Enterprise Strategy Group™

BRIEF

SEPTEMBER 2025

Red Hat InstructLab on IBM Cloud: Democratizing Enterprise AI Development Through Simplified Model Customization

Torsten Volk, Principal Analyst

Abstract: Enabling application developers to easily, rapidly, and cost effectively add generative AI-driven capabilities to their software is quickly becoming a critical success factor. Red Hat InstructLab on IBM Cloud addresses the critical bottlenecks: skills, resource cost, training speed, and data accessibility to make generative AI available to any application developer.

Key Highlights

- 64% of organizations anticipated generative AI having a significant impact on innovation and creativity.[1]
- Red Hat InstructLab provides simple REST APIs for developers to harness generative AI.
- Organizations are currently in a race to offer more valuable individualized user experiences to their customer base and internal staff.

AI Adoption Barriers

Enterprises are in a race to unlock innovation, productivity, delivery speed, and customer satisfaction through the broad adoption of generative AI-driven product capabilities. In response, vendors of developer platforms have focused on providing comprehensive tools and integrations that streamline the development, deployment, and management of generative AI applications.

At the same time application developers still struggle with the upfront cost, skill requirements, access to training data, and the overall duration of the training process.

Generative AI requires high upfront investments in infrastructure due to its intensive computational needs, specialized hardware requirements, and substantial data storage and management demands to effectively train and deploy advanced AI models.

Skill requirements are often prohibitive for developers when adding generative AI capabilities. These technologies demand specialized expertise in complex areas such as advanced machine learning techniques, data science, and model training and fine-tuning. Additionally, developers must understand resource-intensive infrastructure management and navigate the ethical, regulatory, and security considerations specific to generative AI.

Training data requirements are often prohibitive for developers adding generative AI capabilities because obtaining large, diverse, high-quality data sets suitable for effective model training is resource-intensive, costly, and frequently restricted by privacy, compliance, and copyright considerations.

The often-extensive duration of generative AI model training, frequently spanning weeks or even months, can delay the time-to-market to an unacceptable degree.

[1] Source: Enterprise Strategy Group Research Report, *The State of the Generative AI Market: Widespread Transformation Continues*, September 2024. All research references in this Brief have been taken from this report.

Addressing these adoption barriers requires a radically simplified approach toward making generative AI capabilities available to application developers, ideally from within the comfort of their integrated development environment.

Market Insight

37% of organizations named "data quality" as a critical challenge specifically in implementing generative AI.

Red Hat InstructLab on IBM Cloud Broadens Developer Access to Generative AI

Red Hat InstructLab is now available as a fully managed service on IBM Cloud. The service provides enterprises with control over the location and use of their private data, while abstracting away the operational complexity of having to manage the model infrastructure stack and works with IBM's open source Granite models and other popular open source models such as Llama and Mistral.

Analyst Insight

Red Hat InstructLab on IBM Cloud impressively demonstrates IBM's ability to solve the hard problems by putting its research chops to work. Making large language models (LLMs) accessible to the average developer is one of these challenges and IBM's multi-pronged answer is impressive:

1. **Solving the data bottleneck:** LAB (Large-scale Alignment for ChatBots) addresses the often-insurmountable challenge of the lack of availability of training data. Created by MIT-IBM Watson AI Lab and IBM Research, LAB uses a specially trained LLM to understand the domain-specific concepts of a small number of question answer pairs. It then systematically creates variations that maintain semantic accuracy while expanding coverage.
2. **Simplified model customization:** Instead of having to train their own model to complete tasks or answer questions specific to a narrow domain or an organization, Red Hat InstructLab enables incremental model training via API. This lets developers directly collect training data from subject matter experts to then send it right to the Red Hat InstructLab API to enhance the model.

Conclusion

Providing development teams with the ability to experiment with and prototype generative AI-powered application capabilities represents a significant milestone in mainstreaming LLM adoption. The key breakthrough is eliminating two major barriers: the need for data scientist expertise and significant upfront costs.

Development teams often abandon their AI ambitions when confronted with LLMs' inability to reliably answer domain or organization-specific questions. This frustration typically ends AI exploration before it really begins.

InstructLab changes this dynamic by enabling developers to quickly create the required training data directly from subject matter expert input. They can then incrementally train a standard open source LLM until it delivers the desired answers. This capability should give a much wider range of developers the confidence and enthusiasm needed to fully commit to building AI-powered applications.

> "The ability to synthesize training data to then incrementally train the LLM and see instant results is a big deal when it comes to broadening AI adoption."
>
> **- Torsten Volk, Principal Analyst**, *Enterprise Strategy Group*

About Enterprise Strategy Group
Enterprise Strategy Group, now part of Omdia, provides focused and actionable market intelligence, demand-side research, analyst advisory services, GTM strategy guidance, solution validations, and custom content supporting enterprise technology buying and selling.

contact@esg-global.com
www.esg-global.com