

**Domaine :** Intelligence Artificielle - **Thématique(s) :** IA, Data et calcul  
DIPLÔMES D'UNIVERSITÉ (DU, DIU)

## MACHINE LEARNING ET INTELLIGENCE ARTIFICIELLE

**🕒 Durée de la formation :** 98h

**📅 Dates :** Voir le calendrier

**📍 Lieu :** Campus Pierre et Marie Curie – Paris (Jussieu)

**€ Tarif :** Voir ci-dessous

**Modalité :** Présentiel

**CPF :** Éligible

**Formation :** Diplômante

8000€ + 243€ frais inscriptions administrative en 2022-2023

### OBJECTIFS

La certification Machine Learning et Intelligence artificielle a pour objectif de compléter et enrichir les compétences des salariés des secteurs de l'exploitation de données massives et de permettre la réalisation d'analyses prédictives répondant à la problématique de l'entreprise.

La spécificité de cette formation tient dans l'association à parts égales entre informatique et mathématiques, tout en s'attachant à conserver un fort aspect professionnalisant.

Cette formation doit permettre à ses diplômés une évolution professionnelle rapide dans les secteurs de l'analyse des données et des technologies de l'information. En associant étroitement une dimension informatique, mathématique, et connaissance métier, elle confèrera aux stagiaires une capacité à la fois à modéliser, à coder, et à analyser les résultats et conséquences d'une démarche de data scientist. Elle apportera notamment une capacité à ses diplômés à se mouvoir dans le secteur du conseil (où une demande croissante en experts des données est ressentie), ou à développer de nouvelles approches industrielles au sein d'une entreprise qui cherche à valoriser ses données.

Cette certification s'appuie sur l'un des 34 plans de la Nouvelle France Industrielle consacré au Big Data qui s'est transformé en solution Economie des Données dont l'objectif 2020 est de créer et consolider 137000 emplois.

### COMPÉTENCES VISÉES

À l'obtention de la certification, le stagiaire sera en capacité de :

- Maîtriser les fondements mathématiques et informatiques pour le traitement de données
- Associer des tâches de machine learning à des problématiques métier
- Identifier les besoins algorithmiques pour appréhender ces tâches
- Développer une architecture de Deep learning
- Comprendre les enjeux sociétaux de l'intelligence artificielle
- Développer une architecture Big Data
- Mener à bien un projet data

### ACCOMPAGNEMENT DE LA FORMATION

Suivi personnalisé, articulé autour de séances de regroupement en présence d'un ou plusieurs enseignants référents.

### INFORMATIONS

Finançable par le CPF

**Session 1** DU 27/01/2023 AU 07/07/2023

#### CONTACT

📞 01 44 27 82 55

✉ sciences-ftlv-fpc@sorbonne-  
universite.fr

Contact :  
roselyne.friedenberg\_remy@sorbonne-  
universite.fr

## PUBLIC VISÉ ET PRÉ-REQUIS

### Public

Cette formation s'adresse à des profils variés, tous réunis autour d'une forte motivation à acquérir des compétences en science des données.

Cette formation s'adresse à des techniciens et cadres possédant un bon niveau en mathématiques et/ou en informatique, et souhaitant développer des compétences en machine learning, en deep learning, en intelligence artificielle ou en traitement de données massives.

## PROGRAMME

### Module 1 : Outils de Mathématiques et d'informatique

L'objectif de ce module est de remettre à niveau les étudiants en terme de compétences en mathématiques et informatique. Les aspects abordés sont l'algorithmique de base, la maîtrise du langage Python et des bibliothèques de traitement de données (numpy, pandas, matplotlib, bokeh, ...), ainsi que la gestion des entrées/sorties ainsi que l'analyse (intégration), l'algèbre linéaire (produit et inversion de matrices, valeurs propres et singulières), les probabilités et statistiques de base (espace probabilisé, échantillonnage, modèle statistique, inférence).

**Organisation** : 2 jours de conférence/mise en situation.

### Module 2 : Apprentissage statistique supervisé

Ce module dresse un panorama de l'apprentissage statistique aujourd'hui. Il aborde successivement les grandes problématiques du domaine et en présente les avancées majeures des dix dernières années avec un focus sur les modèles supervisés.

Cadre général de l'apprentissage supervisé

Classifieur bayésien naïf

Sur-apprentissage et validation croisée

Arbres : CART, forêts aléatoire et boosting

Minimisation du risque empirique

Régression logistique et machines à vecteurs supports

Perceptron, réseaux de neurones

Approches régularisées

Traitement de données incertaines et incomplètes

**Organisation** : 2 jours de conférence/mise en situation.

### Module 3 : Apprentissage non-supervisé et apprentissage profond

Cet enseignement se focalise sur les modèles de réseaux de neurones profonds et les outils informatiques associés:

Algorithmes de descente de gradient, back-propagation et variantes

Réseaux de neurones convolutionnels et réseaux récurrents

Outils: Plateforme déclarative (TensorFlow) et algorithmique (PyTorch)

**Organisation** : 2 jours de conférence/mise en situation.

### Module 4 : Apprentissage profond et réseaux de neurones pour les images

L'objectif de ce module est d'étudier les réseaux de neurones profonds pour la reconnaissance automatique d'images, ainsi que les algorithmiques classiques utilisés en apprentissage non-supervisé.

Descripteurs SIFT et architectures de réseaux de neurones appliqués aux images (réseaux convolutifs, ...)

Clustering: classification hiérarchique, k-means et ses variantes, modèles de mélange gaussiens, méthodes spectrales

Réduction de dimension et visualisation : décomposition en valeurs singulières, ACP et ses variantes, ACM, MDS, ISOMAP, t-SNE

**Organisation** : 2 jours de conférence/mise en situation.

### Module 5 : Réseaux de neurones pour le traitement du langage et environnement des données

Ce module vise étudier les réseaux de neurones appliqués aux données textuelles et à appréhender le contexte global dans lequel s'inscrivent les

approches industrielles liées au Big Data. Il sensibilise notamment les participants aux notions de réglementation et d'éthique, et insiste sur l'intégration des méthodologies de data science dans un environnement professionnel

Représentations de mots et architectures de réseaux de neurones appliqués au texte (réseaux récurrents, d'attention et « adversariaux »).

Gouvernance et éthique des données

Cyber-sécurité, business Analytics

Communication, « data story telling »

**Organisation** : 2 jours de conférence/mise en situation.

### **Module 6 : Cloud Computing et Big Data**

L'objectif de ce module est de former les étudiants aux outils autour de la gestion de données massives. C'est un enseignement orienté outils informatique et infrastructure qui recouvre les aspects:

Bases de données SQL et NoSQL

Hadoop, Spark, MapReduce

Cloud Computing et virtualisation

Pig, Hive et SPARQL

**Organisation** : 2 jours de conférence/mise en situation.

### **Module 7 : Applications (et déclinaisons par secteur d'activité)**

L'objectif de ce module est d'aborder des applications particulières nécessitant la mise en oeuvre d'outils mathématiques et informatique particuliers. Ce module sera organisé sous forme de conférences, en fonction du profil des étudiants.

Les thèmes seront choisis parmi les applications suivantes :

Recommandation

Graphes et réseaux sociaux

Text Mining

Publicité

Intelligence Artificielle

Actuariat

Finance

Gestion des risques

...

**Organisation** : 2 jours de conférence.

### **Module 8 : Projet Data**

L'objectif de l'UE est d'encadrer les étudiants autour de projets de traitement de données. Les sujets des projets et les données à traiter seront proposés par les étudiants eux-mêmes. Un effort particulier sera porté sur la restitution (Data story telling) des résultats et des modèles proposés.

**Organisation** : Suivi personnalisé, articulé autour de séances de regroupement en présence d'un ou plusieurs enseignants référents.

---

## MODALITÉS D'ÉVALUATION

Les modalités d'évaluation de la certification s'appuient d'une part sur le contrôle continu des connaissances dans chacun des 8 modules et d'autre part sur la réalisation d'un projet et sa soutenance devant un jury.

Le contrôle continu s'effectue sous forme de quiz ou exercices notés. Il est nécessaire d'obtenir une note supérieure ou égale à 10/20 à l'évaluation des connaissances et une note supérieure ou égale à 10/20 à l'évaluation du projet pour se voir attribuer la certification.

### **Validation**

La certification, délivrée le jury de la formation, est un diplôme d'Université de Sorbonne Université

---

## DÉBOUCHÉS

Cette formation permet aux individus de sécuriser leur parcours professionnel en leur donnant les compétences nécessaires pour accompagner les entreprises dans les enjeux liés à leur secteur d'activité et s'adapter aux évolutions technologiques associées.

## LES + DE LA FORMATION

Cette formation s'adresse à des profils variés, tous réunis autour d'une forte motivation à acquérir des compétences en science des données.

### POUR CANDIDATER

Les participants seront sélectionnés sur dossier et au besoin sur entretien individuel.

Ledit dossier devra contenir un CV, un questionnaire relatif à l'entreprise et au projet professionnel (à demander lors de votre prise de contact par courriel) ainsi qu'une lettre de motivation.

La commission se réunit deux fois par mois pour statuer sur les candidatures.

Prise en charge CPF : RS5566 – Sciences des données et apprentissage profond au service de l'intelligence artificielle

Lien d'accès : <https://www.francecompetences.fr/recherche/rs/5566/>

### CALENDRIER

**Durée de la formation** : 98h

**Rythme** : 2 jours /mois

• 2 jours /mois vendredi et samedi • Cours et mise en situation

**SESSION 1**

du 27/01/2023  
au 07/07/2023

Campus Pierre et Marie Curie

Présentiel