

# Ingénieur IA, parcours de 7 mois en continu tutorat inclus

Formation en ligne - 20h

Réf : 3BZ - Prix 2024 : 8 500€ HT

ORSYS et DataScientest, leader dans le domaine de la data science en France, proposent un parcours en ligne de 620 heures pour vous exercer au métier d'ingénieur IA. Vous y apprendrez notamment à récupérer et analyser des données liées au processus de l'entreprise, à élaborer des modèles prédictifs afin d'anticiper l'évolution des données et les tendances relatives à l'activité de l'entreprise, à modéliser des résultats pour les rendre plus lisibles et exploitables. Ce parcours en ligne aborde les fondamentaux du langage Python, l'usage de ses principales bibliothèques pour la data science, ainsi que les technologies autour de l'analyse de la donnée comme la data visualisation, le machine learning, le big data, le deep learning ou les systèmes complexes et l'intelligence artificielle (IA). Il vous permettra en plus d'obtenir la certification AWS Cloud Practitioner.

## OBJECTIFS PÉDAGOGIQUES

À l'issue de la formation l'apprenant sera en mesure de :

Connaître les fondamentaux du langage Python

Déployer son propre modèle de machine learning

Modéliser des résultats d'analyse des données pour les rendre lisibles et exploitables

Utiliser les technologies autour de la data science (data visualisation, machine learning, big data, deep learning, etc.)

Comprendre les systèmes complexes et l'IA

Créer des API avec Flask et FastAPI

Maîtriser les outils de versioning Git/GitHub et gérer des tests unitaires

Automatiser des tâches précises grâce à la maîtrise d'Apache Airflow

Maîtriser les techniques et outils de conteneurisation et d'orchestration des conteneurs

Faire preuve d'agilité et mener un projet data et une équipe

## PÉDAGOGIE ET PRATIQUES

Parcours en ligne basé sur une pédagogie active et conçu avec des experts. Une combinaison de théorie, de démonstrations, de mises en pratique, de partages d'expérience et de bonnes pratiques. Un test de positionnement, un accompagnement sur mesure dès le début du parcours, un projet fil rouge et des séquences pédagogiques de courte durée permettent de renforcer l'apprentissage et d'évaluer l'apprenant tout au long de son parcours. En cas de besoin technique, une cellule support est disponible en ligne 5 jours sur 7 de 9 h à 18 h 30. À l'issue de son parcours en ligne, l'apprenant soutient son projet devant un jury de 3 personnes. Il reçoit ensuite un certificat délivré par MINES ParisTech | PSL Executive Education, ce qui lui permet de bénéficier de la reconnaissance d'un acteur de référence dans le domaine. Il pourra aussi valider le titre RNCP "Chef de projet en Intelligence Artificielle".

## ACTIVITÉS DIGITALES

Test de positionnement sous forme de QCM d'une heure, séance d'introduction à la plateforme à distance, classes collectives, classe de soutien sur mesure, simulation et codage en direct, exercices, fiches de synthèse, projet fil rouge dédié, social learning, échanges avec data scientists, soutenance de projet devant un jury de 3 personnes.

## PARTICIPANTS

Toutes les personnes ayant une appétence pour la data souhaitant se reconverter ou faire évoluer ses compétences.

## PRÉREQUIS

Avoir un niveau Bac +3 en mathématiques ou un niveau Bac +5 en sciences. Des notions de marketing et de communication sont un plus.

## COMPÉTENCES DU FORMATEUR

Les experts qui ont conçu la formation et qui accompagnent les apprenants dans le cadre d'un tutorat sont des spécialistes des sujets traités. Ils ont été validés par nos équipes pédagogiques tant sur le plan des connaissances métiers que sur celui de la pédagogie, et ce pour chaque cours. Ils ont au minimum cinq à dix années d'expérience dans leur domaine et occupent ou ont occupé des postes à responsabilité en entreprise.

## MODALITÉS D'ÉVALUATION

La progression de l'apprenant est évaluée tout au long de sa formation au moyen de QCM, d'exercices pratiques, de tests ou d'échanges pédagogiques. Sa satisfaction est aussi évaluée à l'issue de sa formation grâce à un questionnaire.

## MOYENS PÉDAGOGIQUES ET TECHNIQUES

Les moyens pédagogiques et les méthodes d'enseignement utilisés sont principalement : documentation et support de cours, exercices pratiques d'application et corrigés des exercices, études de cas ou présentation de cas réels. ORSYS fournit aux participants un questionnaire d'évaluation du cours qui est ensuite analysé par nos équipes pédagogiques. Une attestation de fin de formation est fournie si l'apprenant a bien suivi la totalité de la formation.

## MODALITÉS ET DÉLAIS D'ACCÈS

L'inscription doit être finalisée 24 heures avant le début de la formation.

## ACCESSIBILITÉ AUX PERSONNES HANDICAPÉES

Vous avez un besoin spécifique d'accessibilité ? Contactez Mme FOSSE, référente handicap, à l'adresse suivante psh-accueil@orsys.fr pour étudier au mieux votre demande et sa faisabilité.

# LE PROGRAMME

---

dernière mise à jour : 09/2023

## 1) Fondamentaux de Python

- Découverte des différents types de variables, des listes et des tuples.
- Présentation des divers opérateurs et structures de contrôle.
- Initiation au concept de boucle et ses différents types.
- Introduction aux fonctions et leur documentation.
- Instanciation de classes et utilisation de modules.

## 2) Bibliothèques NumPy et Pandas

- Création et manipulation d'un tableau NumPy array.
- Présentation des opérations matricielles et manipulation d'un NumPy array.
- Création d'un indicateur statistique et opérations sur un NumPy array.
- Introduction à la bibliothèque Pandas.
- Chargement et première exploration d'un jeu de données.
- Introduction au data cleaning.
- Introduction au data processing.
- En option : data quality (principes, variables numériques et textuelles, nettoyage de données, valeurs manquantes).

## 3) Matplotlib et Seaborn

- Matplotlib, présentation des différents types de graphes.
- Matplotlib, introduction à la personnalisation des graphiques.
- Seaborn, maîtrise de l'analyse des distributions.
- Seaborn, mise en place de l'analyse statistique.
- Seaborn, initiation à l'analyse multivariée.
- En option : Bokeh (graphiques interactifs et intégrables sur page web, données géographiques, widgets).

## 4) Machine learning supervisé

- Introduction à Scikit-learn et présentation de la classification support vector machines (SVM).
- Application de la méthode des k plus proches voisins (k-nearest neighbors, ou KNN)
- Sélection de modèles et classification semi-supervisée.
- Présentation des algorithmes des k-moyennes.
- Découvertes des algorithmes de mean shift.
- Introduction à la régression linéaire simple et multiple.
- Introduction à la régression linéaire régularisée.
- Feature selection process.
- Initiation aux analyses en composantes principales.
- Application de l'approche manifold learning.

## 5) Machine learning avancé

- Découverte des modèles de base.
- Découverte des modèles ARIMA.
- Text mining, introduction aux expressions régulières.
- Text mining, gestion de données textuelles.
- Text mining, création de word clouds.

## 6) Machine learning et théorie des graphes avec NetworkX

- Introduction à la théorie des graphes.
- Application des algorithmes fondamentaux : Kruskal et Dijkstra.

- Détection de communautés.
- Application de l'algorithme PageRank au classement de pages web.

## 7) Big data, database

- Data engineering et big data, introduction aux 3V du big data (volume, variété, vélocité).
- Data engineering et big data, présentation des bases de données et des architectures.
- Data engineering et big data, mise en production de données.
- Data processing et machine learning sur des grandes bases de données, introduction à PySpark.
- Découverte des fonctions de PySpark : data processing, DataFrames, régression avec PySpark, ML pipelines, model turning.

## 8) Deep learning

- Deep learning avec le framework Keras (Dense Neural Networks, Convolutional Neural Networks)
- Deep learning avec le framework Keras (Architecture LeNet, Transfer Learning).
- Tensorflow 2.0, mise en relation de Tensorflow et Keras.
- Tensorflow 2.0, application du word embedding avec Word2vec.
- Tensorflow 2.0, présentation du Recurrent Neural Network.
- Tensorflow 2.0, présentation du Generative adversarial Network

## 9) Système complexe et IA

- Définition des éléments de l'apprentissage par renforcement.
- Mise au point sur les fondamentaux mathématiques pour le reinforcement learning.
- Présentation de la principale famille d'algorithmes de reinforcement learning : Monte-Carlo.
- Application des algorithmes d'apprentissage par différence temporelle : SARSA.
- Application des algorithmes d'apprentissage par différence temporelle : Q-learning.
- Comparaison entre apprentissage par différence temporelle et Monte-Carlo.
- Introduction au deep reinforcement learning.
- Compréhension des fondements des méthodes basées sur la valeur.
- Compréhension des problèmes de variance élevée, des risques de surestimation, application des méthodes de résolution.
- Introduction à la méthode de policy gradient.
- Principes fondamentaux de la structure Acteur-Critique utilisée dans l'apprentissage en deep reinforcement learning.
- Application à des business cases.

## 10) Programmation avancée

- Webscraping, introduction au langage web (HTML, CSS).
- Webscraping, extraction de contenu web avec BeautifulSoup.
- Webscraping, application du scrapping sur Google.
- Présentation des systèmes Linux.
- Prise en main et utilisation d'un terminal.
- Mise en place de scripts Bash.
- Introduction au système de gestion de version Git.
- Initialisation d'un dépôt Git.
- Présentation et approfondissement des concepts Git (Branches, Tag, Merge).
- Découverte de la plateforme GitHub pour le travail collaboratif sur Git.
- Présentation des fonctionnalités majeures de GitHub (Fork, Pull Request, Issues).
- Partage des modifications avec pull et push.
- Participation à l'amélioration de projet public (open source).
- Présentation des principaux workflows Git.
- Introduction aux APIs et découverte des architectures micro-services.
- Présentation des différentes méthodes HTTP et de leurs fonctions.
- Utilisation des bibliothèques FastAPI et Flask pour développer des API RESTful.
- Documentation d'une API avec la spécification OpenAPI.

- Gestion des erreurs et des performances d'une API.
- Présentation de la conteneurisation et de son utilité par rapport à la virtualisation.
- Initiation au fonctionnement de Docker.
- Manipulation des images et des conteneurs.
- Communication avec les conteneurs.
- Persistance des données grâce aux volumes.
- Création d'une image Docker via un Dockerfile.
- Partage des images sur le Dockerhub .

### 11) DataOps, orchestration

- Kubernetes : déployer et gérer des conteneurs.
- Airflow : orchestration et utilité, graphe orienté acyclique ou DAG (Directed Acyclic Graphs), Opérateurs.
- Gestion de tâches par le biais d'Operators spécifiques.
- Monitoring des DAG via l'interface graphique d'Airflow.

### 12) ModelOps

- MLflow tracking : chargement du projet, importation des données à tracker, automatisation et UI.
- MLflow projects : empaquetage de code dans un format réutilisable et reproductible, Git et collaboration.
- MLflow models : déployer le modèle localement et dans un serveur HTTP simple.
- MLflow Registry : création du répertoire, enregistrement du meilleur modèle, archivage et annotation du modèle.
- Intégration : système de stockage, API et Spark.

### 13) Business intelligence

- Introduction à la BI.
- Power BI.
- Make.

### 14) Data gouvernance

- Gouvernance des données.
- RGPD et éthique dans la data.

### 15) Chefferie de projet

- Gestion de projet et conduite du changement.
- Méthode Agile.
- Application.

## NOS POINTS FORTS

---

- Séquences de courte durée
- Activités digitales variées
- Accès illimité pendant 1 an ou pendant la durée du parcours
- Tutorat personnalisé inclus ou en option
- Accès multi-device (smartphone, tablette ou ordinateur)