Get your data [AI-ready] - what CDOs need to know

The role of data product marketplaces in accelerating AI and business success

[-]luwise

Turning AI promise into reality

No senior business leader needs to be reminded of the potential of AI, and the spread of the technology across organizations and industries. According to McKinsey, over three-quarters of organizations now use AI in at least one business function, reporting an average of a 5% increase in revenues and 10% reduction in costs.

But often this theoretical promise doesn't translate into actual business benefits at scale. Research from BCG finds that 74% of companies struggle to achieve and scale AI value.

Why is this? **The key reason for AI failure comes down to data**, which is often overlooked in the rush to pilot AI projects. Issues such as poor quality, unavailable, or untrustworthy data were identified by 93% of executives interviewed by Wavestone as the biggest barrier to their AI success.

Chief Data Officers and other data leaders recognize the problem. 79% of organizational leaders in Wipro research said AI was essential to their company's future, but just 14% believed that their data maturity could support AI at scale. Data problems impact the training of AI models, prevent accurate predictive analytics, lead to misinformed decision-making and now threaten to derail the introduction of agentic AI.

All of this means that focusing on building trust in data, and providing fast, straightforward access to it is essential to reaping the benefits of AI. The advantages extend far beyond AI too. Reliable high-quality and usable data is at the heart of increasing human consumption, driving data democratization.

CDOs know that they need to act now, to ensure their data is AI-ready to increase usage and drive value. This guide therefore outlines how they can achieve this through a clear data strategy combined with emerging approaches such as data products and data product marketplaces, backed by collaboration with the business, strong governance and integrated data stacks.

Contents

1 No data, [no AI]

Organizations are currently sitting on treasure troves of data which can potentially power their AI programs. Internal and proprietary data, such as customer information, can help differentiate their AI systems, whether by training Large Language Models (LLMs) or providing fuel for AI agents. Equally, external data can provide context, but is often not centralized or accessible to AI.

Yet, the rush to AI has led to many organizations building their projects on shaky or even non-existent foundations. Of the five key reasons for AI project failure listed by the Rand Corporation, **two directly link to data – a lack of necessary data and inadequate data infrastructure.**

What are Large Language Models and Agentic AI?

A Large Language Model (LLM) is an advanced artificial intelligence system trained on vast amounts of data to understand, generate, and predict actions. It works by using deep neural networks to recognize patterns in data and then using these patterns to generate contextually relevant responses that drive prediction, automation, insights, and innovation across business functions. Commercial LLMs have been trained on publicly available data, such as from the internet – increasingly organizations are adding value by training such models using their own, internal data to provide differentiated, customized results.

Agentic AI (or AI agents) are AI systems designed to act autonomously to plan, reason, interact with the world or other systems and take actions and make decisions toward achieving set objectives with minimal human intervention. Fed with data, AI agents work by combining large language models with decision-making, memory, and tool-use capabilities, allowing them to dynamically analyze information, execute tasks, and adapt behavior to changing business contexts by learning from experience. Essentially, they bring AI to the real-world, both physical and digital, performing human tasks and offering an infinite number of possibilities for benefiting from AI by handling complex, multi-step problems.

The unbreakable link between data and AI

Essentially AI systems learn patterns from data or take actions based on analyzing information. That means that the quality of that data directly impacts their accuracy, fairness, and usefulness. Seamless and rapid access to a wide variety and volume of high quality, reliable, timely, trustworthy and comprehensive data is therefore critical to AI success. No quality data, no AI.

Even worse, without reliable data, decisions made by AI are skewed or biased, leading to:

- **Poor decision making:** Three-quarters (76%) of data leaders in a Precisely study say data-driven decision-making is their primary goal for data programs, yet 67% don't completely trust their data.

- **Compliance and legal issues:** Algorithms that have been trained on poor quality or incomplete data can lead to biased or discriminatory results, leading to costly legal cases and regulatory action.

- **Worsening AI accuracy:** AI models learn and continually improve when fed quality data. Unreliable data means their accuracy degrades over time, preventing them from benefiting the business.

- **Inability to generalize:** Incomplete data leads to AI systems that are only accurate under certain conditions. This undermines trust in results, particularly when scaling from pilot to production.

- **Security failings:** Poor data governance processes risk confidential data being made available outside the business, such as being fed into Generative AI (GenAI) models, breaching security and governance rules.

- **Higher costs:** AI projects may fail completely, wasting company investment or require additional time and resources to relabel data and retrain models.

- **Overfitting and underfitting:** Inconsistent data leads to AI models providing random results, with no way to understand how they have reached their conclusions.

- **Reputational damage:** As the examples in the next section demonstrate, AI failures directly impact corporate reputation, undermine trust in brands, and cause share prices to fall.

Data-led AI failures

High profile, publicly known cases include:

Amazon's hiring algorithm was trained on ten years of predominantly male resumes. The result was that it consistently downgraded any applications that contained the word "women".

Housing company Zillow used AI models to predict housing prices for its home-buying business. The model relied on imperfect housing data and failed to capture market volatility, leading to the company losing over $500 million and having to shut down the division.

In the public sector many governments have applied AI to try and transform and modernize processes but have relied on incomplete or poor-quality data. For example, The Dutch Tax and Customer Administration agency used AI to detect childcare benefit fraud, but use of sensitive data and opaque and biased algorithms led to many false positives. Thousands of families were wrongly accused of fraud, leading to financial devastation, broken homes, and even suicides, as well as forcing the resignation of the Dutch government.

The consequences of bad data on agentic AI

While these examples all relate to AI LLMs, the consequences of poor data in emerging agentic AI could be even worse.

Poor-quality data in classic AI leads to **bad predictions**, but in agentic AI leads to **bad actions**, which can cascade into real-world consequences, such as financial loss, safety risks, the spread of misinformation or breakdowns in trust:

- As AI agents take a sequence of actions, bad data might cause a **snowball effect**. For example, a financial trading agent acting on inaccurate market signals could trigger a chain of bad trades, amplifying losses and causing issues in the wider market.
- Poor data may cause agentic AI to optimize for the wrong outcomes, with physical consequences.
- Incorrect data can cause AI agents to hallucinate, providing inaccurate information either to humans or other agents without challenge.
- Issues caused by bad data could impact physical security and safety, such as by affecting the operations of self-driving cars or automated factories.
- Introducing unreliable data to agents, either deliberately or accidentally, leads to feedback loop corruption, potentially hijacking systems and introducing or reinforcing bad behaviors. This has already been seen in chatbots which serve as basic AI agents.
- A lack of reliable data to feed AI agents will mean they fail to deliver value and waste resources.

2 [The challenges] to AI-ready data

The growth in the volume, variety and velocity of data is not a new phenomenon triggered by AI. It is a trend that CDOs and other data leaders have been grappling with years, first looking to better manage data and then aiming to turn it into value for the business.

Successfully managing increasing data volumes has been complicated by a range of factors:

- **Data silos across businesses**, with information trapped in departmental systems
- **Complex data infrastructures** made up of multiple, incompatible and overlapping tools
- **The need to ensure compliance with regulations** such as the GDPR and CCPA, as well as protecting the security of data, especially confidential and personal information
- **An inability to create a single version of the truth** and to make reliable, consistent data available across the organization

All of these factors prevent the widespread sharing of trusted data, preventing its universal consumption and the creation of a data-centric organization. 68% of available data remains unused in organizations, according to IDC research.

AI has exposed the issue of poor-quality, unavailable and unreliable data. and made solving it a C-level priority. Without access to accurate, trusted internal data, organizational AI projects simply will not scale beyond pilots or deliver their potential ROI. When it comes to AI models, companies will be forced to rely on undifferentiated external data, impacting their success and limiting results.

On the positive side this is focusing efforts and resources on overcoming organizational data issues. That's why 75% of organizations are increasing their technology investments around data management due to Gen AI, according to a Deloitte study. However, it also increases the pressure on CDOs and data leaders, meaning they need to deliver reliable, AI-ready data to the business as quickly as possible to enable the scaling of AI, and realizing the competitive benefits it brings.

Success requires CDOs to overcome 8 key challenges:

[1] Data integration issues

Data is scattered across organizations, in a range of systems and formats, from business applications to cloud storage. New types of data, such as from sensors or unstructured information is constantly being added to the data estate. All of this needs to be connected, integrated and converted to standard formats by building robust, automated and compliant pipelines.

[2] Poor data quality

Alongside integrating data from multiple systems, data teams need to understand and enhance its quality. What are the issues with particular data sources and how can they be fixed through processes such as data cleansing? Demonstrating the scale of the issue, 55% of respondents in the 2025 Beyond Big Data report say data quality is their top barrier to unlocking AI's full potential.

[3] Data inconsistency

Data may appear to meet quality standards in terms of formatting, but it can be inconsistent, with gaps in datasets or covering a wide range of values. Ensuring consistency requires an understanding of the source of the data, along with the ability to highlight and remove any values that are outliers that will skew the results of analysis.

[4] Data duplication

Duplicate data stored in different systems pushes up costs and increases management overheads. Adding to this, training AI models using multiple copies of the same data leads to inaccurate results that limit the applicability of AI.

[5] Incomplete data governance

Consistent, organization-wide data governance is critical to protecting data and achieving compliant, AI-ready data. However, silos and an partial picture of the data estate mean that often data governance is incomplete or is seen as a bureaucratic barrier by the business. This can mean data is made available without access controls to audit and protect confidential information, potentially leading to models and results that infringe personal data regulations.

[6] Slow availability of data

AI requires up-to-date, fresh data to deliver relevant results. Often, the processes used to integrate, cleanse and make data available can be lengthy, meaning models and agents are relying on old data to make decisions. This leads to inaccurate conclusions that aren't based on the latest information.

[7] Lack of context to data

Single, raw datasets or assets can only give a partial picture. As they are produced by one source they lack context, making it difficult for humans or AI models to gain a complete, actionable view. They need to be enriched with a combination of other internal datasets and external reference data (such as statistical or geospatial information) in order to provide sufficient context for accurate results.

[8] Data decay

Data has a shelf-life, becoming less useful the longer ago that it was generated or collected. This data decay means that models need to be constantly updated with fresh information to prevent the accuracy of their results deteriorating over time as data ages.

3 [Solving the barriers] to AI-ready data

Successfully delivering AI-ready data requires taking a step beyond traditional data management approaches, ensuring that data and processes all align with organizational needs in order to turn data into business value.

Gartner five steps

Gartner lists five key steps to follow to bridge the gap and make your data AI-ready:

[1] Assess your data needs based on AI use cases:

Ensure data is able to meet AI use case requirements, especially around quality, trust, semantics, confidence and governance.

[2] Gain board buy-in:

Educate the board on requirements to deliver AI-ready data, reiterate the value it will unlock and set clear goals for data transformation.

[3] Evolve data management practices:

Focus processes on specific AI capabilities and use cases, and ensure data is fully enriched with metadata to deliver context and aid discovery.

[4] Extend the data management ecosystem:

Ensure you have AI, data literacy and metadata skills and the right tools in place to manage your data.

[5] Scale and govern:

Identify and extend existing data governance to include AI use cases, monitor forthcoming regulations and put in place strong AI ethics and responsible use rules.

The benefits of an AI-ready data strategy

Ensuring your data is AI-ready enables the more accurate and tailored training of AI models and feeds AI agents with reliable, actionable information. At the same time AI-ready data provided through a data product marketplace enables humans to discover, access and consume data, maximizing its value. Key benefits include:

- **Greater operational performance**, efficiency and productivity through automation
- **Better decision-making** and predictive analytics through more accurate insights
- **Increased trust** in data from business and AI, driving usage
- **Enhanced transparency** around organizational activities for the public, partners and regulators
- **Development of new business services** through differentiation and monetization
- **Enabling greater innovation** by providing relevant data to AI and humans

The foundations of an AI-ready data program

Building on these five steps, organizations require six key foundations to be in place in order to create and deliver reliable, accurate AI-ready data to where it is needed to deliver business value.

[1] Set a clear data strategy

Begin by auditing your current data landscape, gaps and actions to be taken to optimize data for specific AI use cases. AI works by analyzing numerical values, so ensure that your data has been quantified by turning raw information into relevant, accurate, quantitative numbers. Also check that it meets standards around semantics, quality, trust and diversity and that it is AI-ready in terms of labelling, metadata, bias mitigation and lineage.

AI needs to solve a specific business problem or challenge if it is to deliver ROI. Therefore, understand the most pressing business needs and where a combination of AI and data can help deliver maximum ROI quickly. Ensure you have buy-in from both business teams and the board for your strategy, sufficient resources to implement it, and clear success metrics to measure progress against.

[2] Focus on data products

Data products are specific, high-value, highly consumable data assets. They contain everything needed for use by non-experts, without requiring training or support. Governed by clear data contracts that outline what they can be used for and provide SLAs on quality and reliability, they are designed for large-scale, ongoing usage by both humans and AI.

By bringing together and enriching multiple datasets and making them accessible through an intuitive interface, machine-readable data products solve issues about context, accessibility, trust, lineage, accountability and observability. You can see what data has been used to create a product and where it has come from, which is vital for monitoring and checking AI algorithms and agentic AI processes.

> *"Raw data, however abundant, is not yet an asset. To become useful, it must be shaped into a data product — a curated, reusable data set or service that delivers high-quality, ready-to-use information for a clearly defined business need."*
>
> McKinsey

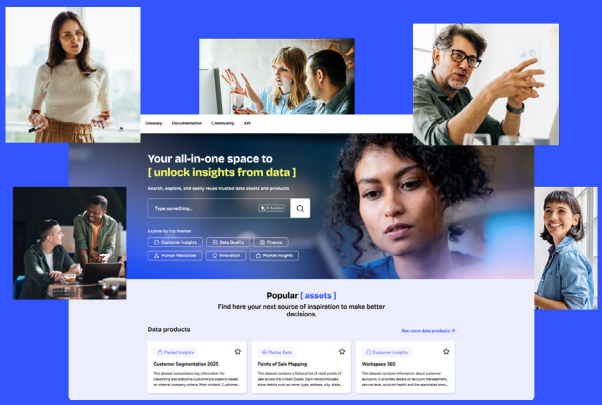
[3] Deliver data through a data product marketplace

If data assets, including data products cannot be easily governed, discovered, accessed and consumed then they simply will not be used by either humans or AI. They need to be available via a secure, self-service data product marketplace that contains all data assets in one place, providing a single source of truth. With an experience based on e-commerce product marketplaces, users and AI models can quickly find and incorporate data products into their activities.

A data product marketplace provides the consumption layer for AI-ready data, ensuring that it is well-organized, accessible, machine-readable, and reliable, with granular access management capabilities to meet governance and security considerations, and in-depth lineage to show what data (products) have been used by LLMs/agentic AI.

[4] Reinforce data governance standards

Data governance covers how you identify, organize, handle, manage, and use data collected in your organization, reducing risk and enabling agility. Effective governance solves issues around compliance, regulations, security, and consistency. AI-ready data requires governance to be extended to ensure that it sets and enforces standards around areas such as metadata, and the responsible and ethical use of data, particularly personal data, to train AI models. Good governance is vital to effective AI. According to Precisely, companies solving governance challenges deploy AI 3x faster with 60% higher success rates than their peers.



[5] Build an agile, comprehensive tech stack

Years of investment in standalone tools has created complex, sprawling data and tech stacks made up of a combination of point solutions that don't easily integrate and/or duplicate functions. This makes data management difficult and reduces the ability for organizations to seamlessly deliver trustworthy, reliable, AI-ready data in the form of data products.

To overcome this sprawl, Gartner recommends that organizations should look to streamline tech stacks with new approaches such as Data Management Platforms (DMPs). These are converged, end-to-end solutions that provide core capabilities in a single product and replace point tools. Importantly, these need to be combined and integrated with best-of-breed data product marketplace solutions to deliver accessible, consumable data assets to AI and human users.

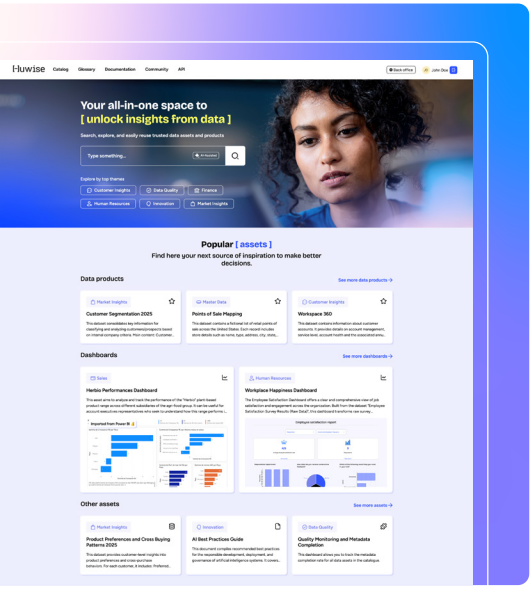
[6] Deepen collaboration with the business

AI and data must solve key business problems if they are to deliver value and ROI. This requires greater collaboration between data/IT/governance teams, business users and data owners across the organization. Working together in cross-functional teams, all of these groups can break down silos and balance centralized control and independence through federated data mesh models, with the key aim of increasing data consumption and value.

Data product marketplaces facilitate this teamwork by providing a centralized space to share, govern, consume, manage and consume data through an intuitive experience that meets the needs of all groups. Users can give feedback and rate data, while data owners can answer queries and administrators grant access to specific data products for relevant users and AI.

4

[The data product marketplace] at the heart of AI-ready data



At Huwise, we are focused on turning data into a source of insight, innovation, and performance, for both humans and AI. Our data product marketplace enables the seamless discovery of relevant data products and other data assets, providing secure access to the information needed to power AI models and agentic agents.

Today, more than 350 clients across 25 countries rely on Huwise to accelerate their data initiatives, with over 3,000 platforms already deployed worldwide. Backed by 14 years of expertise in data management, we deliver tailored support to address the business use cases of our clients.

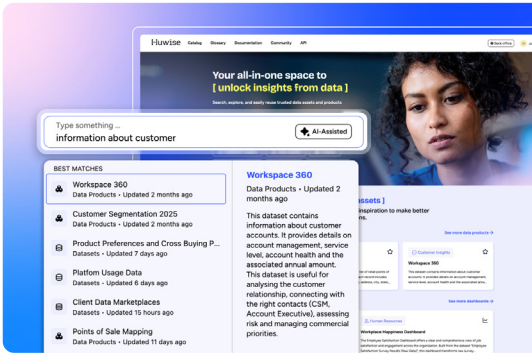
Huwise's strengths in data product marketplaces have been recognized by Gartner, which has included us as a Sample Vendor in the Data Marketplace and Exchange (DME) category in its Hype Cycle for Data Management 2025.

Gartner®

Hype Cycle 2025, Data Management

How a data product marketplace delivers AI-ready data

The Huwise data marketplace solution enables you to create an intuitive, centralized space that provides a single source of truth across the organization, for both humans and AI.



Intuitive, AI-driven experience

Huwise's data product marketplaces feature an AI-powered search engine to drive data discoverability and reuse.

Through AI-assisted search, users and AI agents can quickly discover the most relevant data for their needs. AI also suggests similar assets and enables the automated creation of visualizations through prompts and questions.

Every data asset includes a machine-readable page that displays full information, including its source, owner, description, and full metadata, including theme, creator, modification date, primary contact, and technical identifier. This ensures it is AI-ready, and can be confidently understood and incorporated into LLMs and agentic AI processes.

Powerful features to prepare data products

Huwise's data product marketplace solution not only shares data, but also shapes it to prepare ready-to-use data products. Native connectors and APIs seamlessly connect to all data sources and data management solutions to centralize information and make it simple to combine data assets into data products, backed by support for data contracts.

The solution also incorporates a wide range of processors to improve the quality and context of data, through geographical mapping, processing of dates, text transformations and generic treatments. Huwise's Data Hub of nearly 50,000 public datasets is also available to customers, enabling them to connect external data sources quickly and increase context within their data products.

Data can be shared in multiple, machine-ready formats: pages, datasets, applications, APIs, or downloads, underpinning your strategy and deployment of AI models and agents. Huwise's built-in Model Context Protocol (MCP) server makes AI operational through access to trustworthy data, linking directly to enterprise applications and AI used by the business, ensuring data is instantly usable across all business channels.

AI-ready governance and access management

AI brings increased risks and concerns around ethics, data access, transparency and compliance. Huwise's data product marketplace guarantees and optimizes AI governance through a range of capabilities that integrate with overall governance strategies.

Data access is managed at a granular, role-based level, controlling exactly who can find, preview, or consume each data product, with native request and approval workflows to minimize administration and speed data access approvals.

Data lineage capabilities provide end-to-end analysis of how data is reused within data products and AI. Mapping models the trajectory of a given asset from its point of origin to its point of destination and delivers a complete view of how it has been reused, both internally and externally. This audit trail increases transparency, ensures ethical use of data and delivers full compliance and governance.

AI + data = ROI

By centralizing data and making it available in easily consumable, machine-readable ways through a data product marketplace, your data become AI-ready and can be seamlessly discovered and ingested by AI models and agents, as well as being incorporated into AI systems by developers.

With a data product marketplace you can:

- → Deliver reliable, accurate AI-ready data across your entire organization
- → Accelerate internal AI deployment by training models and agents with high quality data
- → Ensure data accuracy, data governance, and ethical use at scale

Huwise's turnkey MCP server links your data to real-world AI, driving agentic AI deployments and automating data product marketplace management to increase efficiency. With proven compatibility with major AI environments, inwcluding Claude Desktop, OpenAI, Mistral AI, and Microsoft Copilot, it seamlessly integrates with your strategic technology choices to build your AI stack.

5 Successfully delivering data to [power AI]

Harnessing AI to drive greater insights, improved efficiency and increased innovation is at the top of C-level business priorities. However, many organizations have struggled to scale AI projects beyond the pilot stage due to inadequate data foundations, leading to lost opportunities and wasted resources.

Reliable, machine-readable data is key to AI success, both when training LLMs and underpinning fast-emerging agentic AI implementations. Now is the time for CDOs and other data leaders to focus their efforts and investment on building a data stack that supports the creation, discovery and consumption of reliable, accessible data products for humans as well as AI. Data product marketplaces are central to this strategy, providing a centralized, intuitive, self-service space for access to trusted data that powers AI success and makes data available securely to all. Together, data product marketplaces and AI underpin data democratization, elevating thinking and delivering insight, innovation, and performance inside the business and wider society.

Huwise

Huwise helps data leaders transform their data into a source of knowledge, innovation, and performance.

We design a SaaS data product marketplace solution tailored to business teams and non-expert users, making data accessible, actionable, and value-generating. With Huwise, organizations provide employees and partners with secure, self-service access to their data products through a single platform.

In just a few weeks, we can deploy any type of data product marketplace, whether internal or external.

Today, more than 350 clients across 25 countries rely on Huwise to accelerate their data initiatives, with over 3,000 platforms already deployed worldwide. Backed by 14 years of expertise in data management, we deliver tailored support to address the concrete use cases of our clients.

Data made yours.

www.huwise.com