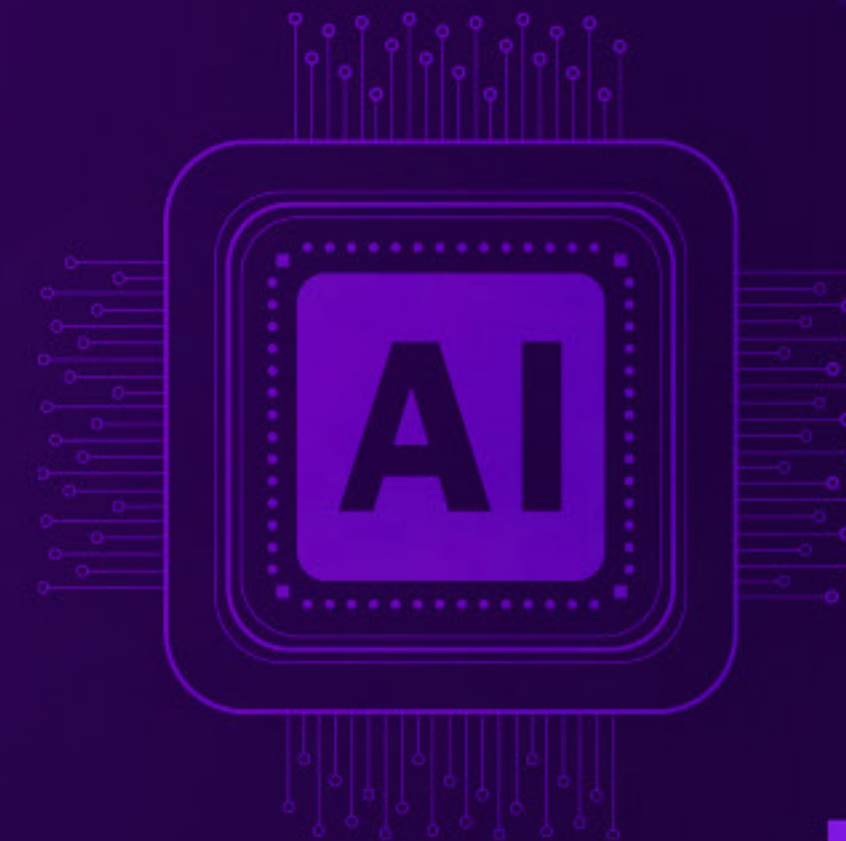# precisely

Trusted AI 101:
Tips for Getting Your Data AI-Ready

AI

Successful AI initiatives rely on trusted data

With competition fiercer than ever, it's easy to understand why businesses are racing to harness the power of artificial intelligence (AI) to increase productivity and efficiency, attract and retain customers with personalized service, and create new ideas that provide a competitive edge.

However, there are risks to rushing into AI without the proper preparations. In 2023, we saw AI failures ranging from **AI-written briefs**[1] containing fake citations to renowned consulting firms implicated in **non-existent scandals**[2] and many others.

In each case, the model that produced the bogus results had inadequate training data for the intended purpose, leading to skewed and flawed outputs, underscoring the need for AI powered by trusted data.

And trusted data requires data integrity — or data with maximum accuracy, consistency, and context. Think of it this way: your AI outputs will only be as strong as the data feeding them.

However, organizations often face data integrity challenges, like struggling to integrate data fast enough, understand and govern its responsible use, observe and improve its quality, enrich it for deeper context, and guarantee security and privacy.

The benefits are limitless if your AI applications learn from trusted, AI-ready data. And yet, the current reality for many organizations is sobering.

In the 2023 Gartner IT Symposium Research Super Focus Group, only 4% say their data is AI-ready.

With all this in mind, let's explore valuable AI use cases and the data integrity fundamentals you need to ensure trust and success in your results.

AI in action: Six use cases that succeed with data integrity

AI applications were previously reserved for automation and predictive analytics. With generative AI (GenAI), they push the boundaries of new, imaginative use cases, enabling the creation of content, ideas, and data that can significantly enhance your competitive advantage. In an October 2023 **study by Fortune/ Deloitte**[3], 79% of CEOs said, "Accelerating innovation is one of the top use cases for implementing GenAI."

A wealth of new possibilities can finally be realized through intuitive, natural language-based access to corporate data – making previously unattainable use cases a reality.

The following six use cases are examples of AI succeeding with data integrity.

1 USE CASE
AI recommendations

The benefits

Faster, more personalized recommendations.

How it's done

An AI recommender system is a sophisticated technology that leverages AI and vast amounts of user data — like past preferences, behaviors, and interactions — to suggest tailored products, content, or services.

Why data integrity is essential

Recommendations will only be accurate and relevant if the data has integrity, especially the critical element of data quality. Data integrity also powers more effective training of models, resulting in improved recommendation performance, increased user satisfaction, and higher engagement levels.

2 USE CASE
AI-powered workflows

The benefits

Increased productivity, higher efficiency, and lower costs.

How it's done

Workflows can be automated with AI models that process data in real time. The applications range from automating sales and marketing campaigns and project management workflows to coding assistants, where the benefits are pronounced. According to a **Stack Overflow survey**[4] of over 90,000 developers, almost 70% use or plan to use AI tools.

Why data integrity is essential

Integrated systems often use multiple data sources, so you must combine critical data from all relevant sources, including complex transactional systems. Implementing data quality and governance practices improves data reliability, reduces errors, and ensures accurate results — leading to efficient workflows. For smooth operations and high-quality results, data integrity is vital.

3

Machine learning applications

The benefits

Accelerated business processes with greater accuracy.

How it's done

Machine Learning (ML) applications enable computers to learn from data and make predictions or decisions autonomously, like generating fast pricing quotes and delivering greater customer satisfaction — but only when data engineers train models with high-integrity data.

Why data integrity is essential

To dramatically reduce data prep time and increase accuracy, you must integrate data, ensure quality and governance, and add context through enrichment and spatial analytics. Together, these steps ensure fresh data that provides trust in your ML models' outputs.

4 Foundation Model training

The benefits

Natural language processing abilities enable Foundation Models (FMs) to generate content and code, summarize text, analyze sentiment, answer questions, and more.

How it's done

A Foundation Model (FM) is an ML model pre-trained on large datasets and designed to capture general patterns and features.

However, a significant challenge with FMs is the potential for learned bias. For example, FMs used by global banking organizations to process loan applications for minority-owned or home-based businesses are at risk if the data contains inherent biases that reflect societal prejudices, stereotypes, and disparities.

When FMs generate text or provide responses, they may inadvertently replicate and amplify existing social, gender, or racial biases, leading to discriminatory outputs that exacerbate inequalities in various domains.

Why data integrity is essential

To prevent bias, you must train GenAI models on data carefully curated with relevant datasets from across systems, preprocessed to ensure quality, and enriched with third-party attributes to guarantee that all groups are accurately represented.

⑤ Chatbots

The benefits

Efficient and personalized assistance that increases user engagement.

How it's done

Chatbots built on large language models (LLMs) can deliver natural, contextually rich responses to user prompts. An LLM is a type of FM trained on vast amounts of textual data to understand and generate human-like language.

A chatbot's ability to dynamically generate responses based on the ongoing conversation sets them apart, enhancing user engagement across multiple industries and use cases like customer support.

GenAI's impact on customer service is already being felt. The National Bureau of Economic Research **surveyed**[5] 5,179 customer support agents and found an average productivity increase of 14% when exposed to AI tools. This number goes up to 34% for novice workers.

Why data integrity is essential

High-quality chatbot responses require LLMs trained on high-quality, complete data. An FM will take the data as-is to generate a response — it won't correct issues. Data integrity is essential to bringing together all relevant data for the GenAI model and ensuring it's accurate, consistent, and contextualized.

6 AI assistants using RAG

The benefits

Enhanced contextual understanding for AI assistants.

How it's done

AI assistants can be built using retrieval augmented generation (RAG), which combines FMs with existing external knowledge to improve model performance and create more helpful and relevant results. RAG-powered systems excel in question-answering scenarios, providing context-aware and detailed answers from extensive knowledge bases.

Why data integrity is essential

Since data is the driving force of the model's response, the integrity of this data is critical. Inaccurate data can result in false responses. Similarly, incomplete or poor-quality data may result in hallucinations.

For domain-specific scenarios in industries like real estate and insurance, enrichment with third-party data provides additional context to enhance the specificity and relevance of the AI assistant's response, such as integrating property attribute data into a RAG-based GenAI application that produces descriptions of real estate assets .

Solve Top AI Challenges with Data Integrity

Incomplete data, compliance struggles, a lack of context for AI outputs… do these challenges and others like them familiar? If so, the good news is that they're all solvable.

To reap the many benefits of AI, like the ones we've covered above, you can take a proactive approach by investing in robust, tailored data integrity capabilities. Let's explore common challenges in more detail.
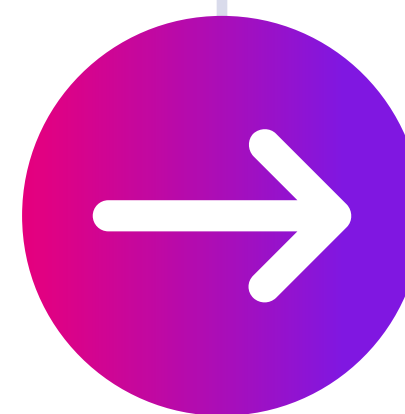
THE CHALLENGE
Narrow, biased results

Incomplete datasets and separate data infrastructure stacks limit an AI's understanding of the questions it's asked and produce biased, unreliable results.

Because enterprise data is generated and stored in various legacy systems, not all relevant and critical data may be available on the (cloud) platform where AI is run. Lack of access to a complete view of data can result in bias.

THE SOLUTION
Data integration

Incomplete datasets and separate data infrastructure stacks limit an AI's understanding of the questions it's asked and produce biased, unreliable results.

Because enterprise data is generated and stored in various legacy systems, not all relevant and critical data may be available on the (cloud) platform where AI is run. Lack of access to a complete view of data can result in bias.

Untrustworthy results

Inaccurate predictions and recommendations from your AI lead to a lack of trust and can potentially prevent further adoption of these technologies. These are some of the consequences of poor data quality.

The demand for data that's accurate, consistent, and fit for purpose by AI applications is driving a rethinking of the data quality domain. Without reliable data, advanced AI models are of little use. The stakes are high, and overcoming data quality roadblocks has become a priority.

Additionally, governments across the globe have been stepping up efforts to ensure compliance with regulations aimed at protecting data privacy, which puts more pressure on businesses to comply with data sovereignty laws.
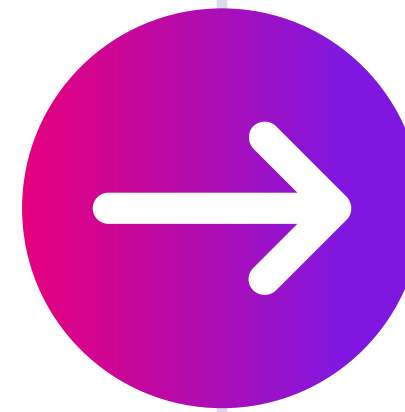
Data quality and governance

AI initiatives require a new approach to data quality to ensure they're using data that's accurate, consistent, and fit for purpose. This will often come in the form of core data quality and business rules, automated validation and cleansing, and integration with data observability and data governance solutions.

Proactive data quality tools can monitor data pipelines, use advanced ML techniques to quickly identify anomalies and outliers, and employ AI to recommend or create rules that ensure issues are remediated before they reach downstream systems.

Data governance provides a clear understanding of how the data you use in AI applications is:

- **Collected:** what are the data types?
- **Stored:** where is it located?
- **Used:** who has access to it?

This unified understanding helps you enforce policies and procedures that protect your data.

You need a proactive approach to data security and compliance to protect sensitive data, safeguard your reputation, and get the most out of AI applications. Implementing robust data governance measures and staying current with the latest regulations is critical.

THE CHALLENGE

Lack of contextual relevance

Without context in the nuances and dependencies of a given real-world scenario, the AI bases its inference or recommendations on only a small portion of the bigger picture. This can lead to incomplete, inaccurate, or contextually irrelevant results with potentially dangerous downstream impacts.
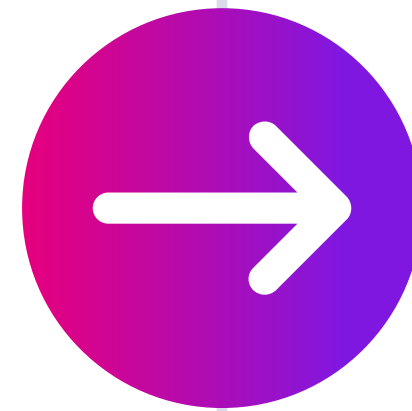
THE SOLUTION

Spatial analysis and data enrichment

For more accurate and contextualized AI outputs, you must enrich the data that fuels them with trusted third-party data and spatial insights. For instance, you may want to combine address details and environmental risk factors with your property portfolio data for more accurate predictive modeling and natural catastrophe insights.

This process requires a thorough spatial analytics and data enrichment strategy to accurately map your data to real-world scenarios. With your AI and ML models trained on accurate and relevant data, you'll be ready to produce bigger and better outcomes.

Three Data Integrity for AI Considerations

Many executives consider integrating new technologies into their business models a top strategic priority – and the focus is on transforming their infrastructure through AI.

However, AI applications based on probability tend to hallucinate, with varying results based on how the prompt was engineered. Comprehensive data integrity can go a long way in building trust.

Training your AI applications with accurate, consistent, and contextualized data is key to reliable results that fuel success. It comes down to three primary data integrity considerations for AI. Here's what you need to know and how to achieve your desired results.

1

A more complete dataset helps you realize the full potential of your AI

What you achieve with data integrity

Training AI models with all relevant critical data on-premises, in the cloud, and hybrid environments—including complex data residing on your mainframe—can minimize bias, improve accuracy and reliability, and enhance understanding.

How you achieve it

Break down data silos and bring fresh data to your AI development environment – fast – with modern data pipelines from Precisely. Tap into the broadest and deepest functionality for compute and scalable cloud storage with cloud instances optimized for training and inference. By providing your team with cloud resources to derive value from massive datasets, you make AI work for your business.

2

Fuel your AI applications with trusted data to power reliable results

What you achieve with data integrity

Ensure AI outcomes you can trust. For accurate predictions, recommendations, and effective process automation, models must be trained by data with integrity delivered via cloud-native solutions.

For trusted AI outcomes, your data needs to meet rigorous quality metrics; it needs to be accurate, complete, validly structured, standardized, and free of duplicates. High-integrity data should also be timely, governed using a robust framework, and observed for changes and anomalies.

How you achieve it

When you have transparency into the data lineage in your pipelines, you can improve and observe its quality and govern your data and AI models. You can streamline this process using the Precisely Data Integrity Suite, which runs data integrity processes where your data resides – across operational and analytical systems. Using data with integrity for model training, deployment, inference, and monitoring makes it easier to build trust in your AI results.

Using trusted data to train and fine-tune your ML and GenAI models is essential. An automated data integrity program can help you better understand your data and automate steps needed to improve its quality. The services of the Precisely Data Integrity Suite provide the required tools to achieve this goal.

When dealing with the high volume of data required to train models, you need a cloud-native approach that scales elastically and cost-effectively. Monitoring and proactive anomaly detection are necessary to ensure continuous data quality and visibility into your entire pipeline. For example, automated alerts can prevent bad data from being sent to your AI development environment or other services for training.

3 Add context to your data for more relevant and nuanced responses

What you achieve with data integrity

Enriching the data that fuels your AI applications with trusted third-party data and spatial insights can boost their accuracy and contextual relevance.

How you achieve it

To accelerate the development of AI applications and increase their adoption, add context to your data that allows the AI to grasp nuances, maintain coherence, and generate contextually relevant responses.

You can achieve this by bringing third-party data and spatial insights from Precisely to your AI development environments. Curated, authoritative datasets like this add more detail to what you and your AI know about

places, people, properties, businesses, and environmental risk factors.

Many cloud service providers offer AI and machine learning (ML) services, infrastructure, and implementation resources to help you at every stage of your AI adoption journey. By building AI applications using high-integrity data and these services, you increase your AI adoption and ensure that your applications are accurate and relevant.

Make your data AI-ready and maximize the potential of your AI-based solutions using the Precisely Data Integrity Suite.

Precisely helps you gain trusted AI outcomes by bringing together critical data and ensuring its optimal quality — governed by a robust framework, observed to detect degradation, and enriched with essential context derived from spatial insights and third-party datasets.

These considerations ensure your data is of high integrity and that the resulting AI applications can be delivered with complete trust and reliability. Focus on increasing data integrity will accelerate the development and adoption of AI within your organization.

1. **Judge sanctions lawyers for brief written by A.I. with fake citations**
   https://www.cnbc.com/2023/06/22/judge-sanctions-lawyers-whose-ai-written-filing-contained-fake-citations.html

2. **Academics apologise for AI blunder implicating Big Four**
   https://www.accountingweb.co.uk/tech/tech-pulse/academics-apologise-for-ai-blunder-implicating-big-four
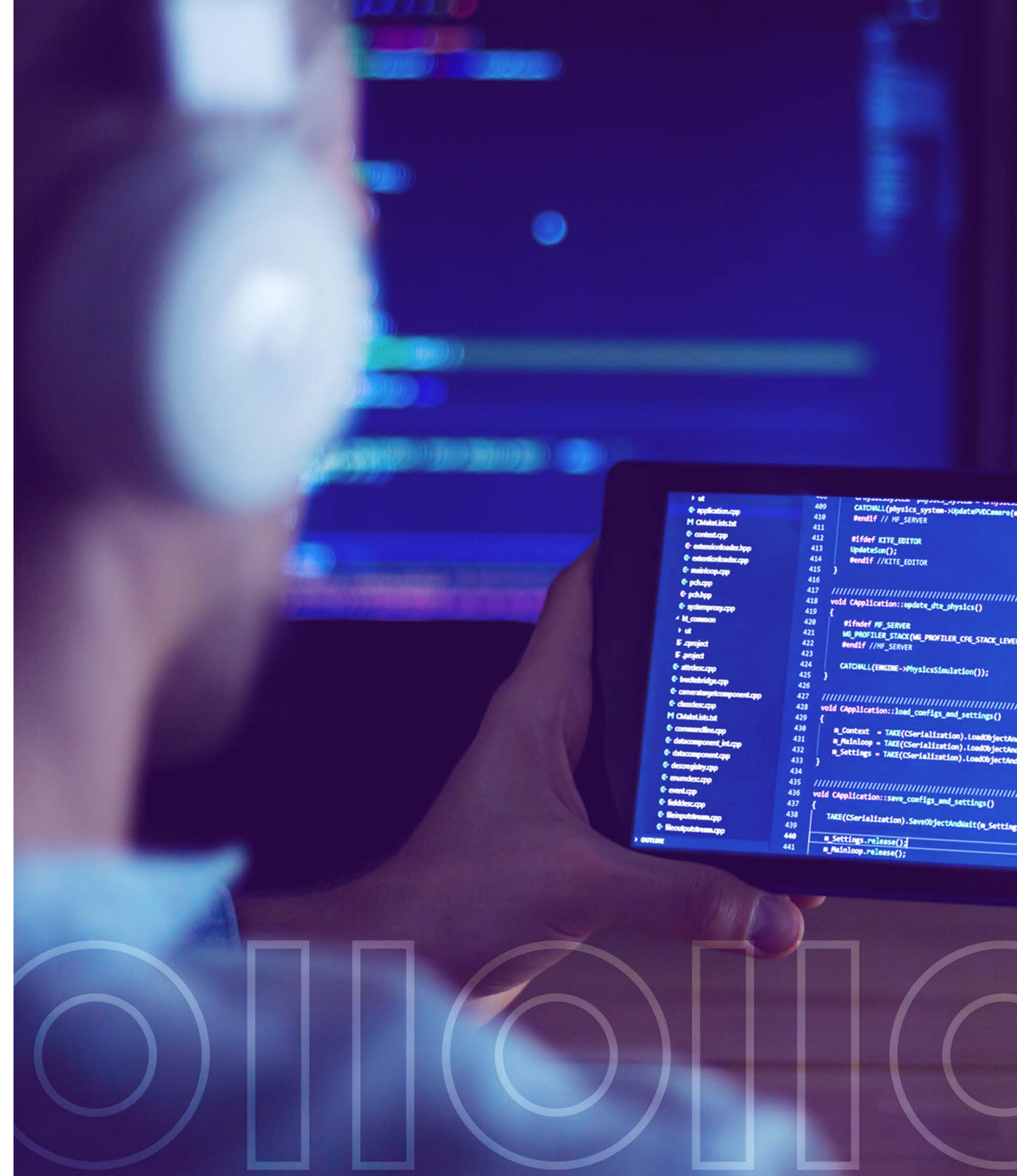
3. **Fall 2023 Fortune/Deloitte CEO Survey Insights**
   https://www2.deloitte.com/us/en/pages/chief-executive-officer/articles/ceo-survey.html

4. **Developer sentiment around AI/ML**
   https://stackoverflow.co/labs/developer-sentiment-ai-ml

5. **Generative AI at Work**
   https://www.nber.org/papers/w31161

Summary

With the rise of GenAI applications, prioritizing data integrity has never been more crucial.

For AI initiatives that are high-performing, reliable, and produce quality outputs, your data must be complete and accurate, trusted and consistent, and contextualized. This helps you overcome the barriers standing in your way, making fostering trust and adoption easier and accelerating AI development and innovation.

What are your top use cases for AI? What could you achieve if you unlock its full potential? Whatever your unique goals, remember that trusted AI starts with trusted data. Future-proof your AI by starting the data integrity journey today.

www.precisely.com/suite

Data integrity for AI resources

PRODUCTS
The Precisely Data Integrity Suite

WEBINAR
AI You Can Trust: Embracing Data Integrity Throughout the Development Lifecycle

BLOGS
Trusted Generative AI using Precisely and AWS

Use Data Enrichment to Supercharge AI

The Power of AI in Precisely Software

What's Ahead in Automation

Tackling Top Data Issues with the Precisely Data Integrity Suite

precisely

Precisely is the global leader in data integrity, providing accuracy and consistency in data for 12,000 customers in more than 100 countries, including 99 of the Fortune 100. Precisely's data integration, data quality, data governance, location intelligence, and data enrichment products power better business decisions to create better outcomes.

Learn more at www.precisely.com.