Informatica

Infographic

# IT Leaders' Checklist: Data Requirements for Predictive AI and GenAI

You're ready to embrace artificial intelligence (AI) — now what? Take the first step in your AI journey: establish data requirements for predictive AI and generative AI (GenAI). Doing so is crucial for enhancing the accuracy, reliability and overall performance of your AI models.

Get your data ready for AI with this comprehensive checklist.



☑ **High-quality and trusted**
Deliver data that is correct and precise for model accuracy. Ensure all necessary data fields are complete, with no missing values. The data must be suitable for the prediction timeframe and consistent across datasets and time periods.

**Large in volume and high in veracity**
Use a dataset that's sufficiently large to capture common patterns and variability for accurate predictions. It should include a representative sample, such as the entire population or specific scenarios in the model.

☐





☐ **Clean**
Eliminate random errors and irrelevant information from your data. Appropriately manage data outliers by either removing them or applying treatments.

**Well-structured**
Organize structured data into tables, using databases or spreadsheets. Conversely, unstructured data — including text, images and other formats — typically requires preprocessing. For instance, an extra step may be necessary when preparing text data for natural language processing.

☐





☐ **Compliant with privacy and security measures**
Verify that your data collection and usage adhere to applicable privacy laws and regulations specific to your country or industry. Remember to protect the privacy of individuals by anonymizing sensitive data.

**Governed for access and control**
Establish governance protocols to guarantee only authorized personnel access data and all usage complies with company policies and legal requirements. Maintain data integrity and compliance with regulatory standards through effective management of data access rights and governance policies.

☐





☐ **Integrated for advanced analytics and decision-making**
Combine batch and streaming data sources to gain a comprehensive view of information, which enhances real-time analytics and decision-making capabilities. This integration enables dynamic updates and insights, essential for developing adaptive and responsive predictive models.

**Relevant**
Align your data specifically with the task for which you're training the GenAI model.

☐





☐ **Domain-specific**
Incorporate domain-specific data into your dataset to tackle specialized tasks. For example, in the healthcare sector, this might include medical images, patient histories and symptoms to personalize treatments.

**Contextual**
Tailor data to fit your use case by integrating models with enterprise data to yield relevant insights. Instead of relying on a single large language model (LLM), capitalize on the strengths of different models by combining multiple GenAI base models.

☐



**Meet essential data requirements to secure AI readiness.**

GET EBOOK

Source: Informatica, AI Playbook for IT Leaders: 4 Strategic Moves to Drive AI Innovation, 2024

**Where data & AI come to**
LIFE
Informatica

Informatica (NYSE: INFA), a leader in enterprise AI-powered cloud data management, brings data and AI to life by empowering businesses to realize the transformative power of their most critical assets. We have created a new category of software, the Informatica Intelligent Data Management Cloud™ (IDMC), powered by AI and an end-to-end data management platform that connects, manages and unifies data across virtually any multi-cloud, hybrid system, democratizing data and enabling enterprises to modernize their business strategies. Customers in approximately 100 countries and more than 80 of the Fortune 100 rely on Informatica to drive data-led digital transformation. **Informatica. Where data and AI come to life.™**

IN20-5063-1024

© Copyright Informatica LLC 2024. Informatica and the Informatica logo are trademarks or registered trademarks of Informatica LLC in the United States and other countries.
A current list of Informatica trademarks is available on the web at https://www.informatica.com/trademarks.html. Other company and product names may be trade names or trademarks of their respective owners.
The information in this documentation is subject to change without notice and provided "AS IS" without warranty of any kind, express or implied.