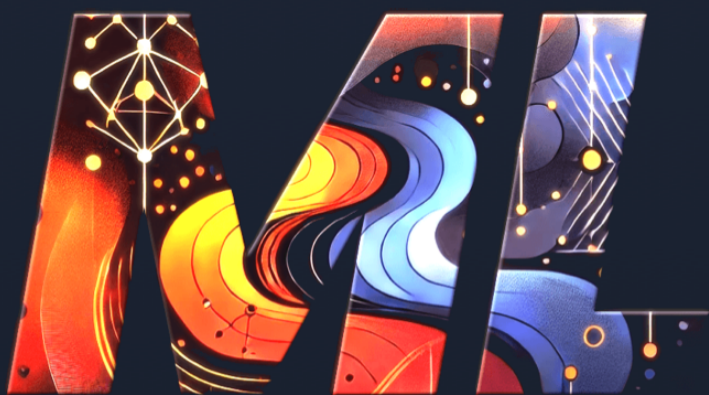


**ML-Pro**

Machine Learnia



# ML-Pro

---

Formation ML-PRO de Machine Learnia

Formateur : Guillaume Saint-Cirgue

# MODULES

**01**

Les 6 piliers de l'IA

**02**

Mathématiques

**03**

Machine Learning

**04**

Deep Learning

**05**

ML-OPS

**06**

Data Engineering

**07**

Cloud

**08**

Apprentissage par renforcement

**09**

IA générative

**10**

Aide à l'emploi

# Module 01 : Les 6 piliers de l'IA

Le **socle** de la formation. On y acquiert le raisonnement scientifique, l'approche de travail et la méthodologie pour résoudre des problèmes de Data Science, ainsi que la maîtrise des outils fondamentaux du métier.

**Bonus : Hacker votre apprentissage.** Les méthodes d'apprentissage efficaces pour progresser rapidement et durablement tout au long de la formation.

**Cours 1 : Prise en main de sa machine.** Comprendre son ordinateur et son système d'exploitation, prendre en main Linux et le terminal, gérer ses environnements virtuels, configurer un IDE (VSCode, PyCharm) et installer Git et GitHub. Mise en pratique.

**Cours 2 : La méthodologie de l'IA.** Le coeur méthodologique du métier de Data Scientist : analyse technique des données, analyse univariée et multivariée, tests statistiques, preprocessing, modélisation et évaluation. Cette méthodologie sert de fil conducteur à tous les projets de la formation.

**Cours 3 : Formation Python.** Apprentissage complet du langage Python, des bases jusqu'aux concepts avancés, avec de nombreux exercices de mise en pratique.

**Cours 4 : Formation SQL.** Manipulation des bases de données relationnelles : CRUD, calculs, jointures, fonctions, filtres et conditions, avec mise en pratique sur SQLite3.

**Cours 5 : Les frameworks de la data science.** Prise en main des bibliothèques essentielles de l'écosystème Python : Numpy, Pandas, Scipy, Matplotlib, Seaborn, Plotly, Polars, Streamlit, Scikit-learn, PyTorch, Keras et DuckDB.

**Cours 6 : Formation Git.** Gérer ses commits et ses branches, travailler avec des dépôts distants, résoudre les conflits, utiliser les issues, le fetch, les pull requests, le rebase et les tags, et collaborer efficacement en équipe.

# Module 02 : Mathématiques

Les **fondations mathématiques** nécessaires pour comprendre en profondeur le fonctionnement des algorithmes d'IA, avec mise en pratique et visualisations.

**Bonus : Prise en main de Manim.** Découverte de la bibliothèque d'animation mathématique Manim pour visualiser les concepts.

**Cours 1 : Les mathématiques de l'IA.** Introduction et panorama des mathématiques indispensables à l'intelligence artificielle.

**Cours 2 : Analyse et calcul mathématique.** Fonctions, dérivées et intégrales, suites, séries, récurrence et notions de distances.

**Cours 3 : Algèbre linéaire.** Vecteurs, matrices et espaces vectoriels, réduction de matrices, décomposition en valeurs singulières (SVD). Mise en pratique et applications de visualisation avec Manim.

**Cours 4 : Statistiques et probabilités.** Statistiques descriptives, lois de probabilités et variables aléatoires, avec exercices et mise en pratique.

# Module 03 : Machine Learning

Un cours détaillé sur le **fonctionnement des principaux algorithmes de Machine Learning**, avec leur théorie et leur mise en pratique avec Scikit-learn.

**Cours 1 : Modèles linéaires.** Régression linéaire simple et multiple, régularisation Lasso, Ridge et ElasticNet, et régression logistique.

**Cours 2 : Arbres de décision.** Algorithmes CART, ID3 et C4.5, critères du Gini et de l'entropie, déviance de Poisson, techniques de pre-pruning et post-pruning.

**Cours 3 : Les plus proches voisins.** Classification par KNN et RNN, régression par KNR et RNR.

**Cours 4 : Support Vector Machines.** SVC linéaire et à noyaux, SVC multi-classes, SVR linéaire et à noyaux, et nu-SVM.

**Cours 5 : Méthodes ensemblistes.** Random Forest, Extra Trees, AdaBoost, Gradient Boosting, XGBoost, LightGBM et Stacking.

**Cours 6 : Clustering.** KMeans, MeanShift, DBSCAN, Affinity Propagation, Spectral Clustering et Hierarchical Clustering.

**Cours 7 : Réduction de dimension.** PCA, t-SNE et Isomap.

# Module 04 : Deep Learning

De la consolidation des bases des **réseaux de neurones** jusqu'aux architectures modernes, avec mise en pratique sur TensorFlow, Keras et PyTorch.

**Cours 1 : Réseaux de neurones artificiels (ANN).** Batch learning, fonctions d'activation, méthodes d'initialisation, Batch Normalization, régularisation, algorithmes d'optimisation, multi-classification et régression.

**Cours 2 : Réseaux de neurones convolutifs (CNN).** Convolution, stride, padding et pooling, data augmentation et transfer learning, pour la vision par ordinateur.

**Cours 3 : Réseaux de neurones récurrents (RNN).** RNN, Deep RNN, Bidirectional RNN, GRU et LSTM, pour le traitement de séquences.

**Cours 4 : Les transformers.** Embedding, self-attention et cross-attention, connexions résiduelles.

**Cours 5 : Les autoencodeurs.** Variational, Denoising, Convolutional, Sparse et Contractive AutoEncoders.

# Module 05 : ML-OPS

Déployer, industrialiser et maintenir des modèles de Machine Learning en **production**.

**Cours 1 : Tour d'horizon du ML-OPS.** Objectifs du ML-OPS, DevOps, gestion des données, déploiement, surveillance et architectures.

**Cours 2 : Conteneurisation avec Docker.** Dockerfile, volumes, ports, Docker Compose, microservices, health checks, builds multi-stage, hot-reload, devcontainers et sécurité.

**Cours 3 : FastAPI.** Création d'APIs, notions HTTP, query et path parameters, documentation automatique, SQLAlchemy, Pydantic, authentification et JWT.

**Cours 4 : MLFlow.** Tracking des expériences, gestion des modèles et registry.

**Cours 5 : EvidentlyAI.** Monitoring des modèles et détection de data drift.

# Module 06 : Data Engineering

Concevoir et opérer des **pipelines de données** robustes. Un projet fil rouge, le Projet Vélib, accompagne l'ensemble du module.

**Cours 1 : Introduction au Data Engineering.** Outils typiques du métier, pipelines ETL et ELT, architectures, bases de données et data quality.

**Cours 2 : Stockage avec PostgreSQL.** SQL avec Postgres, vues, index, rôles et permissions, installation.

**Cours 3 : Transformation des données avec DBT.** Initialisation, modèles, sources, Jinja, matérialisation, configuration, documentation et tests, macros.

**Cours 4 : Orchestration avec Airflow.** DAGs, automatisation, alertes, assets et interface.

**Cours 5 : Data Viz, Storytelling et Dashboarding.** Les différents types de graphiques, les outils de dashboarding, et la construction du dashboard du projet Vélib.

**Cours 6 : Big Data avec Spark.** MapReduce, architecture, RDD, lazy evaluation et PySpark.

# Module 07 : Cloud

**Développer, déployer et maintenir** vos modèles de Machine Learning dans le monde réel, sur les grandes plateformes cloud.

**Cours 1 : Introduction au Cloud.** Les différents types de services (IaaS, PaaS, SaaS, FaaS, XaaS), les providers, les VPS, les data centers, le réseau, les serveurs, l'infrastructure et les services.

**Cours 2 : Amazon Web Services (AWS).** IAM, EC2, load balancing et Auto Scaling Groups, S3, bases de données, ECS, Lambda et monitoring.

**Cours 3 : Google Cloud Platform (GCP).** Services de Computer, Storage et BigQuery

# Module 08 : Apprentissage par renforcement

Les fondements et les algorithmes de l'**apprentissage par renforcement**.

**Cours 1 : Q-Learning**. Chaînes de Markov et processus de décision markoviens, les politiques, l'équation de Bellman, la Q-table et la stratégie epsilon-greedy.

**Cours 2 : Deep Q-Network (DQN) et Double DQN**. Approximation de la fonction de valeur par un réseau de neurones, experience replay, réseau cible et correction du biais de surestimation.

**Cours 3 : Les méthodes de policy gradient : A2C et PPO**. Optimisation directe de la politique, architectures acteur-critique, fonction d'avantage, Advantage Actor-Critic (A2C) et Proximal Policy Optimization (PPO).

# Module 09 : IA générative

Comprendre et exploiter les **modèles génératifs** modernes, avec utilisation et déploiement de modèles open source.

**Cours 1 : Large Language Models (LLMs).** Tokenisation, embeddings et mécanisme d'attention, inférence et paramètres de génération.

**Cours 2 : Les chaînes conversationnelles.** Prompts, mémoire de conversation, chaînage d'appels et orchestration avec LangChain.

**Cours 3 : Le RAG (Retrieval Augmented Generation).** Découpage des documents, embeddings, bases de données vectorielles, recherche sémantique et génération augmentée par la récupération.

**Cours 4 : Les agents.** Raisonnement, appel d'outils, planification, mémoire et enchaînement d'actions autonomes.

**Cours 5 : Les protocoles.** MCP (Model Context Protocol) et A2A (Agent to Agent) : standardiser la connexion des modèles aux outils, aux données et entre agents.

# Module 10 : Aide à l'emploi

Transformer **vos compétences en emploi**. Un accompagnement en continu, complété par des sessions en live, du choix de votre objectif jusqu'au débrief de vos entretiens.

**Étape 1 : Définir votre objectif.** Le métier visé, le secteur et la région : le point de départ qui conditionne tout le reste.

**Étape 2 : Ce que le marché demande vraiment.** Des études menées par nos équipes sur les offres d'emploi actuelles, pour identifier les compétences réellement attendues.

**Étape 3 : Les projets qui servent de preuve.** Choisir et construire les projets qui démontrent concrètement ce que vous savez faire.

**Étape 4 : Le CV.** Traduire votre parcours et vos projets en un CV lisible par les recruteurs de votre secteur.

**Étape 5 : Les candidatures.** Où postuler, comment, à quel rythme, et comment suivre vos candidatures.

**Étape 6 : Les entretiens et le débrief.** Préparer chaque entretien, puis le débriefer ensemble pour ajuster avant le suivant.

...