

Programmes de formations



IA & Data Science

Version 17/04/2026

1. Introduction à l'intelligence artificielle générative

Public : tous public

Durée : 2 jours, adaptable en fonction du besoin

Pré-requis : Pas de pré-requis

Méthodologie et Evaluation :

- Interventions sous formes d'exposés
- Mise en application de la théorie via des ateliers
- Questionnaire d'évaluation des connaissances
- Support de cours et exercices

Objectifs de la formation :

- Connaître le fonctionnement de l'IA
- Utiliser les outils de générations de texte, d'images et présentations
- Ecrire les prompts adéquats

Programme :

Fonctionnement de l'IA générative

- Définition de l'IA Générative
- Exemples d'applications
- Classification des outils les plus connus et utilis ations
- Biais algorithmiques et éthique
- Ecrire les bons prompts
- Réglementations

Génération de texte

- Découverte des outils de génération de texte (Bing / Perplexity / POE / ChatGpt)
- Comment affiner son résultat
- Application des concepts sur un cas concret

Workshop Pratique Génération d'images

- Découverte des outils de génération d'image
- La génération d'images avec Canva
- Comment construire un bon prompt
- Quels paramètres utiliser pour affiner son résultat
- Présentation d'une App ChatGPT dédiée à la génération d'image
- Application des concepts sur un cas concret

Workshop Pratique Génération de support type PPT



- Découverte des outils de génération de support de présentations (Tome)
- Découverte et test de différents outils
- Application des concepts sur un cas concret

2. Intelligence Artificielle pour Data Engineer

Public : Data Engineer et experts techniques

Durée : 2 jours, adaptable en fonction du besoin

Pré-requis : Pas de pré-requis

Méthodologie et Evaluation :

- Interventions sous formes d'exposés
- Mise en application de la théorie via des ateliers
- Questionnaire d'évaluation des connaissances
- Support de cours et exercices

Objectifs de la formation :

- Comprendre les fondamentaux de l'IA et du Machine Learning appliqués à la data engineering
- Intégrer des modèles d'IA dans des pipelines de données
- Utiliser des API d'IA (LLM, embeddings) dans un contexte data
- Automatiser des traitements data enrichis par l'IA
- Mettre en place des architectures data augmentées par l'IA
- Identifier les limites, risques et bonnes pratiques (coûts, sécurité, biais)

Programme :

Jour 1

1 Rôle du Data Engineer à l'ère de l'IA

- Évolution du rôle : du ETL → AI-ready pipelines
- Différence entre Data Engineer / ML Engineer / AI Engineer
- Cas d'usage concrets :
 - Enrichissement de données
 - Classification automatique
 - Data quality intelligente
- Brainstormings et réflexions

2 Fondamentaux IA utiles au Data Engineer

- ML vs Deep Learning vs LLM
- Concepts clés :
 - Features / labels
 - Inférence vs entraînement
- Focus LLM :
 - Prompting
 - Tokens / coût

- Latence
- Atelier : Test d'une API IA (classification simple de données)

3 Intégrer l'IA dans un pipeline data

- Architecture type : ingestion → transformation → enrichissement IA → stockage
- Appels API dans un pipeline
- Gestion des erreurs et retry
- Traitement batch vs temps réel
- Atelier pratique : Pipeline Python

4 Embeddings & recherche sémantique

- Qu'est-ce qu'un embedding ?
- Cas d'usage : recherche sémantique
 - déduplication
 - clustering
- Introduction aux vector databases
- Atelier : générer des embeddings, faire une recherche sémantique simple

5 Architectures data + IA

- Data Lake + IA
- Feature store
- RAG (Retrieval Augmented Generation)
- Orchestration (Airflow-like)
- Étude de cas : Architecture d'un pipeline IA complet

6 Industrialiser l'IA dans les pipelines

- Monitoring des appels IA
- Gestion des coûts (tokens, API)
- Caching des résultats
- Versioning des modèles / prompts
- Atelier : Optimisation d'un pipeline IA (coût + perf)

7 Sécurité, gouvernance et limites

- Données sensibles et IA
- RGPD & data privacy
- Biais des modèles
- Hallucinations
- Cas pratique : analyse de risques sur un pipeline IA

3. IA pour Data Analyst

Public : Data Analysts et experts techniques

Durée : 2 jours, adaptable en fonction du besoin

Pré-requis : Pas de pré-requis

Méthodologie et Evaluation :

- Interventions sous formes d'exposés
- Mise en application de la théorie via des ateliers
- Questionnaire d'évaluation des connaissances
- Support de cours et exercices

Objectifs de la formation :

- Comprendre les apports de l'IA dans l'analyse de données
- Utiliser des outils d'IA pour automatiser l'analyse
- Exploiter les LLM pour interroger, transformer et analyser des données
- Générer des insights automatiquement à partir de datasets
- Créer des analyses augmentées (text, classification, clustering)
- Identifier les limites, biais et risques liés à l'IA

Programme :

1 Le Data Analyst augmenté par l'IA

- Évolution du rôle du Data Analyst
- IA comme copilote d'analyse
- Cas d'usage : génération d'insights
- analyse exploratoire automatisée
- synthèse de rapports
- Réflexions autour des tâches automatisables avec l'IA

2 Comprendre les LLM pour l'analyse de données

- Fonctionnement simplifié des LLM
- Prompting pour analystes
- Structuration des requêtes (ex : "analyse ce dataset...")
- Limites (hallucinations, biais)
- Atelier : Générer une analyse automatique à partir d'un dataset

3 IA pour l'analyse exploratoire (EDA)

- Détection de patterns
- Analyse descriptive automatisée

- Détection d'anomalies
- Génération de visualisations assistées
- Atelier pratique : Utilisation Python + 1 outil dataviz et exploitation IA Gen

4 Génération d'insights et storytelling

- Génération automatique de commentaires
- Synthèse de dashboards
- Storytelling assisté par IA
- Génération de rapports
- Exercices pratiques : transformer un dashboard en rapport automatisé

5 Automatiser les analyses avec l'IA

- Pipelines d'analyse
- Automatisation de rapports
- Requêtes dynamiques avec IA
- Intégration dans outils BI
- Projet de script d'analyse automatisée (dataset → insights)

6 Classification, clustering et NLP

- Classification de données texte
- Clustering automatique
- Analyse de sentiments
- Tagging automatique
- Atelier : Classifier des données clients / tickets

7 Recherche intelligente et analyse sémantique

- Recherche sémantique
- Embeddings simplifiés
- Analyse de documents
- Cas d'usage métier
- Projet : recherche intelligente dans un dataset texte

4. Python pour la Data Science

Public : Data Scientist, Data Engineer, Data analyst et experts techniques

Durée : 2 jours, adaptable en fonction du besoin

Pré-requis : Pas de pré-requis

Méthodologie et Evaluation :

- Interventions sous formes d'exposés
- Mise en application de la théorie via des ateliers
- Questionnaire d'évaluation des connaissances
- Support de cours et exercices

Objectifs de la formation :

- Connaître l'intérêt de Python, les principes fondamentaux et son application à l'analyse de données
- Connaître les différents types de données et opérateurs
- Appliquer des techniques de nettoyage et pré-traitement des données
- Créer et interpréter des visualisations de données
- Développer des premiers modèles de Machine Learning

Programme :

Introduction

- Les bases de la programmation Python
- Introduction à la syntaxe de Python
- Vue d'ensemble des environnements de développement Python (Jupyter Notebooks, PyCharm).
- Configuration de l'environnement Python
- Installation de Python et des bibliothèques essentielles

Manipulation de données avec Pandas

- Exploration des objets Series et DataFrame.
- Indexation, sélection, filtrage et tri des données.
- Fusionner, joindre et concaténer des données.
- Techniques de nettoyage des données
- Gestion des valeurs manquantes, des données dupliquées et des types de données incorrects.
- Traitement des données textuelles et encodage des données catégorielles

Visualiser des données

- Introduction à Matplotlib, Seaborn et Plotly

- Création de graphiques de base : graphiques linéaires, diagrammes à barres et histogrammes.
- Visualisations avancées : diagrammes de dispersion, diagrammes en boîte et cartes thermiques.
- Analyse visuelle des données
- Utiliser les visualisations pour comprendre les distributions et les relations entre les données.

Les bases de l'apprentissage automatique

- Les différents opérateurs (addition, égalité...).
- Prétraitement des données pour l'apprentissage automatique.
- Vue d'ensemble de l'apprentissage supervisé et non supervisé.
- Modèles de régression, de classification et de clustering
- Séries temporelles