

Generative AI and ML for the enterprise



Contents



01 →
Introduction

02 →
Generative AI,
foundation models
and ML

03 →
Advantages of
foundation models
and ML

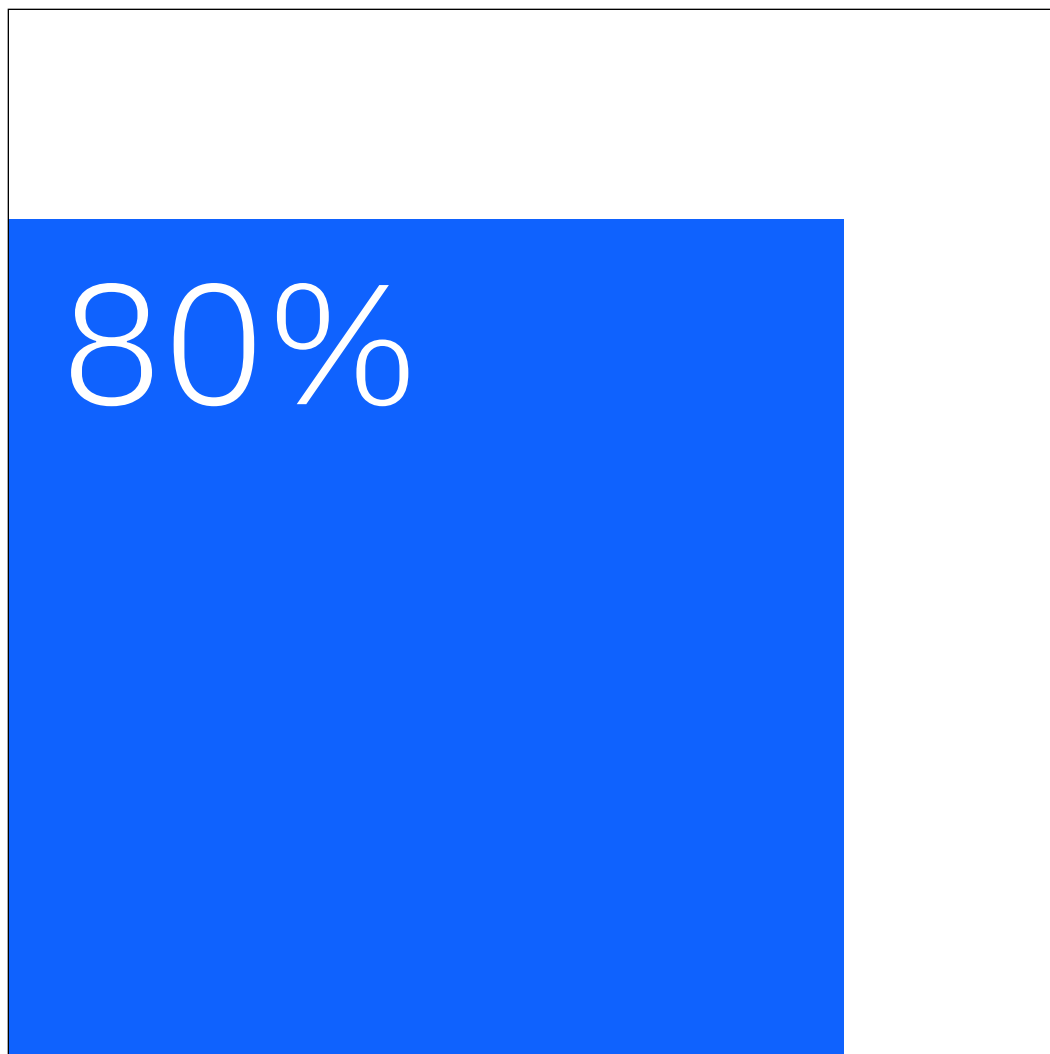
04 →
Common use cases
and tasks

05 →
What to look for
in an AI platform

06 →
IBM watsonx.ai
studio for AI builders

07 →
Next steps

Introduction



Over 80% of enterprises are working with or planning to adopt generative AI.¹

Public adoption of generative AI has shown everyone how transformative the technology can be.

It holds the potential to create competitive advantage by driving the speed and efficiency of operations ranging from HR and talent to application modernization and customer care—and will likely transform entire industries.

By adopting AI that's fine-tuned to your organization's specific needs, you can achieve productivity gains, enhance outcomes and fast-track innovation at scale.

But it's only possible if the models are trained in your industry, your organization and its specific use cases. This process is no easy task.

While over 80% of enterprises are working with or planning to adopt generative AI,¹ concerns about data privacy, trust, security and skills need to be addressed to achieve ROI.

Common AI adoption challenges include:

- Choosing the right use case
- Integrating data and AI across any cloud
- Accessing the right skills and open-source tools or platforms to develop models
- Reducing bias, ensuring security and compliance

“We are especially interested in leveraging the summarization features in watsonx.ai to greatly improve service delivery.”

David Tan
CTO
CrushBank

The bottom line is there's just too much data and complexity to treat AI projects in isolation. The success of AI adoption at scale hinges on alignment with a mature data strategy and integration with existing AI and machine learning (ML) initiatives.

This ebook will help you understand how you can sensibly incorporate generative AI into your existing environment, diving deep into ways you can combine it with traditional ML for greater performance.

Why do leaders want to adopt generative AI?

- Make better and faster decisions by applying generative AI to their data
- Use AI to automate decision-making
- Democratize AI for all AI value creators and builders—regardless of skill-level or role
- Amplify growth by automating manual tasks, uncover hidden patterns in their data, and create new business lines, products and services
- Support the next-generation of the workforce with data-driven insights
- Enhance customer experiences with more hyper-personalized interactions

USTA partnered with IBM to add AI-generated commentary—built with watsonx—to match highlights in the US Open app.

Generative AI, foundation models and ML



Generative AI is a collection of algorithms that produce seemingly new content by interpreting and manipulating preexisting data. This type of AI can generate original images, text, music, speech and even code.

While not a new concept, the latest AI approach is based on a neural network architecture, coined a “transformer.” The most powerful algorithms are those that employ foundation models—a new form of generative AI that combines transformer architecture with unsupervised learning, a ML function.

Foundation models are large AI neural networks trained on extensive unlabeled data and fine-tuned for a variety of functions. These models undergo self-supervised training to uncover inherent patterns that can be applied across a broad spectrum of tasks—and, for reasons we’ll explore, they’re key to developing more advanced and complex models tailored to your specific business needs.

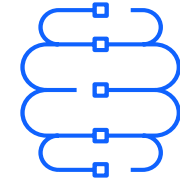
Powered by foundation models, generative AI enables semantic search, code generation, email routing, customer service and improved automation for businesses everywhere.

Using newer AI techniques like foundation models alongside traditional ML techniques can position your organization to get the most from its data.

ML techniques are great for performing mathematical and statistical equations, including making predictions and forecasts, and detecting patterns, trends and anomalies.

Generative AI can even be used to generate more data, known as synthetic data, that can be fed back to traditional ML models to make them more accurate. In the following chapters, we'll look at the benefits of combining generative AI with traditional ML.





Traditional AI and analytics can be augmented with generative AI

When to use traditional AI capabilities

Predictive and prescriptive

Structured data analysis, predictions and forecasting

Directed conversational AI

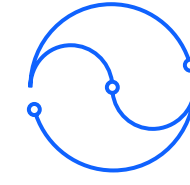
Deterministic dialog flows for AI-driven conversational AI

Computer vision AI

Machine vision for object and anomaly detection

Process automation

Robotic process automation, process reengineering and optimization



When to use generative AI capabilities

Summarization

Documents, such as user manuals, asset notes and financial reports

Conversational search

Standard operating procedures and troubleshooting instructions

Content creation

Personas, user stories, synthetic data, generating images, personalized UI, marketing copy, email and social responses

Code creation

Code creation and conversion, technical documentation and test cases

Most use cases with text, image, video and code generation are good candidates for generative AI. On the other hand, traditional AI tends to be better for most structured data analysis, prediction and prescription use cases. And finally, generative AI can augment traditional AI use cases to enhance natural language interactions and summarization.

“Seismic, also an early adopter of generative AI capabilities, is excited to explore [the ability of watsonx.ai studio] to improve content discoverability and summarize and classify data to enhance productivity.”

Krish Mantripragada
Chief Product Officer
Seismic

Advantages of foundation models and ML



Powered by foundation models trained on the full breadth of enterprise data, generative AI has the potential to unlock trillions in economic value over the next decade. And it can improve the speed and efficiency of existing ML efforts.

For example, generative AI could help you quickly build and deploy traditional ML models for demand forecasting, to predict and solve various business problems—such as supply chain outages or sales patterns—and act on new opportunities, such as building personalized recommendation engines.

Let's look at the reasons why foundation models have established a new paradigm for AI capabilities →

3 ways foundation models add value

1

Get more accurate results tailored to your business domain

Earlier neural networks were narrowly tuned for individual tasks. They required a very specific data set with structured, labeled data to go into a very specific AI model. Now multiply that by the number of tasks you'd like to delegate to AI and the number of data sets it would take to get there.

Foundation models, however, can be adapted to accomplish a broad range of tasks and tailor outputs to your organization's unique data and domain knowledge with a specificity that was previously impossible.

You can conduct various types of tasks based on how you prompt. And with a little fine-tuning, foundation models can handle jobs that range from translating text to analyzing medical images—and with better accuracy.

2

Save time and money on training

The substantial amount of data and processing power it normally takes to train an AI model can easily exceed the ROI potential of typical business applications—but not foundation models.

They're trained on massive unlabeled data sets with unsupervised learning, which cuts down the time and expense significantly. In fact, foundation models can make AI more scalable by reducing the cost and effort of model training by up to 70%.²

3

Bring traditional AI into production faster

With the right platform, you can use data science tools and ML techniques to simplify model production and automate model retraining to make informed, insightful business decisions, including:

Model training and development

- Build experiments quickly and enhance training by optimizing pipelines and identifying the right combination of data.
- Use predictions to optimize decisions, create and edit models in Python, open programming language (OPL) or with natural language.

Integrated visual modeling

- Prepare data quickly and develop models visually.
- Uncover correlations.
- Validate insights for hypotheses.
- Find relationships and connections within the data.

“Watsonx.ai allows us to slash delivery time from 3–4 months down to 3–4 weeks for many customer care use cases.”

Marc Martina

Owner, President & CEO
Technology Dynamics, Inc.

Common use cases and tasks

Curious about where to start? Here are some of the proven use cases and common tasks where generative AI, powered by foundational models, can be combined with traditional ML to improve performance and the accuracy of AI outputs—especially when data and resources are limited.

Content creation

Generative AI can assist in generating creative content, such as product designs or marketing materials, while traditional ML algorithms can analyze customer engagement and feedback data to optimize the generated content for better performance.

Personalization

Traditional ML algorithms can analyze customer behavior and preferences to provide personalized recommendations. Generative AI models can further enhance personalization by generating bespoke content, such as product descriptions or creative marketing materials, tailored to each customer's interests.



Anomaly detection

Traditional ML models can be used to build anomaly detection systems to identify unusual patterns or behaviors in data. Generative AI models can be used to learn the normal data distribution and generate synthetic data samples. The generated samples can then be compared to real data to identify anomalies and make predictions more effectively. Potential business applications could include network fraud detection or credit card fraud detection.

Forecasting

Traditional ML techniques can be used to analyze historical data and find trends hidden in your data—including forecasting customer churn, the likelihood of default loans, sales cycles and volumes, then generative AI can be utilized to summarize insights found, generate customer emails and personalize recommendations in marketing outbound efforts.

Data augmentation with synthetic data

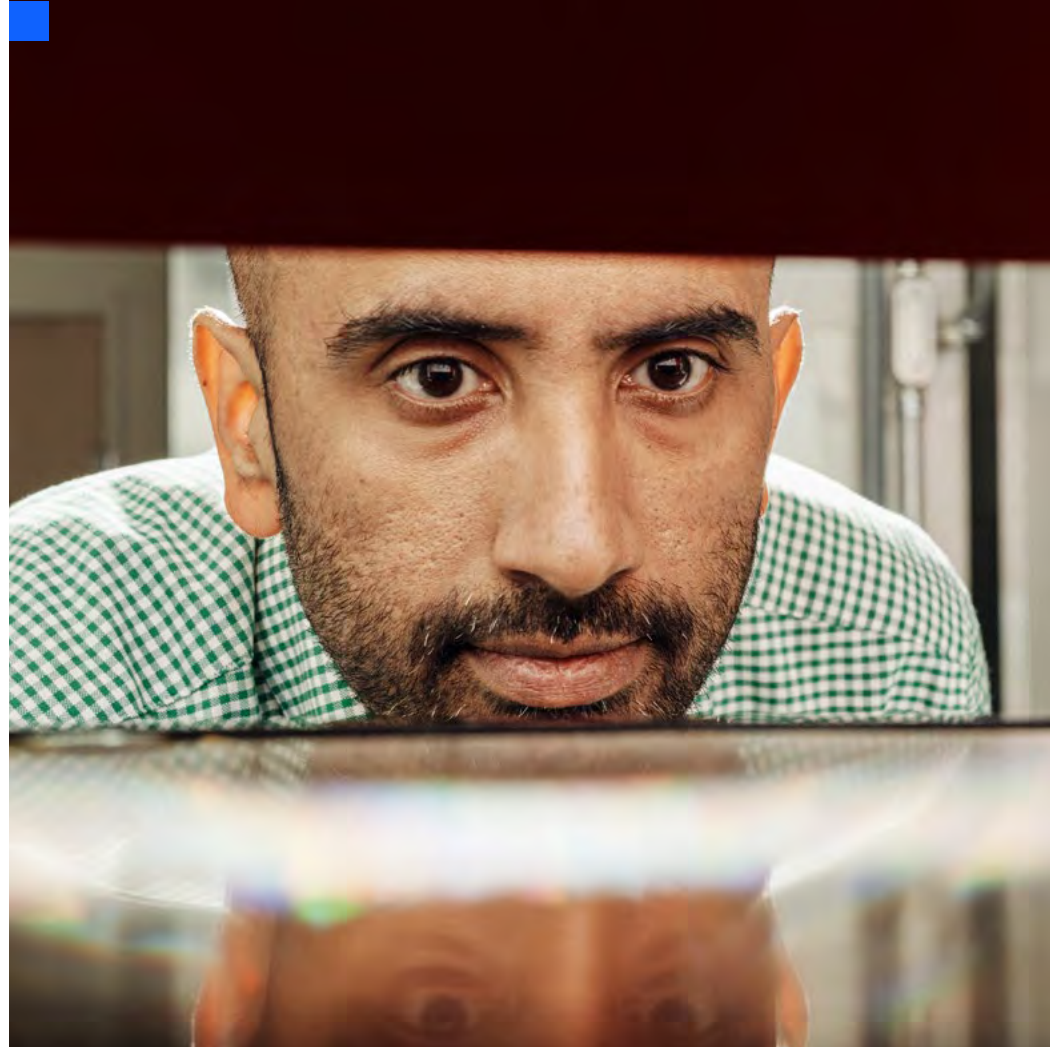
Generative AI can be used to create synthetic tabular data that's similar to the real data, which can then be combined with the existing data set to enhance the performance of traditional ML models. This feature is useful when the real data is limited or imbalanced. Potential applications include use in medical research, such as detecting rare medical conditions, or object detection in autonomous vehicles.



“IBM’s launch of watsonx was an awakening, and it has inspired us to explore the immense potential [of the generative AI capabilities of watsonx.ai] to deliver unprecedented innovations for our clients.”

Sean Im
President & CEO
Samsung SDS America

What to look for in an AI platform



To be most impactful, AI should integrate into existing workflows and systems, automating and augmenting key processes. Selecting the right AI platform is key to making this integration happen and capturing the true potential of scaled AI for business.

To help you make the most from your data, an enterprise AI platform should combine the most important aspects of AI—predictive, classification and ML—for a holistic approach. The ideal solution is backed with four key elements, which we'll explore on the next page.



If you're considering generative AI, your organization will need to comply with legal and regulatory obligations, which are still in flux. The right solution can help you do it.

Four key elements of the ideal enterprise AI platform

1. Open

When it comes to selecting an AI model, it's best to have a variety of sizes and architectures so you can pick what's best suited to your needs. Your data and AI platform should provide the ability to build, train, validate tune and deploy AI using both curated open-source models and private foundation models that can bring in business domain context.

2. Trusted

Before deploying a model, you need to ensure that it delivers reliable results and helps keep the data—and your intellectual property (IP)—secure. AI should be explainable, fair, robust and transparent. All models should offer transparency into the data's origin, bias and governance. The solution should track data, curating methods and models, enabling AI that can be updated to meet evolving business and regulatory requirements.

3. Targeted

Consumer AI isn't the same as enterprise AI. Models should be highly curated to business domains and use cases that drive quick gains in productivity and time to value—such as augmenting and automating HR, customer service and code generation.

4. Empowering

Your generative AI and ML platform should empower you to be an AI value creator, not just a user. You shouldn't be limited to just prompting someone else's AI model with no control over the model or the data. Regardless of if you're a data scientist or a business user, the platform should allow you to build, train, tune and deploy, and govern the data and AI models you're using.

To realize maximum business value from AI, it's crucial for organizations to provide their data scientists, developers and business analysts access to high-quality, trustworthy data, anywhere—with a unified AI and data platform for building and refining AI models at a pace to drive rapid business innovation.

■ Watsonx.ai studio is helping companies custom build AI solutions to suit their specific needs.

Finally, let's acknowledge the concerns about generative AI adoption. Information accuracy, security, privacy, ethics and IP have emerged as key causes for concern regarding AI. That's quite a list. If you're considering generative AI, then you're keenly aware that you'll need to demonstrate your organization's effort to comply with legal and regulatory obligations—which are themselves in flux right now. The right solution can help you do it. That's where IBM comes in.



IBM watsonx.ai studio for AI builders

IBM® watsonx.ai™ is part of the watsonx™ AI and data platform, with a set of AI assistants designed to help you scale and accelerate the impact of AI with trusted data across your business.

The core components include: a studio for new foundation models, generative AI and ML; a fit-for-purpose data store built on an open data lakehouse architecture; and a toolkit to accelerate AI workflows that are built with responsibility, transparency and explainability.

What is watsonx.ai?

Watsonx.ai is an enterprise-ready next-generation studio for ML and generative AI, powered by foundation models spanning the AI lifecycle.

With the watsonx.ai studio, AI builders can train, validate, tune and deploy traditional ML and new generative AI capabilities with ease and build AI applications in a fraction of the time with a fraction of the data.

These models combine best-in-class architectures with a rigorous focus on data acquisition, provenance and quality to serve enterprise needs.

IBM's open, hybrid, full-stack approach includes:

- Foundation model library with the IBM® Granite™ family of models built for business and select open-source models from Hugging Face
- Prompt Lab to experiment with foundation models and build prompts for various use cases and tasks
- Tuning Studio to support prompt tuning and tailor models using enterprise data
- Data science and MLOps to build ML models automatically with model training, development, visual modeling and synthetic data generation

Next steps

You can immediately get started with the watsonx.ai studio for free with our trial experience.

Your organization can co-create with IBM Consulting™ and the 21,000 data, AI and automation consultants, in addition to the Center of Excellence for Generative AI comprised of more than 1,000 consultants with specialized generative AI expertise. These experts can work with you to help tune and operationalize models for targeted use cases aligned to your specific business requirements.



[Start your free trial](#) →

[Explore watsonx.ai](#) →



1. Linthwaite, Rachel. “Overcome Obstacles To Get To AI At Scale.” Forrester, January 2020
2. What is generative AI, what are foundation models, and why do they matter? IBM blog, March 2023. <https://www.ibm.com/blog/what-is-generative-ai-what-are-foundation-models-and-why-do-they-matter/>

© Copyright IBM Corporation 2024

IBM Corporation
New Orchard Road
Armonk, NY 10504

Produced in the United States of America
July 2024

IBM, the IBM logo, IBM Consulting, watsonx, and watsonx.ai are trademarks or registered trademarks of International Business Machines Corporation, in the United States and/or other countries. Other product and service names might be trademarks of IBM or other companies. A current list of IBM trademarks is available on [ibm.com/trademark](https://www.ibm.com/trademark).

This document is current as of the initial date of publication and may be changed by IBM at any time. Not all offerings are available in every country in which IBM operates.

It is the user’s responsibility to evaluate and verify the operation of any other products or programs with IBM products and programs.

It is the user’s responsibility to evaluate and verify the operation of any other products or programs with IBM products and programs. THE INFORMATION IN THIS DOCUMENT IS PROVIDED “AS IS” WITHOUT ANY WARRANTY, EXPRESS OR IMPLIED, INCLUDING WITHOUT ANY WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND ANY WARRANTY OR CONDITION OF NON-INFRINGEMENT. IBM products are warranted according to the terms and conditions of the agreements under which they are provided.

Statement of Good Security Practices: No IT system or product should be considered completely secure, and no single product, service or security measure can be completely effective in preventing improper use or access. IBM does not warrant that any systems, products or services are immune from, or will make your enterprise immune from, the malicious or illegal conduct of any party.

The client is responsible for ensuring compliance with all applicable laws and regulations. IBM does not provide legal advice nor represent or warrant that its services or products will ensure that the client is compliant with any law or regulation.