

The AI Platform with X Factor?

A new AI platform stepped out onto the stage at IBM's technology showcase Think conference back in May: IBM watsonx, the "next-generation AI and data platform to scale and accelerate AI." Well maybe it didn't quite step out, it was more of a preview performance (more on that in a bit), but we got to hear some pretty compelling things about a product that IBM has been preparing for the big stage for some time — months of design, strategy, rehearsals, feedback, fine-tuning, because in July 2023, this unknown performer is making its world debut! So, what do we have to look forward to? Does it have the "X-Factor", that exceptional quality that sets it apart and contributes to its success and appeal? IBM certainly believes it does, so let's dive in.

The preview

OK, so what did we learn about this new superstar from the "preview performance" at Think?

Arvind Krishna introduces it as follows [1]: "We're excited to announce the launch of watsonx, a groundbreaking data & AI platform that offers foundation models and generative AI technology. Clients will have access to a toolset, technology, infrastructure, and consulting expertise to build their own — or fine-tune and adapt available AI models — on their data and deploy them at scale in a more trustworthy and open environment to drive business success."

Let's put a pin in "generative AI" and "scale in a more trustworthy and open environment" as we'll come back to those in a moment.

Importantly, we learned that watsonx is not a solo act. It is, in fact, a trio of products, each with their own distinctive role to play in the band [2].



watsonx.ai — the lead singer with the X Factor in this new AI tech (Gencraft generated)

watsonx.ai

Without a doubt, the lead singer/lead guitarist in the band, is watsonx.ai. It struts its stuff with the self-belief of knowing that AI is no longer the emerging hit, it is mainstream! This is its time to shine! IBM describes it as “a next generation enterprise studio...for AI builders to train, test, tune, and deploy both traditional machine learning and new generative AI capabilities.” It’s tooling for both creating and remixing AI models to implement your business’ AI use cases. So, expect all the AutoML capabilities that streamline and automate model training, testing and deployment, allowing your small data science team to be super productive, and maybe even allowing your data science-savvy “business people” to step over the IT boundary line and produce some models of their own. But note the “generative AI” reference again. This is not last year’s AI platform before ChatGPT took the world by storm. This is an AI platform of its time, including foundational models [3] that power generative AI, and allowing them to be remixed as part of your own AI use cases. This is significant. Training AI models requires significant amounts of data. And creating these large and seemingly all-powerful models also requires a significant amount of raw processing power – not to mention storage. This is well beyond the ROI-reach of your average or even high-end business use cases. Foundational models are pre-trained, ready to “remix.” This is why everyone is so excited about the realistic potential for generative AI to positively transform mainstream business. IBM is very smart to cover this in their platform.



watsonx.data
— the drummer
powering this
AI performance
(Gencraft
generated)

watsonx.data

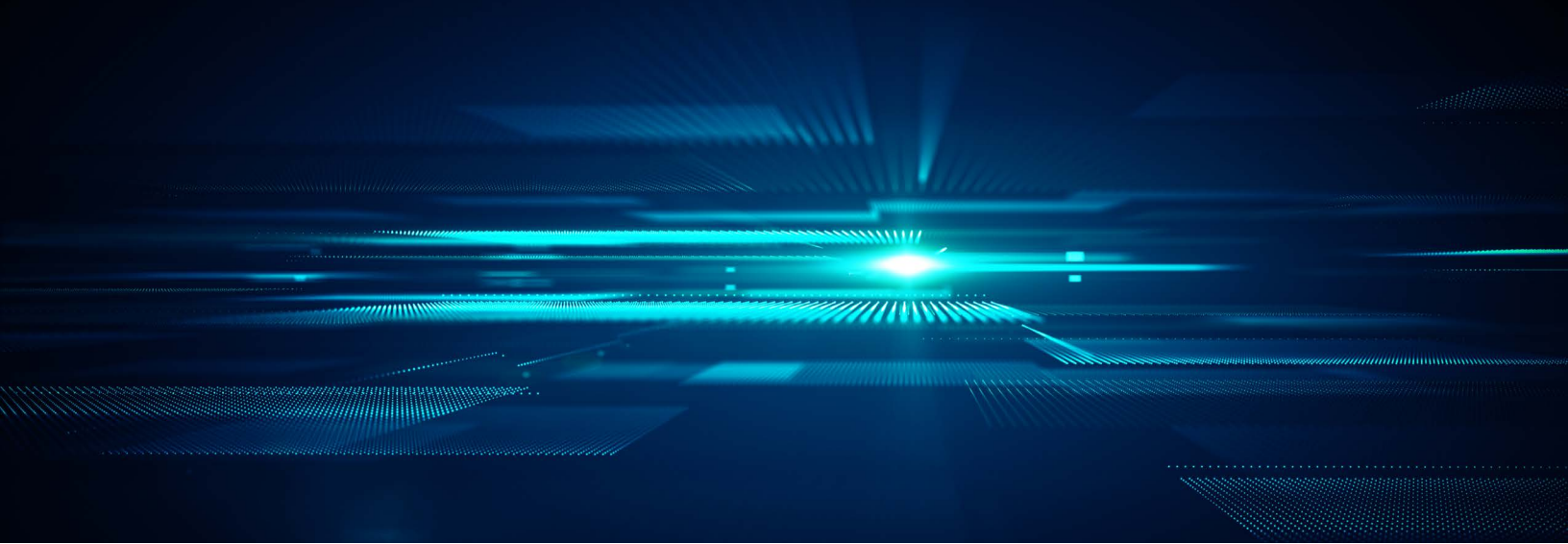
This is the drummer, the bass guitarist, the beat powering the performance. As we all know, data is essential for AI, and this is where that data is brought together to create your AI virtuoso compositions. IBM describes **watsonx.data** as “a fit-for-purpose data store built on open lakehouse architecture that is optimized for governed data and AI workloads, supported by querying, governance, and open data formats to access and share data.” OK, so lakehouses [4] are not a new thing, and some may say that IBM is late to the party here. But maybe they got their timing just right. Let’s look at two big players in this space: Snowflake (a cloud data warehouse rather than a lakehouse), and Databricks (a true lakehouse). Both have seen fairly meteoric rises in popularity due to the fundamental need for a solid data platform that forms the core of any data-driven business, bringing in raw data and turning it into data assets that are able to create and power AI models. So, the need is certainly there. But Snowflake does not claim to provide an AI platform, it just focuses on the data. Databricks can’t claim to have invented the lakehouse concept, although it has certainly massively popularized it over the last few years. Databricks also focuses on “traditional machine learning” rather than foundational models. So possibly the world is ready for a new entrant in the lakehouse space.



watsonx.governance — The control and oversight for the AI “band” (DALL-E generated)

watsonx.governance

Less of a performer role, more of a producer/mixer role. The role of watsonx.governance is to keep the band’s performance in key, on tempo, but also in tune with the target audience. When it comes to AI, audiences need things like ethical decision making, bias and fairness, transparency and explainability, robustness and safety, oversight and control. IBM describes watsonx.governance as “an AI governance toolkit to enable trusted AI workflows.” Hmm. Let’s let IBM expand on that: “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.” This role is about removing unwanted imperfections and giving direction, while taking care of some of the deep technical complexities. More of a behind-the-scenes Brian Eno [5] role rather than a front-and-center Prince. Less is known about this member of the band, and it’s due to be released two months after initial release, but this could really be the critical piece that elevates the whole platform. How important? Think U2 after Brian Eno versus U2 before Brian Eno [6]. Let’s keep in mind how quickly the first optimistic media blushes of ChatGPT were followed by the scary headlines of “AI taking over your job”. And then how quickly did that escalate into “AI resulting in the end of humanity”? AI governance is clearly a need that is coming mainstream at a rapid rate and will surely be an essential component of any AI platform.



IBM — a strong label pedigree

OK, let us put the band/music analogy to one side. What makes IBM a good bet for a new data & AI platform? Well, they've certainly had some good hits (sorry) in this space before.

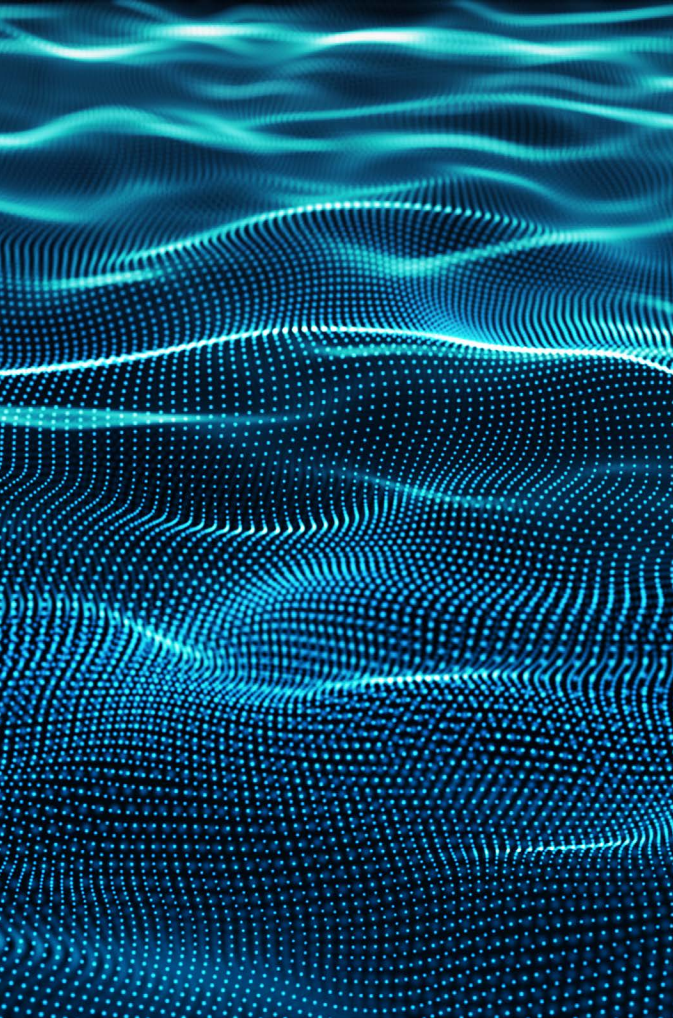
Let's rewind all the way to 2011 and remind ourselves of the story of Watson, the AI that beat the best human Jeopardy player and "showcased the power of artificial intelligence... the beginning of a technological revolution about to sweep through society" [7]. Watson may have been a false dawn for IBM, but it did showcase the value generated from the significant investments made at IBM Research. Fast forward to the new, more confident IBM under Arvind Krishna, and maybe IBM has learned its lessons on how to take its technology pearls to market in a more business-focused way.

Further back, there was a notable period of increased investment in data products by IBM in the early 2000's. During this time, IBM recognized the growing importance of data and analytics in the business world. They saw the potential for businesses to gain insights and make better decisions by leveraging data effectively, investing heavily in developing data-focused products and services. In 2005, IBM launched its "Information on Demand" initiative, aimed at helping businesses harness the power of their data and turn it into valuable insights — data

management, data integration, and analytics capabilities. They further acquired companies like Cognos, SPSS, and Netezza, which enhanced their capabilities in data analytics, business intelligence, and data warehousing. These are all strong products that have made a significant impact on the market.

More recently, IBM has shown it's willing to create rather than acquire with their Cloud Pak for Data platform [8] — a bold step in creating a born-on-the-cloud platform that tied together much of its heritage IP along with new components that are cloud-native. Significantly, it's also portable across on-prem or any cloud with its OpenShift foundation. Also significantly, this has allowed IBM to offer SaaS/PaaS incarnations of their products, a move that is long overdue and more in step with their competitors. A quick note — I understand from IBM that watsonx is not a replacement for CP4D.

So in summary, IBM has had some successes in the data & AI space, and seems to be trending in the right direction in terms of how to package capability and take it to market in a way that suits their functional as well as non-functional needs. So, do they have their timing right?



Scale is important when it comes to AI, and more specifically, data. We all know that training AI requires oodles of data. And this is why you need a lakehouse — bringing together all the data you need for data science. But this is typically the same sort of data you would put in a data warehouse with its more traditional BI/reporting use cases. A lakehouse means you don't need two copies of all this data — warehouse for BI, lake for AI — just put them in a lakehouse which is good enough for meeting both needs. Running both a warehouse and a datalake can result in significant costs. Simple math tells us that replacing with a lakehouse is good for the bank balance! Also, without naming any names, organizations are realizing that their warehouse/datalake/lakehouse costs keep going up as more data moves in. One more interesting point — watsonx.data promises that it won't require you to make a copy of the data in order to have it available on the platform. This brings the 2-in-1 cost benefits of the lakehouse with the further cost reduction of leaving certain data where it is.

In tune with the current hits?

What's the mood music in the room?

- Increasing scale
- Reducing cost
- ChatGPT-style human capabilities
- But avoiding redundancy/extinction-by-AI

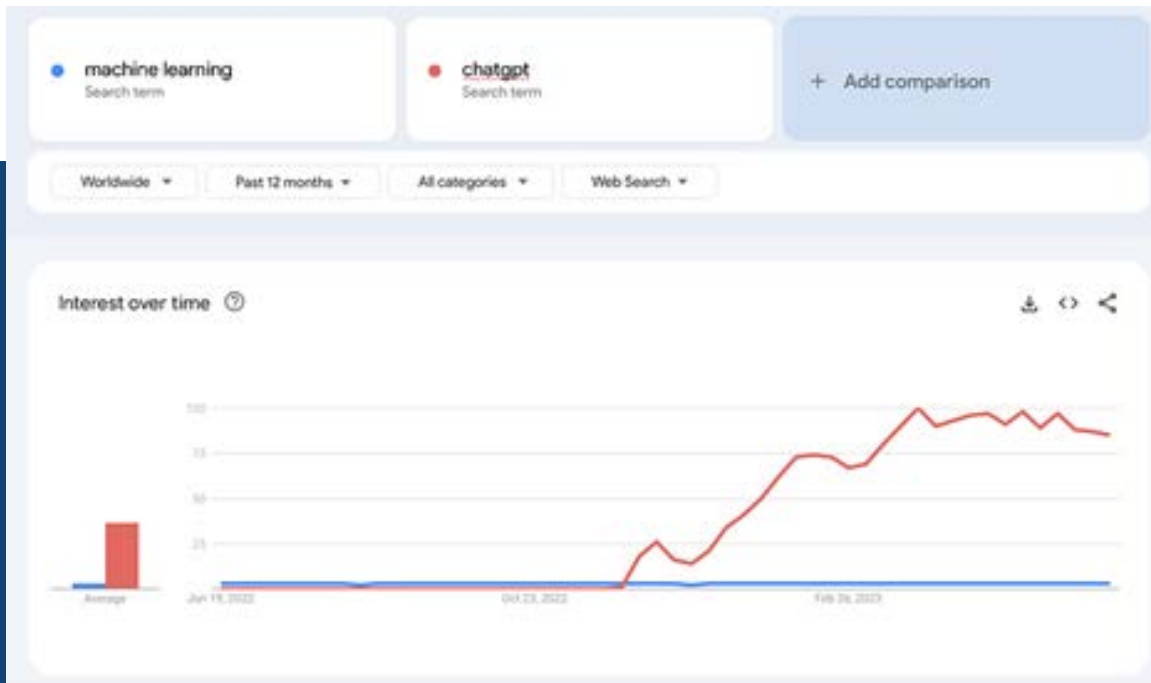
And on the face of it, IBM has read the top 10 current hits very well.

There is a lot of architectural choice and flexibility packed into the IBM platform. Let's dip our toe into one technical detail — IBM's watsonx.data is based on Apache Iceberg lakehouse technology as opposed to the leading alternative which is Delta Lake. I'll lift a quote: "While Delta Lake is mostly backed by Databricks, Iceberg is backed by many companies, including Netflix, Adobe, Alibaba, and many others. This means that Iceberg is becoming a standard in the industry. Wider open source commitment and adoption are huge by the industry. Many vendors are already baking Iceberg support" [9]. Time will tell whether Iceberg becomes the Betamax or the VHS of lakehouse technologies, but IBM is clearly putting themselves on a path of differentiation.

So watsonx.data is architected to scale to AI levels.

Increasing scale...tick! Reducing cost...tick!

Yesterday's machine learning excitement has been quickly eclipsed by generative AI [10].



Anyone heard of ChatGPT? [screenshot from trends.google.com]

I previously mentioned foundational models are a smart inclusion in this platform. Models that will generate language, models that will generate code, models that will understand the world around us. The bar for AI jumped up considerably over the last 12 months — and so has expectations. What was sci-fi just last year is a minimum expectation today.

Of watsonx.ai, IBM tell us:

“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: Models 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.”

With watsonx.ai you get generative AI-powering foundational models included with the platform.

Human capabilities in a box...tick!

The other smart inclusion is AI governance. This is probably the most topical, although at levels approaching hysteria. Once all the hysteria dies down (pun not intended), there will be a very real set of AI governance requirements — ethical decision making, bias and fairness, transparency and explainability, robustness and safety, oversight and control — which makes governance an essential part of any AI platform. IBM already has a strong pedigree of data governance products, and AI governance is a logical inclusion.

So, built-in AI failsafe mechanism (aka AI governance)...tick!



“Give it back!”
No – don’t worry. AI hasn’t taken over... yet (Tongue firmly in cheek.)

With watsonx, IBM looks to have read the mood music, leant on the best of its back catalog, internalized the current greatest hits, and possibly produced a new hit record that is on-trend, and seemingly of its time.

So when will the record be on the shelves and ready to purchase? (yes, yes, I’m dating myself — make that “When will it be available for streaming/download?”)

Release week? First live appearance!

If all of this has got you excited to be in the crowd for the first live performance, the good news is you don’t have to wait! IBM has recently dropped the first two components (watsonx.ai and watsonx.data), with watsonx.governance due “later this year.”

We have a backstage pass for you! Fill out this quick form for more information.



Greg Hodgkinson is Prolifics’ Chief Technology Officer and Worldwide Head of Engineering, and an IBM Lifetime Champion. As a technology leader, he’s responsible for innovative cross-practice solutions for our customers, creating a foundation for innovation in the company, and driving improvements in the art of software development and delivery throughout Prolifics.

- [1] A word from Arvind
- [2] Read all about it!
- [3] They are foundational
- [4] Lakehouses, a short history
- [5] Brian who?
- [6] watsonx wouldn’t be the same without governance
- [7] IBM Watson
- [8] Catch up on CP4D
- [9] Iceberg support?
- [10] ChatGPT vs ML, trends