IBM Unveils the Watsonx Platform to Power Next-Generation Foundation Models for Business

- Watsonx is a new platform to be released for foundation models and generative AI, offering a studio, data store, and governance toolkit

- New Watson products infused with foundation models and generative AI to be launched for code, AIOps, digital labor, security, and sustainability

- New collaboration with Hugging Face will work to bring the best of open-source AI models to the enterprise on the watsonx platform

- IBM Consulting announces a Center of Excellence for generative AI with over 1000 AI experts ready to implement clients' business transformation with enterprise-grade AI

ARMONK, N.Y., May 9, 2023 /PRNewswire/ -- Today, IBM (NYSE: IBM) at its annual Think conference, announced IBM watsonx, a new AI and data platform to be released that will enable enterprises to scale and accelerate the impact of the most advanced AI with trusted data. Enterprises turning to AI today need access to a full technology stack that enables them to train, tune and deploy AI models, including foundation models and machine learning capabilities, across their organization with trusted data, speed, and governance - all in one place and to run across any cloud environment.

Also at Think, IBM is announcing further planned advancements, including a GPU-as-a-service infrastructure offering designed to support AI-intensive workloads, an AI-powered dashboard to measure, track, manage, and help report on cloud carbon emissions, and a new practice for watsonx and generative AI from IBM Consulting that will support client deployment of AI.

With watsonx, IBM is offering an AI development studio with access to IBM-curated and trained foundation models and open-source models, access to a data store to enable the gathering and cleansing of training and tuning data, and a toolkit for governance of AI into the hands of businesses that will provide a seamless end-to-end AI workflow that will make AI easier to adapt and scale.

"With the development of foundation models, AI for business is more powerful than ever," said Arvind Krishna, IBM Chairman and CEO. "Foundation models make deploying AI significantly more scalable, affordable, and efficient. We built IBM watsonx for the needs of enterprises, so that clients can be more than just users, they can become AI advantaged. With IBM watsonx, clients can quickly train and deploy custom AI capabilities across their entire business, all while retaining full control of their data."

Clients will have access to the toolset, technology, infrastructure, and consulting expertise to build their own — or fine-tune and adapt available AI models — on their own data and deploy them at scale in a more trustworthy and open environment to drive business success. Competitive differentiation and unique business value will be able to be increasingly derived from how adaptable an AI model can be to an enterprise's unique data and domain knowledge.

The IBM watsonx platform consists of three unique product sets to address these needs:

IBM watsonx.ai: A next generation enterprise studio, expected to be generally available in July 2023, for AI builders to train, test, tune, and deploy both traditional machine learning and new generative AI capabilities powered by foundation models through an open and intuitive user interface.

- The AI studio provides a range of foundation models, training and tuning tools, and cost-effective infrastructure that facilitate the entire data and AI lifecycle, from data preparation to model development, deployment, and monitoring.
- The studio also includes a foundation model library that gives users easy access to IBM curated and trained foundation models. The IBM foundation models use a large, curated set of enterprise data backed by a robust filtering and cleansing process and auditable data lineage. These models are being trained not just on language, but on a variety of modalities, including code, time-series data, tabular data, geospatial data, and IT events data. An initial set of foundation models will be made available in beta tech preview to select clients. Examples of model categories include:
    - **fm.code**: Models built to automatically generate code for developers through a natural-language interface to boost developer productivity and enable the automation of many IT tasks.
    - **fm.NLP**: A collection of large language models (LLMs) for specific or industry-specific domains that utilize curated data where bias can be mitigated more easily and can be quickly customized using client data.
    - **fm.geospatial:** Model built on climate and remote sensing data to help organizations understand and plan for changes in natural disaster patterns, biodiversity, land use, and other geophysical processes that could impact their businesses.

- As part of a new collaboration between IBM and Hugging Face, the watsonx.ai studio will build upon Hugging Face's open-source libraries and offer thousands of Hugging Face open models and datasets. This is part of IBM's commitment to delivering to clients an open ecosystem approach that allows them to leverage the best models and architecture for their unique business needs.

IBM watsonx.data: A fit-for-purpose data store built on open lakehouse architecture that is optimized for governed data and AI workloads, supported by querying, governance, and open data formats to access and share data. The solution is expected to be generally available in July 2023 and:

- The solution can manage workloads both on-premise and across multi-cloud environments.
- Through workload optimization, with this solution, an organization can reduce data warehouse costs by up to 50 percent.[1]
- Watsonx.data will allow users to access their increasingly robust data through a single point of entry while applying multiple fit-for-purpose query engines to uncover valuable insights.
- It will also provide built-in governance tools, automation and integrations with an organization's existing databases and tools to simplify set-up and user experience.

IBM watsonx.governance: An AI governance toolkit to enable trusted AI workflows. The solution, expected to be generally available later this year:

- Operationalizes governance to help mitigate the risk, time and cost associated with manual processes and provides the documentation necessary to drive transparent and explainable outcomes.
- Provides the mechanisms to protect customer privacy, proactively detect model bias and drift, and help organizations meet their ethics standards.

With the watsonx platform, clients are enabled to meet the needs of their organization in five key areas of their business: interacting and conversing with customers and employees; automating business workflows and internal processes; automating IT processes; protecting against threats; and tackling sustainability goals.

IBM also plans to infuse watsonx.ai foundation models throughout all its major software products going forward, for example:

- **Watson Code Assistant**: A solution, expected later this year, that taps generative AI to allow developers to generate code with a straightforward English language command.

- **AIOps Insights**: AI Operations (AIOps) capabilities enhanced with foundation models expected for code and NLP to provide greater visibility into performance across IT environments, helping IT operations (ITOps) managers and Site Reliability Engineers (SREs) resolve incidents in a more expedient and cost-efficient way.

- **Watson Assistant and Watson Orchestrate**: IBM's digital labor products are expected to be combined with an NLP foundation model to enable enhanced employee productivity and customer service experiences.

- **Environmental Intelligence Suite**: IBM EIS Builder Edition is planned to be enabled by the geospatial foundation model and available in preview later this year, allowing organizations to create tailored solutions that address and mitigate environmental risks based on their unique goals and needs.

Also at Think 2023, IBM will announce a number of additional upcoming offerings that are planned to help drive AI adoption, including:

- **New GPU offering on IBM Cloud:** Addressing the global need for foundation models, IBM is announcing new GPU offerings on IBM Cloud, an AI-tailored infrastructure designed to support enterprise compute-intensive workloads. Later this year, IBM is expected to offer full stack high-performance, flexible, AI-optimized infrastructure, delivered as a service on IBM Cloud, for both training and serving foundation models.

- **IBM Consulting Center of Excellence for Generative AI:** IBM Consulting announces a Center of Excellence for generative AI with over 1,000 generative AI experts and plans to build a watsonx-focused practice which will actively build and deploy watsonx for clients. IBM Consulting has completed dozens of client engagements infusing generative AI with IBM Watson and a portfolio of ecosystem partners through its proven IBM Garage method.

- **IBM Cloud Carbon Calculator:** An AI-informed dashboard to enable clients to measure, track, manage and help report their carbon emissions associated with their hybrid multi-cloud journey. Based on technology from IBM Research, IBM Cloud Carbon Calculator is expected to be generally available later this year. The dashboard complements IBM's existing sustainability solutions with a comprehensive portfolio of technology and expertise, including the IBM Envizi ESG Suite, IBM Turbonomic, IBM Planning Analytics and IBM LinuxONE, that help organizations accelerate their sustainability and business objectives.

- **New study revealing generative AI is among seven trends shaping business**: A new report from the IBM Institute for Business Value, titled Seven Bets, shares seven trends impacting business today and describes the seven bets worth making to enhance their business, including insights on why businesses should adopt an "AI-first" mindset and how leaders can most effectively capitalize on AI's opportunities now and in the future, as well as manage the enhanced risks across their organizations.

The Think 2023 broadcast can be accessed at www.ibm.com/events/think. During his keynote airing Tuesday, May 9, 8:30 a.m. - 9:15 a.m. EDT, IBM CEO and Chairman Arvind Krishna will be joined on stage by leaders from Delta Airlines, Citi, and Red Hat to discuss their visions for creatively turning challenges into opportunities with hybrid cloud and AI. On Wednesday, May 10, 4:30 p.m. - 5:00 p.m. EDT, Darío Gil, Senior Vice President and Director of IBM Research, will give a keynote on how generative AI can be applied for business use cases.

Following the flagship event in Orlando, Think on Tour will travel to more than a dozen cities around the world. For updates on confirmed cities, please visit: https://www.ibm.com/events/think/on-tour/.

About IBM
IBM is a leading global hybrid cloud and AI, and business services provider, helping clients in more than 175 countries capitalize on insights from their data, streamline business processes, reduce costs and gain the competitive edge in their industries. Nearly 4,000 government and corporate entities in critical infrastructure areas such as financial services, telecommunications and healthcare rely on IBM's hybrid cloud platform and Red Hat OpenShift to affect their digital transformations quickly, efficiently and securely. IBM's breakthrough innovations in AI, quantum computing, industry-specific cloud solutions and business services deliver open and flexible options to our clients. All of this is backed by IBM's legendary commitment to trust, transparency, responsibility, inclusivity and service. For more information, visit www.ibm.com.

Hugging Face and IBM to partner to help democratize access to foundation models

"To benefit from the latest AI capabilities, enterprises want to build on open-source machine learning: open-source models trained on accessible datasets that can run within a secure environment with compliance and proper data governance," said Clem Delangue, co-founder and CEO, Hugging Face. "We are excited to partner with IBM to deliver a top-tier developer experience based on Hugging Face open-source and community-driven machine learning, within the new watsonx platform offering the enterprise-readiness and trustworthiness of IBM."

Meta's PyTorch team and IBM change one line of code to massively improve AI model training

"Businesses today cannot get the most from their AI and cloud technologies unless these innovations are equipped to succeed at scale," Damien Sereni, Engineering Director, Meta. "Meta has enjoyed a multi-year collaboration with IBM to address this key challenge. And, with the release of IBM watsonx, the PyTorch team at Meta and IBM are collaborating further to build a production-ready software stack for end-to-end training, fine-tuning and inference of large-scale foundation models."

How IBM and NASA's AI foundation model can help track and adapt to climate change

"NASA is committed to the full, free and open sharing of over 100 PB of scientific data and information. This is the first-time foundation models have been applied to the immense archive of Earth science data and we are working with IBM to evaluate the foundation model stack within NASA's Science Managed Cloud Environment," said Kevin Murphy, Chief Science Data Officer at NASA. "Next-generation AI technology can expedite our analysis and discovery of mission-critical insights pertaining to Earth and climate-related issues, enabling scientists to gain insights that can help us protect our planet."

WIX leverages IBM large language models to improve customer care

"Wix is proud to leverage state-of-the-art IBM large language models to uncover novel insights from vast amounts of PII-stripped customer care data and thereby increase our support of more than 243 million clients across the globe," said Yoav Abrahami, Chief Architect at Wix and Head of Velo.

IBM's plans, directions, and intentions may change or be withdrawn at any time at IBM's discretion without notice. Information about potential future products and improvements is provided to give a general idea of IBM's goals and objectives and should not be used in making a purchase decision. IBM is not obligated to provide any material, code, or functionality based on this information.

----

[1] When comparing published 2023 list prices normalized for VPC hours of watsonx.data to several major cloud data warehouse vendors. Savings may vary depending on configurations, workloads and vendors.

Media Contact:
Caitlin O'Neill
caitlin.oneill@IBM.com

SOURCE IBM

Additional assets available online: Photos

https://newsroom.ibm.com/2023-05-09-IBM-Unveils-the-Watsonx-Platform-to-Power-Next-Generation-Foundation-Models-for-Business?mhsrc=ibmsearch_a&mhq=watsonx&_gl=1%2A7evjyf%2A_ga%2AMjAzMDY1NjEzOS4xNjg5MTAwOTkx%2A_ga_FYECCCS21D%2AMTY4OTEwMDk5MS4xLjEuMTY4OTEwMTAxNC4wLjAuN