

Quickly build production-ready LLM applications with Weights & Biases

Seamlessly track and evaluate applications to enhance performance

Organizations often struggle to effectively leverage large language models (LLMs) in production environments to power generative AI applications. Key difficulties include debugging inputs, outputs, and traces, which can be complex and time-consuming.

Weights & Biases provides a lightweight toolkit, W&B Weave, that helps track and evaluate LLM applications so you can optimize performance. With W&B Weave on Amazon Web Services (AWS), you can confidently move your LLM-powered proof of concept (POC) to production faster and accelerate application development.

About W&B Weave

W&B Weave helps developers confidently build and iterate on generative AI applications. Using the toolkit, you can:

- Create evaluations to score the performance of any aspect of your application.
- Examine and seamlessly debug failures by inspecting inputs and outputs.
- Ensure continuous performance of AI applications in production.

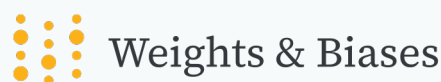
W&B Weave features

Traces

Spend less time debugging errors in your generative AI applications with Traces. W&B Weave automatically captures all input and output data and builds a trace tree to help you understand how data flows through your application.

With these insights, you can dive into failure modes to find hallucinations or inaccurate responses, and analyze how different inputs affect document retrieval, tool use, and custom behaviors from your application.

In collaboration with:



Hear what customers are saying about W&B Weave

"I love Weave for a number of reasons. The fact that I could just add a library to our code and all of a sudden I've got a whole bunch of information about the GenAI portion of our product, in Weights & Biases, which I was already using and very familiar with. All those things that I'm watching for the performance of our AI, I can just now report on quickly and easily with Weave."

—Mike Maloney,
Co-founder and CDO, Neuralift

W&B Weave features (cont.)

Evaluations

Seamlessly review and analyze LLM performance with Evaluations. This feature enables you to quickly organize and compare any performance metric, allowing you to measure and validate model accuracy across various scenarios.

Whether you use built-in scorers or define your own, you have the flexibility to create functions for your specific use case so you can track model behavior. This enables you to gain valuable insights to improve generative AI applications.

Weights & Biases and Amazon Bedrock

W&B Weave integrates with Amazon Bedrock to accelerate the development of LLM-powered applications. Whether you are evaluating different foundational models or retrieval-augmented generation (RAG)-tuned versions of the same model in Amazon Bedrock, W&B Weave provides the tools to seamlessly compare their performance. It also allows you to drill into each model's details to identify the best one for your application.

Outcomes

Accelerated development

Save time developing LLM-powered applications to achieve faster time to market.

Enhanced compliance and governance

Utilize a centralized record of all LLM prompt experiments to ensure a transparent history of AI development.

"I think Weave is just really simple, and what you get is a whole lot of summary metrics that you normally spend a while configuring yourself, and now you get everything recorded. It just make it so much quicker to debug what you're doing and iterate quickly."

—Georgina Corrie,
Applied Scientist, UnlikelyAI

Featured AWS products



Amazon Bedrock

Build and scale generative AI applications with foundation models.



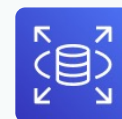
Amazon Elastic Kubernetes Service (Amazon EKS)

Run and manage Kubernetes clusters on AWS without installing and maintaining your own control plane.



Amazon Simple Storage Service (Amazon S3)

Store, manage, analyze, and protect any amount of data for virtually any use case.



Amazon Relational Database Service (Amazon RDS)

Set up, operate, and scale relational database management systems (RDBMS) in the cloud.

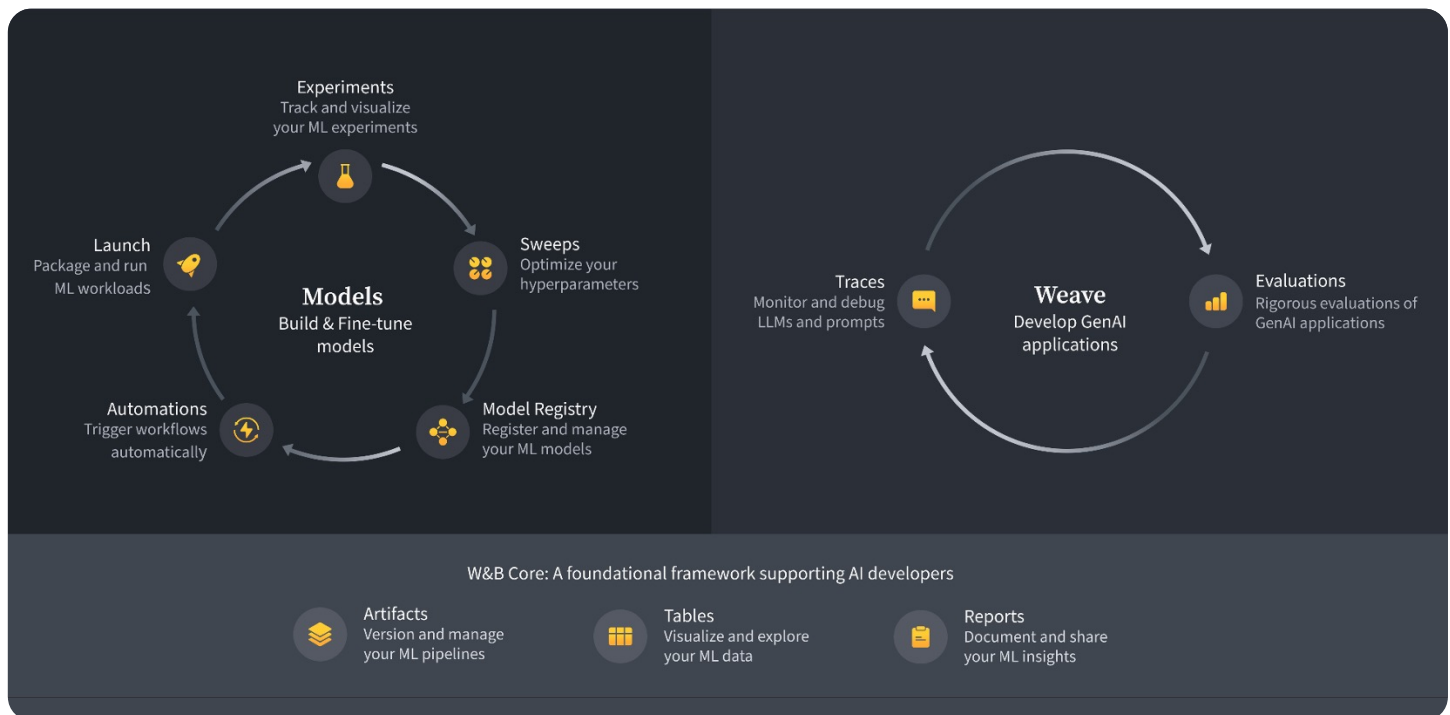


Amazon Elastic Compute Cloud (Amazon EC2)

Get on-demand, scalable computing capacity in the AWS Cloud.

[Learn more ›](#)

Weights & Biases: an AI developer platform



Accelerate LLM-powered app development with W&B Weave on AWS.
Get started developing generative AI applications with confidence.

About Weights & Biases

Weights & Biases is an AI developer platform supporting end-to-end MLOps and LLMOps workflows. Used by over 30 foundation model builders and 1,000 companies to productionize machine learning at scale including teams at OpenAI and Toyota. Weights & Biases is part of the new standard of best practices for machine learning.