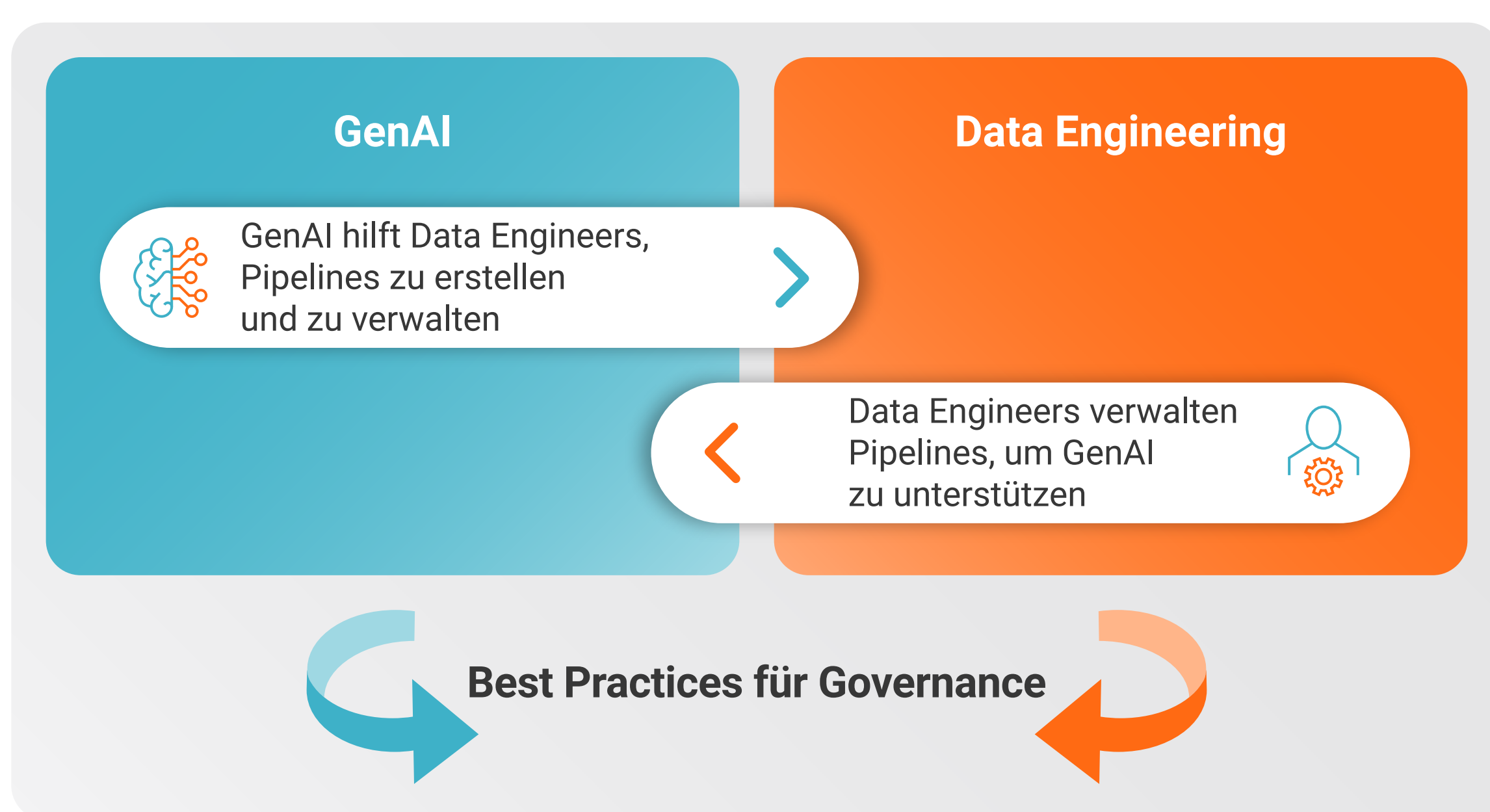


Synergien nutzen: Das Zusammenspiel zwischen GenAI und Data Engineering

Es ist unbestritten, dass Data Engineering nur mit generativer KI funktioniert und generative KI nur mit Data Engineering. Das hat verschiedene Gründe: GenAI erhöht die Produktivität von Data Engineers, während Data Engineering bahnbrechende Innovation für GenAI ermöglicht.

Erfahren Sie, wie Sie das dynamische Zusammenspiel zwischen GenAI und Data Engineering fördern können.

Das Zusammenspiel zwischen GenAI und Data Engineering



Wie GenAI Data Engineering hilft

Das Problem



Data Engineers beschäftigen sich mit Entwurf, Test, Umsetzung, Monitoring und Optimierung von Pipelines, die Daten für Analytics-Zwecke bereitstellen. Der Anstieg an SaaS-Anwendungen, mobilen Apps, IoT-Sensoren, Datenplattformen, Analysetools und Business Usern macht die Verwaltung von Aufgaben zur Einspeisung und Umwandlung von Daten so kompliziert. Es ist sehr schwierig, diese unterschiedlichen Komponenten nahtlos zu integrieren.

Die Lösung



Mit LM-Plattformen (Sprachmodelle) und fortschrittlichen Funktionen innerhalb von Pipeline Tools kann diese Herausforderung gemeistert werden. Sie basieren auf Befehlen in natürlicher Sprache von Data Engineers und generieren Startcode für Datenpipelines, machen Vorschläge, um fehlerhaften Code zu verbessern und dokumentieren die Pipelines und dazugehörigen Datensätze für die Katalogisierung. Darüber hinaus empfehlen LMs Regeln für Überprüfungen der Datenqualität und bewerten verschiedene Architekturansätze für die Planung von Pipelines. Das führt dazu, dass Sie mithilfe von LMs viel Zeit bei Automatisierung, Beschleunigung und Vereinfachung der komplexen und mühsamen Aufgaben sparen, die das Data Engineering beinhaltet.

Wie Data Engineering GenAI hilft

Das Problem



Unternehmen erstellen ihre eigenen GenAI-Anwendungen, die ein LM (oder eine API (Application Programming Interface), die mit einem LM verbunden wird) beinhalten, sowie eine Benutzeroberfläche (UI) für die Kommunikation und zusätzliche Funktionen, die Aufgaben ausführen, die auf LM-Outputs basieren. Um verwertbare Outputs zu generieren, benötigen GenAI-Anwendungen nutzbare Inputs und Pipelines, die unstrukturierte Datenobjekte in numerische Vektoren umwandeln können, mit denen Nutzerbefehle angereichert und LM verbessert werden können.

Die Lösung



Data Engineers können diese Herausforderung meistern, indem Sie Pipelines erstellen, die ETL- oder ELT-Phasen beinhalten (Extraction, Transformation und Loading). Mit der Abfolge des Extrahierens und Ladens bzw. des Transformierens und Ladens können die unternehmensspezifischen Daten eines Unternehmens für die Nutzung durch GenAI-Anwendungen vorbereitet werden.



Extrahieren und Laden

Pipelines extrahieren relevanten Text aus Anwendungen und Dateien und laden ihn in den Landing-Bereich auf Plattformen, wie Databricks Lakehouse oder Snowflake Data Cloud. Um die Genauigkeit von GenAI zu verbessern, sollte dieser Text mit Masterdaten übereinstimmen und Qualitätsstandards entsprechen.



Transformieren

Die Daten werden in den Pipelines umgewandelt, um für die LM-Nutzung vorbereitet zu werden. Danach werden Wörter in numerische Token konvertiert, die wiederum in „Blöcke“ gruppiert werden. Danach werden Vektoren erstellt, die die Bedeutung und Beziehungen dieser Blöcke beschreiben.



Laden

Diese Einbettungen werden durch Pipelines in eine Vektor-Datenbank (wie Pinecone und Weaviate) oder in Vektor-fähige Plattformen (wie Databricks und MongoDB) geladen.

Im nächsten Schritt können Datenteams die vektorisierten Daten verwenden, um GenAI-Anwendungen auf zweierlei Weise zu unterstützen:

- 1 Zur Implementierung von RAG (Retrieval-Augmented Generation), so dass relevante Inhalte in der Vektor-Datenbank gefunden und zu Nutzerbefehlen hinzugefügt werden, damit die Qualität der Antworten des LM steigt.
- 2 Zur Feinabstimmung des LM, indem die Parameter an den vektorisierten Text angepasst werden.

Sowohl RAG als auch die Feinabstimmung sind wichtig, damit GenAI-Anwendungen präzise Antworten erstellen. Durch diese Kombination wird sichergestellt, dass der Geschäftskontext stärker berücksichtigt wird und das Risiko von Datenhalluzinationen sinkt.

Transformieren Sie Ihre Datenstrategie durch GenAI und Data Engineering.

MEHR ERFAHREN

Quelle: Achieving Fusion: How GenAI and Data Engineering Help One Another, 2024

Where data & AI come to



Informatica (NYSE: INFA) ermöglicht es Unternehmen, das transformative Potenzial ihrer Daten und von KI voll auszuschöpfen. Wenn Daten effizient genutzt werden, können sämtliche Datenbestände im Unternehmen demokratisiert werden, so dass aus Chaos Klarheit entsteht. Dank der Informatica Intelligent Data Management Cloud™ können Unternehmen ihre Daten optimal nutzen, um große Ideen voranzutreiben, Prozesse zu verbessern und Kosten zu senken. Diese Cloud beruht auf unserer KI-Engine CLAIRE® und ist die einzige Cloud speziell für die Verwaltung sämtlicher Datentypen und -muster, für Daten jeglicher Komplexität oder für sämtliche Workloads, unabhängig vom Standort – über eine einzige, zentrale Plattform.

IN20-4792-0524

© Copyright Informatica LLC 2024. Informatica und das Logo von Informatica sind Marken oder eingetragene Marken von Informatica LLC in den USA und in anderen Ländern. Die aktuelle Liste mit Marken von Informatica ist hier zu finden: <https://www.informatica.com/de/trademarks.html>. Alle weiteren Firmen- und Produktbezeichnungen können Handelsnamen oder Marken ihrer jeweiligen Inhaber sein. Die in diesem Dokument enthaltenen Informationen können sich ohne vorherige Ankündigung ändern und werden „wie gesehen“ und ohne jegliche ausdrückliche oder stillschweigende Gewährleistung bereitgestellt.