watsonx.ai
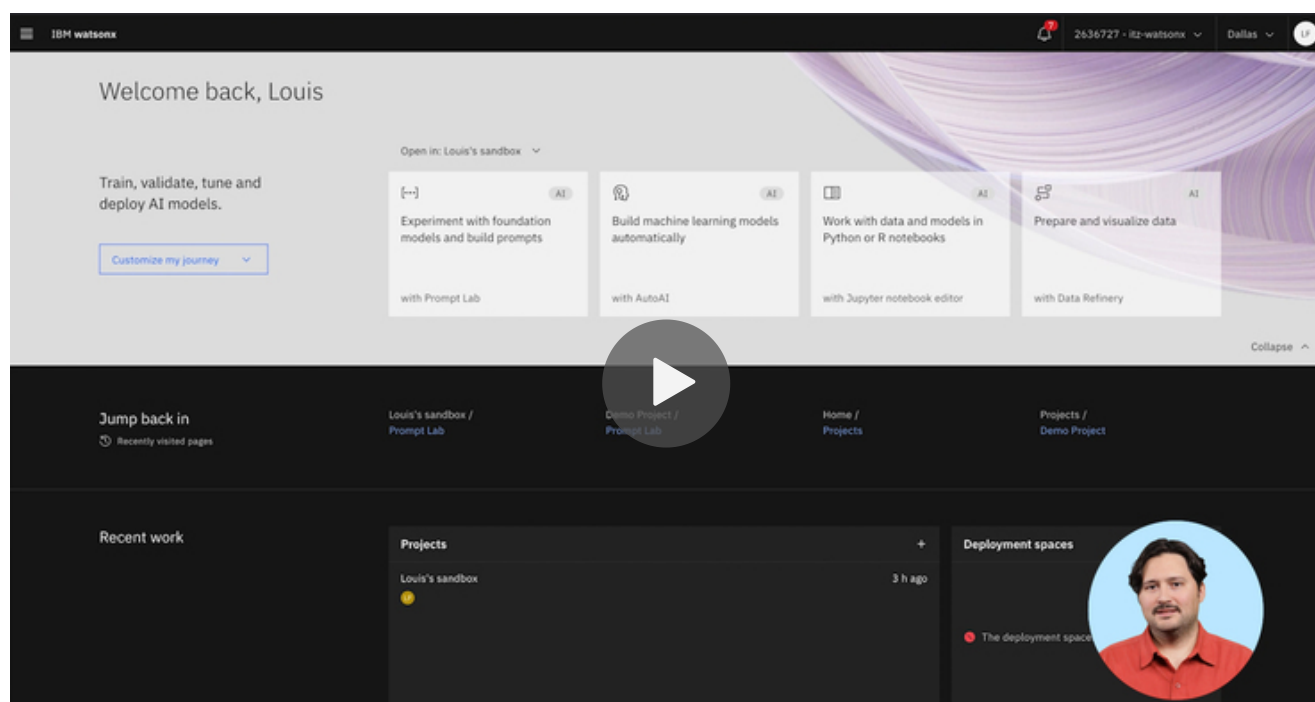
Now available—a next generation enterprise studio for AI builders to train, validate, tune and deploy AI models

Start your free trial →	Explore the watsonx.ai demo →

Our next-generation enterprise studio for AI builders

IBM® watsonx.ai™ AI studio is part of the IBM watsonx™ AI and data platform, bringing together new generative AI (gen AI) capabilities powered by foundation models and traditional machine learning (ML) into a powerful studio spanning the AI lifecycle. Tune and guide models with your enterprise data to meet your needs with easy-to-use tools for building and refining performant prompts.
With watsonx.ai, you can build AI applications in a fraction of the time and with a fraction of the data. Watsonx.ai offers:

- **Multi-model variety and flexibility**: Choose from IBM-developed, open-source and third-party models, or bring your own model.
- **Differentiated client protection**: IBM stands behind IBM-developed models and indemnifies the client against third-party IP claims.
- **End-to-end AI governance**: Enterprises can scale and accelerate the impact of AI with trusted data across the business, using data wherever it resides.
- **Hybrid, multi-cloud deployments**: IBM provides the flexibility to integrate and deploy your AI workloads into your hybrid-cloud stack of choice.

Get the ebook: Generative AI and ML for the enterprise   →

What's new

IBM Offers Meta's Llama 3 Open Models on Watsonx

Read the press release →

Chat directly with an LLM in the watsonx.ai 30-day demo. This demo does not include agents or simultaneous chat with multiple models.

→

IDC Spotlight: The truth about successful generative AI

→

Use cases

Bring your own custom foundation models or work with a suite of curated foundation models offered by IBM. Experiment with open source models through the IBM and Hugging Face partnership to meet the needs of your business.

Knowledge management ⌃

Collect, create and share knowledge

Build a Q&A resource from a broad internal or external knowledge base across your company's proprietary data to share organizational insights. With the help of generative AI, you can analyze multiple documents and data inputs, provide effective responses based on real-time information feeds and improve documentation quality.

Learn more →
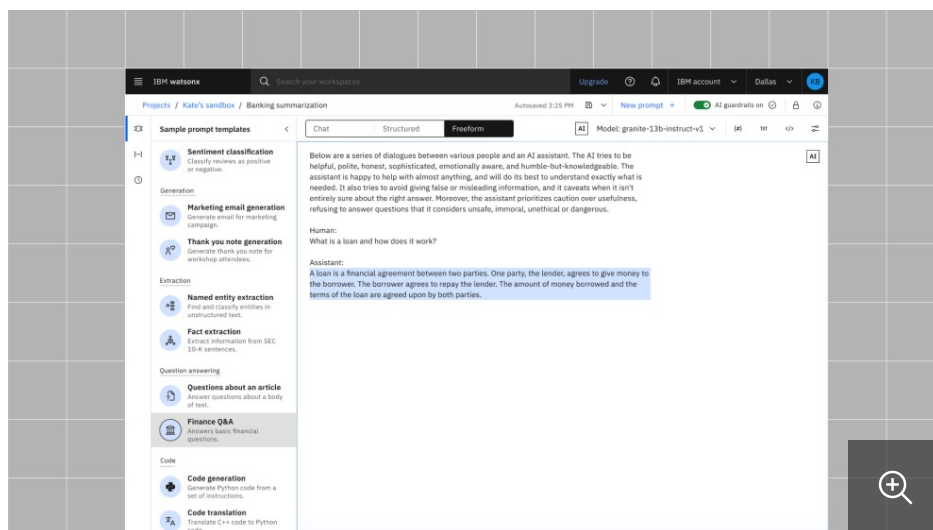
Insight extraction and AI forecasting ⌄

Synthetic data generation ⌄

Content generation ⌄

Benefits
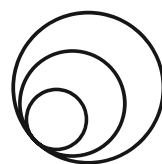
[Discover how to scale AI](#) 📄PDF

Bring together AI builders

Use open-source frameworks and tools for code-based, automated and visual data science capabilities–all in a secure, trusted studio environment.

Tune foundation models for your business

Leverage foundation models and generative AI with minimal data, advanced prompt-tuning capabilities, full SDK and API libraries.
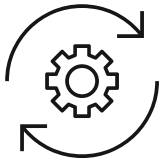
Manage the full AI lifecycle

Accelerate the full AI model lifecycle with all the tools and runtimes in one place to train, validate, tune and deploy AI models across clouds and on-premises environments.

Features

Foundation models

Clients can bring in IBM-supported custom foundation models from Hugging Face and have access to other third-party models including Llama 3 and Mixtral 8x7b. They also have access to a family of IBM-developed foundation models of different sizes and architectures, including open-source Granite models and IBM customized Granite models to support enterprise domains and use cases such as RAG.

See foundation models in watsonx.ai →

Explore how to choose the right foundation model →
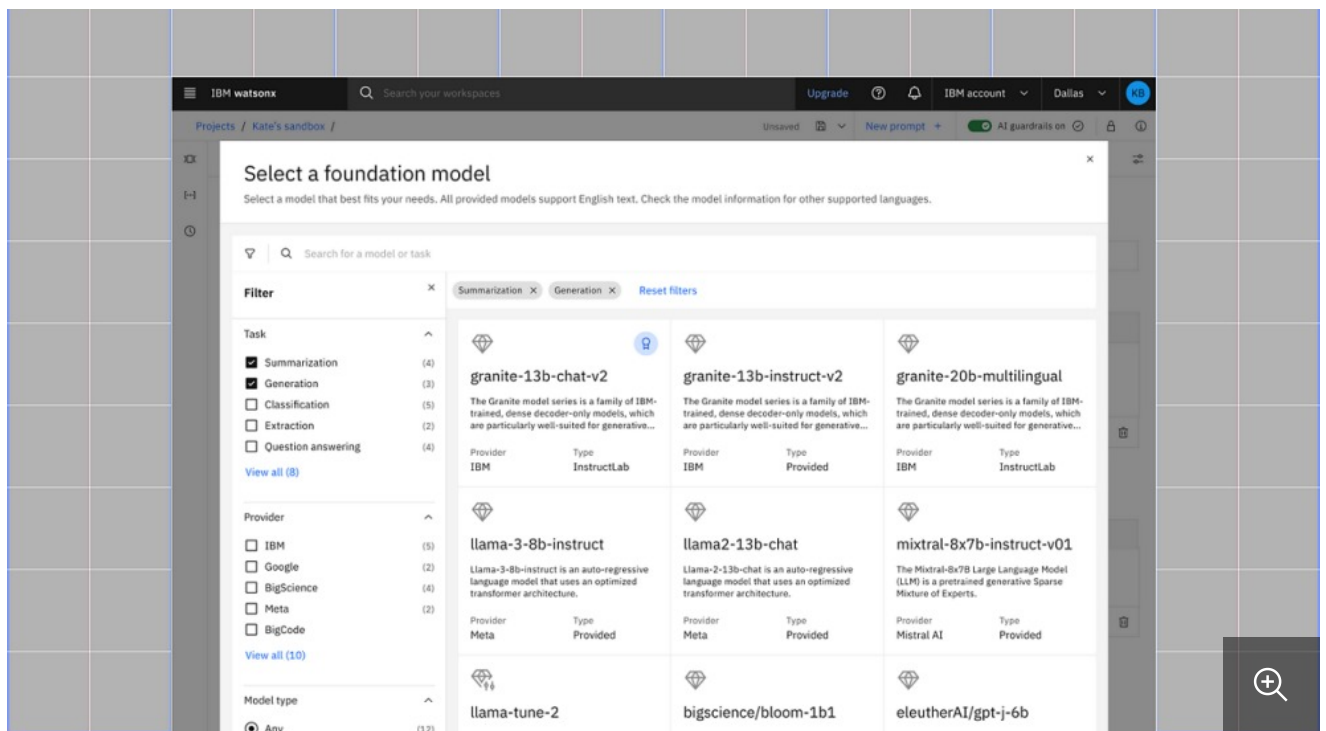
Prompt Lab

Where AI builders can ==work with foundation models and build prompts using prompt engineering==. Within the Prompt Lab, users can interact with foundation models in the prompt editor using the Chat, Freeform or Structured mode. These multiple options will allow you to craft the best model configurations to support a range of Natural Language Processing (NLP) type tasks including question answering, content generation and summarization, text classification and extraction.

Tuning Studio

<mark>Prompt-tune foundation models as part of our Tuning Studio to help tune your foundation models with labeled data for better performance and accuracy</mark>. Prompt-tuning is an efficient, low-cost way of adapting a foundation model to new downstream tasks without retraining the model and updating its weights. Once your model is tuned, you can start using it in the Prompt Lab. Subsequent releases will include capabilities for model fine-tuning as part of our Tuning Studio.

Public preview coming soon

Flows engine

With its intuitive flow language, watsonx.ai flows engine simplifies the development process, allowing generative <mark>AI applications to be effortlessly deployed to the cloud, accessed via API calls or SDKs, and integrated into existing projects and architectures.</mark> The feature provides a range of pre-built AI flows for tasks such as Retrieval Augmented Generation (RAG) and Summarization, allowing developers to embed AI into their business processes seamlessly.

Learn More about flows engine →

Data science and MLOps

All the tools, pipelines and runtimes powered by foundation models, a data scientist would need to support building ML models automatically. <mark>Automate the entire AI model lifecycle from development to deployment with connections to a variety of APIs,</mark>

🔍

Gen AI capabilities

In the Prompt Lab, leverage foundation models to create better AI, faster. Experiment with different prompts for various use cases and tasks. With just a few lines of instruction you can draft job descriptions, classify customer complaints, summarize complex regulatory documents, extract key business information and much more. Quickly tune models for your specific business needs using the latest open source and IBM trained foundation models.

Start your free trial →

Generate ⌃

Generate content, no code required

From writing marketing emails to creating customer personas, watsonx.ai is your new creative partner. Just specify what you want, set the parameters and let the AI do the work.

Classify	⌄
Summarize	⌄
Extract	⌄
Question answering	⌄

MLOps capabilities

<mark>Build models either visually or with code, deploy and monitor with end-to-end lifecycle explainability and fairness.</mark> Use MLOps to simplify model production from any tool and provide automatic model retraining.

Start your free trial →

Orchestration pipelines	⌃

Create automated pipelines

<mark>Single collaborative studio for data scientists to build, train and deploy ML models.</mark> It supports a wide range of data sources enabling teams to streamline their workflows. With advanced features like automated ML and model monitoring, users can manage their models throughout the development and deployment lifecycle.

CPLEX optimization engines	⌄
Visual modeling	⌄
Automated development	⌄
Synthetic data generator	⌄

Partner with us to deliver enhanced commercial solutions, embedded with watsonx.ai, to better address clients' needs.

Explore watsonx technology partners →

Client stories

Businesses are excited about the prospect of tapping foundation models and ML in one place, with their own data, to accelerate generative AI workloads.

Explore more client stories →

AddAI.Life

"Watsonx.ai proved to be very useful. In our research, we liked how it helped our customers (and our development team) to simplify tasks and extend the assistant knowledge without the need to pre-

Minijob-Zentrale

"With this tool, we will need only a quarter of the time compared to before to plan, write and publish an article." — Madeline Scholz

set the whole dialog in advance. It is a next level for us and our customers." — Jindrich Chromy, CEO and co-founder

→

→

Third-party reviews

What IBM watsonx.ai customers are saying on

IBM Watsonx.ai: A Powerful AI Platform for Businesses of All Sizes

★★★⯪☆   Aug 8, 2023

IBM watsonx.ai helped me to solve natural language processing models and machine learning. I really found out that is a versatile platform wide range of business problems. The good thing about

Unlocking AI with IMB Watson AI review

★★★★★   Feb 15, 2024

Its Amazing ML and predictive Analytics and easy integration with other system. Also, it has sound NLP

this is highly scalable and can be...

Siva S.

Automotive Small-Business (50 or fewer emp.)

Amar H.
Project Management- (Mobile Game & Applications Development, Market Research + Data Analytics) | Client Servicing | RFQ Mgt | Customer Relationship Management | Financial Transactions Processing
HFG Entertainments Private Limited
Computer Games Small-Business (50 or fewer emp.)

Read all reviews ⤴

◀ 1 / 5 ▶

Take the next step

Take the next step to start operationalizing and scaling generative AI and ML for business.

Start your free trial →

Explore the interactive demo →

More ways to explore

Become an IBM Business Partner →

Connect with the IBM Community →

Explore AI consulting with IBM →

Explore the interactive demo →

Generative AI learning series →

REST API documentation →

SaaS documentation → See product updates → Software
                                                documentation   →

Subscribe to AI topic → Support   →
updates

Footnotes

[1]IBM's statements regarding its plans, directions, and intent are subject to change or withdrawal without notice at IBM's sole discretion. See Pricing for more detail. Unless otherwise specified under Software pricing, all features, capabilities, and potential updates refer exclusively to SaaS. IBM makes no representation that SaaS and software features and capabilities will be the same.

IBM

United States — English 🌐

About IBM

Overview

Annual report

Corporate social responsibility

Diversity & inclusion

Financing

Investor

Newsroom

Security, privacy & trust

Senior leadership

Careers with IBM

IBM Research

Topics

Artificial intelligence

Machine learning

Conversational AI

AI governance

CSRD

Cybersecurity

Predictive analytics

Quantum computing

Partners

Our strategic partners

Find a partner

Website

Blog

Publications

Collaborate with us

Become a partner - Partner Plus

Partner Plus log in

Engage with IBM

IBM TechXChange Community

LinkedIn

X

Instagram

YouTube

Subscription Center

Participate in user experience research

Contact IBM

Privacy

Terms of use

Accessibility

Cookie Preferences