

Aide-mémoire : Ingénierie des données et IA générative

Outils populaires d'ingénierie des données et d'IA

Outil	Usage	Lien
Robot de données	Apprentissage automatique automatisé et IA	www.datarobot.com
Préfet	Gestion des flux de données	www.prefect.io
Brumeux	Génération de données synthétiques	https://hazy.com/
ChatGPT	Modèle de langage de l'IA	https://openai.com/chatgpt
Données de point	Intégration des données	www.stitchdata.com
Rédacteur de schémas	Génération de schéma de données	schemawriter.ai
donneesuniverselles.io	Génération et augmentation de données	https://www.universaldata.io/
DBdiagram.io	Création d'un diagramme ER	https://dbdiagram.io/

Principaux éléments à prendre en compte pour l'IA générative dans la conception architecturale

Tâche	Rapide
Générer une conception d'architecture de données pour un réseau hospitalier	Créez une conception détaillée de l'architecture de données pour un réseau hospitalier.
Ajouter des composants de modélisation de données pour les données des patients	Créez des composants de modélisation de données supplémentaires pour les données démographiques des patients, les antécédents médicaux, le diagnostic, le traitement et les normes de qualité.
Concevoir pour mettre en œuvre des contrôles d'accès robustes, la confidentialité des données et des mécanismes d'audit	Créez une conception incluant la mise en œuvre de contrôles d'accès robustes, de mécanismes de cryptage et d'audit pour protéger les données des patients contre tout accès non autorisé ou toute violation.
Conception détaillée de l'architecture des données pour la relation client d'une entreprise de vente au détail	Créez une conception d'architecture de données détaillée pour le système de gestion de la relation client d'une entreprise de vente au détail.
Générer un profil client unifié	Créez un profil client unifié avec des attributs tels que les données démographiques, l'historique des achats, le comportement de navigation, les préférences et les interactions.

Conseils importants pour l'IA générative pour la conception de schémas de bases de données et d'entrepôts de données

Tâche	Rapide
Générer un schéma d'entrepôt de données pour un magasin de mode	<p>Créez un schéma d'entrepôt de données détaillé pour un magasin de vente au détail de mode qui devrait contenir :</p> <ul style="list-style-type: none">i) Données sur les employésii) Données sur les ventesiii) Données sur les stocksiv) Profils des clientsv) Informations sur le vendeur

Conseils importants pour l'IA générative en vue de l'anonymisation des données

Tâche	Rapide
Anonymiser les noms	Remplacez les entrées sous l'attribut « Nom » d'un ensemble de données par des pseudonymes tels que « User_i » à l'aide de Python.
Rédiger les adresses e-mail	Écrivez un code Python pour rédiger les entrées sous l'attribut « E-mail » dans un bloc de données afin que seuls les noms d'utilisateur et les premier et dernier caractères du fournisseur de services soient visibles. Tous les autres caractères sont remplacés par le caractère « * ».
Généraliser les âges aux décennies	Écrivez un code Python pour généraliser les entrées sous l'attribut « Âge » d'un bloc de données de sorte que le nombre exact soit converti en une plage générique
Ajoutez du bruit aux numéros de contact	Écrivez un code Python pour ajouter du bruit aléatoire d'une longueur de 5 chiffres à un attribut numérique « Numéro de contact » dans un bloc de données avec toutes les valeurs de longueur de dix chiffres.

Conseils importants pour l'IA générative pour la configuration de l'infrastructure

Tâche	Rapide
Générer l'infrastructure de données de la plateforme e-commerce	<p>Écrivez du code Python qui exécute les tâches suivantes :</p> <p>créez les étapes permettant à la plateforme de commerce électronique d'améliorer son infrastructure de données pour gérer l'augmentation du trafic. Proposez des améliorations dans</p> <ol style="list-style-type: none"> 1. Stockage évolutif 2. Meilleures capacités de traitement 3. Analyses en temps réel
Générer une infrastructure de lac de données pour une plateforme de soins de santé	<p>Écrivez du code Python qui exécute les tâches suivantes :</p> <p>Créez les étapes pour qu'une entreprise de soins de santé mette en place une infrastructure de lac de données capable d'effectuer les tâches suivantes :</p> <ol style="list-style-type: none"> 1. Gestion des Big Data 2. Ingestion de données provenant de diverses sources 3. Transformation des données 4. Sécurité des données et conformité aux directives réglementaires
Générer une infrastructure pour détecter les transactions frauduleuses en temps réel pour une entreprise financière	<p>Écrivez du code Python qui exécute les tâches suivantes :</p> <p>créez les étapes permettant à une société financière de mettre en place son infrastructure si elle souhaite détecter les transactions frauduleuses en temps réel. Proposez des détails en termes de :</p> <ol style="list-style-type: none"> 1. Machines informatiques 2. Pipeline d'ingénierie des fonctionnalités 3. Pipeline de modélisation prédictive 4. Déploiement et surveillance du modèle

Auteur(s)

Réseau de compétences

