-prep-kit-a-comprehensive-cloud-native-toolkit-for-scalable-data-preparation-in-genai-app-daiki-tsuzuku-takuya-goto-ibm> - A session at the Open Source Summit + AI\_dev Japan 2024 discussing Data Prep Kit as a cloud-native toolkit for scalable data preparation in generative AI applications.
5. <https://docs.beeai.dev/introduction/welcome> - The official BeeAI documentation introducing the platform as an open-source solution for discovering, running, and composing AI agents from any framework.
6. <https://github.com/i-am-bee/beeai-framework> - The GitHub repository for BeeAI Framework, providing resources for building production-ready AI agents in Python and TypeScript.
7. <https://news.google.com/rss/articles/CBMiugJBVV95cUxPVTlneWJBbjEtUjFvSXNJcWx3WG5sU0F4SFY3ZzRrZVZzSDNrUWlCYTNoVF9MelpPM2lrM2RuMGprR05NNlRkeVg3QzlPZi15SUlkaGZ4ZkFrV3ZJQWJxc0E1Ynk1VEMtNGhVdUowVHpicS1NNTBSQkJYYzA2RG9faW1BZVpacGxaclJoc1Y3VmR3MUVWX1F1SmNvZm9mV3hLajBhV1RxbDNtVUEtNHdON1ZaeUxKTmZieWFGU2hLRUhvZ0Yyek5jSnFDQ3BkTU9QdlhvQnNYVGFOWEMtVkkxV0xRVmNqVndvUjFJQWZKanlNczBiSk5VZnNlZ3BHQ0pTc1JwenVuNVV0REtzZktSVTdwajZUVDljWml5dWk3VlE2dWNpbV9nM0lJbEhyMHZOMFhaeUhMeDdxUQ?oc=5&hl=en-US&gl=US&ceid=US:en> - Please view link - unable to able to access data