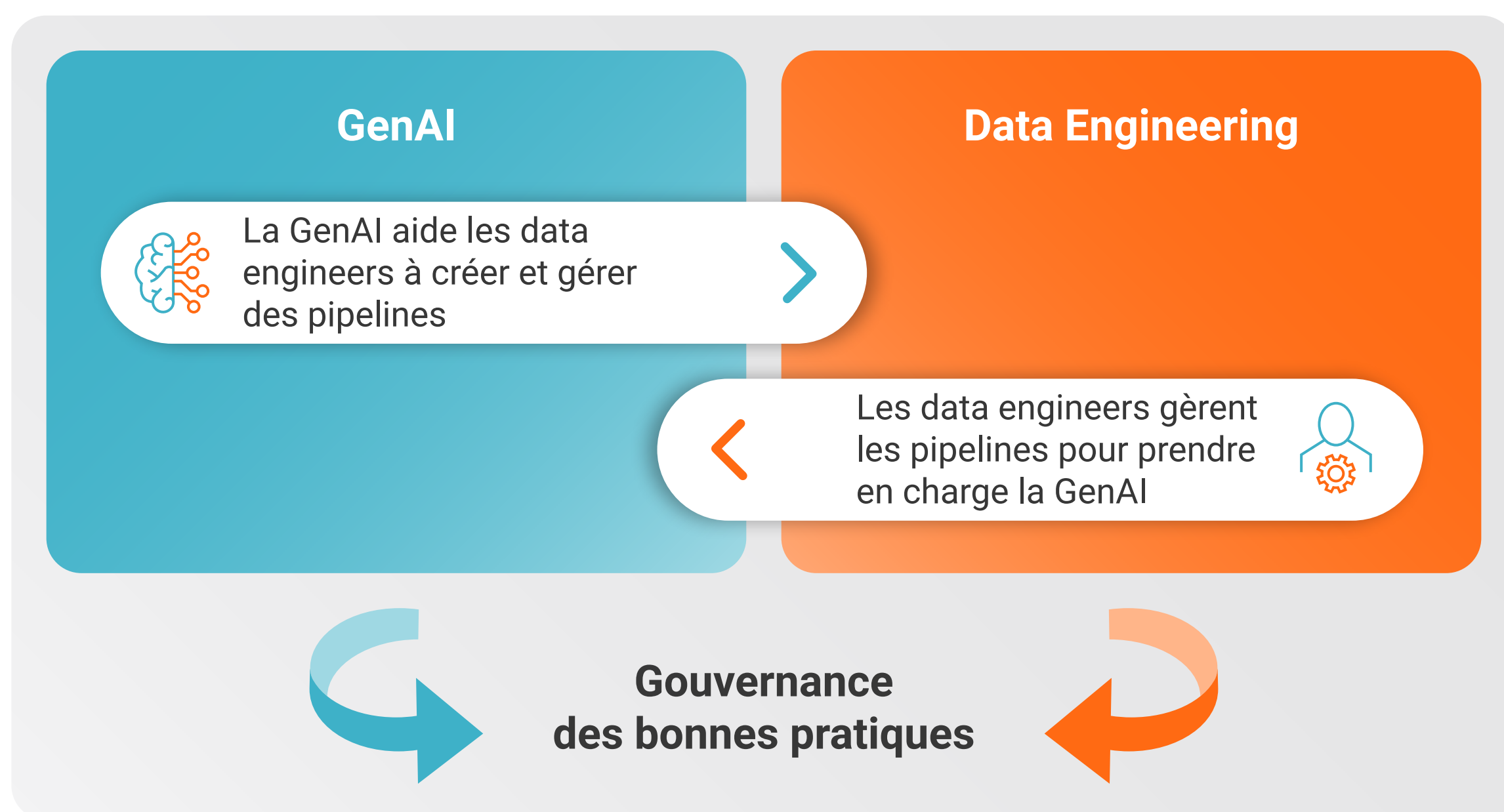


Accomplir la fusion : Synergie entre GenAI et Data Engineering

À n'en pas douter, le Data Engineering a besoin de l'aide de l'intelligence artificielle générative (GenAI) et la GenAI a besoin de l'aide du Data Engineering. En voici les raisons : La GenAI booste la productivité des data engineers, tandis que le Data Engineering offre de nouveaux niveaux d'innovation à la GenAI.

Découvrez comment favoriser une interaction dynamique entre GenAI et Data Engineering.

Fusion entre GenAI et Data Engineering



Comment la GenAI soutient le Data Engineering

Le problème



Les data engineers conçoivent, testent, déploient, surveillent et optimisent les pipelines qui fournissent des données à des fins d'analyse. L'essor des applications SaaS, des applications mobiles, des capteurs IoT, des plateformes de données, des outils d'analyse et des utilisateurs métiers complique la gestion des tâches d'ingestion et de transformation des données, et par voie de conséquence l'intégration transparente de ces composants variés.

La solution



Les plateformes de modèle de langage (LM) et les fonctionnalités avancées des outils de pipeline permettent de relever ce défi. Basées sur les invites en langage naturel des data engineers, elles génèrent le code de démarrage des pipelines de données, suggèrent des moyens de déboguer le code et documentent les pipelines et les jeux de données associés pour le catalogage. De plus, les modèles de langage recommandent des règles pour les contrôles de qualité des données et évaluent différentes approches architecturales pour la conception des pipelines. Par conséquent, les modèles de langage font gagner un temps considérable par l'automatisation, l'accélération et la simplification des tâches complexes et fastidieuses liées au Data Engineering.

Comment le Data Engineering soutient la GenAI

Le problème



Les entreprises commencent à créer leurs propres applications de GenAI qui incluent un modèle de langage (ou une interface de programmation d'applications (API) qui se connecte à un modèle de langage), une interface utilisateur conversationnelle et des fonctionnalités supplémentaires qui exécutent des tâches basées sur les sorties du modèle de langage. Pour fournir des résultats exploitables, les applications de GenAI ont besoin d'entrées et de pipelines capables de transformer des objets de données non structurés en vecteurs numériques qui enrichissent les invites utilisateur et aident à affiner le modèle de langage.

La solution



Les ingénieurs de données peuvent relever ce défi en créant des pipelines qui incluent les étapes d'extraction, de transformation et de chargement (ETL) ou d'ELT. La séquence d'extraction et de chargement, de transformation et de nouveau chargement peut préparer les données propres au domaine d'une entreprise pour leur utilisation par leurs applications de GenAI.



Extraction et chargement

Les pipelines extraient le texte pertinent à partir des applications et des fichiers, puis le chargent dans une zone d'accueil sur des plateformes telles que Databricks Lakehouse ou Snowflake Data Cloud. Pour améliorer la précision de la GenAI, ce texte doit correspondre aux données de référence et répondre aux normes de qualité.



Transformation

Les pipelines transforment les données pour les préparer à leur consommation par le modèle de langage, puis convertissent les mots en jetons numériques, en regroupant les jetons en « segments » et en créant des vecteurs qui décrivent la signification et les relations entre les segments.



Chargement

Les pipelines chargent les intégrations dans une base de données vectorielle (comme Pinecone et Weaviate) ou des plateformes vectorielles (comme Databricks et MongoDB).

Ensuite, les équipes de données peuvent utiliser les données vectorisées pour prendre en charge les applications de GenAI de deux manières principales :

- 1 Mise en œuvre de la génération améliorée par récupération (RAG), qui recherche le contenu pertinent dans la base de données vectorielle et l'ajoute aux invites utilisateur afin que le modèle de langage soit davantage susceptible de fournir des réponses de haute qualité.
- 2 Ajustement du modèle de langage en réglant ses paramètres pour les aligner sur le texte vectorisé.

La génération améliorée par récupération et l'ajustement jouent un rôle déterminant dans l'amélioration de la précision des réponses générées par les applications de GenAI. Cette combinaison garantit une représentation plus précise du contexte métier et limite le risque d'hallucinations de données.

Repensez votre stratégie de données grâce à la puissance combinée de la GenAI et du Data Engineering.

EN SAVOIR PLUS

Source : Accomplir la fusion : comment la GenAI et le Data Engineering s'allient, 2024

Where data & AI come to



Informatica (NYSE : INFA) donne vie aux données et à l'IA en permettant aux entreprises de concrétiser la puissance de transformation de leurs ressources les plus stratégiques. Lorsqu'elles sont correctement traitées, les données deviennent une ressource vivante et fiable qui est démocratisée à l'échelle de votre entreprise, la clarté s'imposant face au chaos. Grâce à l'Intelligent Data Management Cloud™ d'Informatica, les entreprises insufflent de la vie à leurs données afin de développer l'innovation, d'améliorer les processus et de réduire les coûts. Optimisé par CLAIRE®, notre moteur d'IA, IDMC est le seul Cloud dédié à la gestion des données quel qu'en soit le type, le modèle, la complexité ou la charge de travail et indépendamment de l'endroit où elles se trouvent, le tout sur une plateforme unique.

IN20-4792-0524

© Copyright Informatica LLC 2024. Informatica et le logo Informatica sont des marques commerciales ou déposées d'Informatica LLC aux États-Unis et dans d'autres pays. La liste des marques commerciales d'Informatica est disponible sur le Web, à l'adresse: <https://www.informatica.com/fr/trademarks.html>. Plus de 7 000 entreprises du monde entier font appel aux solutions de données d'Informatica. Pour en savoir plus, appelez le +33 01 42 04 89 00 ou visitez notre site www.informatica.com/fr.