

FORMATION PYTHON

perfectionnement et usages avancés

Le langage informatique Python s'est imposé comme langage de référence dans de nombreux domaines d'activité, notamment pour tous les métiers touchant de près ou de loin à la data science. L'apprentissage du Python est donc un investissement très judicieux et rentable rapidement.

Durée : 6 jours – 36 heures

Coût : 4000 € HT / salarié

Format : Inter-entreprise et Intra-entreprise

Localisation : A distance

PRÉ-REQUIS

Les apprenants doivent absolument maîtriser les points suivants :

- Principe de base de la programmation (variables, structures de contrôle, fonctions, etc...)
- Principe de base de la POO (Programmation Orientée Objet)
- Connaissance basique de la syntaxe Python

PUBLIC

La formation s'adresse à des profils :

- Ingénieurs ou spécialistes métier dans les domaines où l'importance du traitement des données est de plus en plus manifeste (actuaire, analystes, développeurs)

OBJECTIFS

À l'issue de cette formation, l'apprenant sera capable de :

- Être autonome en Python, d'implémenter des scripts et des programmes de traitement de données
- Maîtriser quelques bibliothèques adaptées à ces besoins

Classement des langages de programmation ouvrant le plus d'opportunités d'emploi

Rank	Language	Type	Score
1	Python▼	🌐 📱 💻	100.0
2	C▼	📱 💻	98.0
3	Java▼	🌐 📱 📺	97.1
4	Go▼	🌐 📺	87.2
5	C++▼	📱 💻	85.2
6	JavaScript▼	🌐	81.8
7	R▼	📺	80.4
8	HTML▼	🌐	75.7
9	Swift▼	📱 📺	70.1
10	SQL▼	📺	69.4

- Respecter les bonnes pratiques de code (PEP8)

Cette formation vise à valoriser le profil des stagiaires et améliorer leurs perspectives d'évolution professionnelle

MODALITÉS ET DÉLAIS D'ACCÈS

- Validation de l'inscription par bon de commande et convention de formation.
- Compter un délai de deux semaines pour l'organisation d'une formation pour un minimum de 5 apprenants (en-dessous de 5 personnes, sur devis personnalisé).

MÉTHODES MOBILISÉES

- Cours magistral sur support numérique
- Mise en pratique, exercices guidés de programmation
- Mise en situation sur un problème représentatif.

MODALITÉS D'ÉVALUATION

- En amont de la formation : Positionnement et validation des pré-requis et du profil du ou des apprenants (par téléphone ou par mail), confirmation avec un questionnaire numérique de positionnement,
- Evaluation continue par les formateurs par le biais de quizz et de rétroactions avec les apprenants.
- En sortie de formation : QCM d'évaluation.

ACCESSIBILITÉ

- La formation se déroulera à distance dans le cadre d'une formation inter-entreprise
- Dans le cas d'une formation intra-entreprise, la formation peut se dérouler dans les locaux de celle-ci.
- Les formations sont pensées pour être les plus inclusives possible permettant au plus grand nombre d'en tirer profit.

PLAN DE FORMATION :

Introduction : (2 jours)

- Programmation orientée Objet, définitions, rappels et intérêt
- Rappels de Python, syntaxe et concepts clés

- Typage et générateurs
- Gestions des modules (Pypy et virtualenv)
- Notions de programmation concurrente et parallélisation

Manipulation de données : (1 Jours)

- Manipulation de fichiers
- Parsing de xml, json
- Présentation de Pickle

Librairies classiques : (2 Jours)

- Pandas, présentation et utilisation pour la manipulation de données tabulaires
- Numpy, application aux calculs matriciels et vectoriels
- Sklearn, aperçu de méthodes de régression et de classification simples

Bonnes pratiques de code : (1 Jour)

- Documentation avec Doxygen
- Utilisation des exceptions
- Présentation de PEP8

FORMATION IA POUR LE DROIT

Conception, entraînement et mise en production de systèmes de machine learning appliqués au droit.

Durée : 4 jours – 16 heures

Coût : 5000 € HT / salarié

Format : Inter-entreprise et intra-entreprise

OBJECTIFS

- Familiariser les apprenants avec les différentes étapes de la procédure d'ingestion de documents.
- Maîtriser la représentation d'un texte sous forme vectorielle, tant au niveau du vocabulaire que du contexte sémantique
- Initier les apprenants aux concepts et techniques spécifiques, permettant de faire comprendre du texte à une machine
- Former les stagiaires à la mise en production des techniques enseignées

MODALITÉS ET DÉLAIS D'ACCÈS

- Positionnement et validation des pré-requis et du profil du ou des apprenants (par téléphone ou par mail), confirmation avec un questionnaire numérique de positionnement, validation de l'inscription par bon de commande et convention de formation.
- Compter un délai de deux semaines pour l'organisation d'une formation pour un minimum de 5 apprenants (en-dessous de 5 personnes, sur devis personnalisé).

MÉTHODES MOBILISÉES

- Cours magistral sur support numérique
- Mise en pratique, exercices guidés de programmation
- Mise en situation sur un problème représentatif.

MODALITÉS D'ÉVALUATION

- Evaluation continue par les formateurs par le biais de quizz et de rétroactions avec les apprenants.
- En sortie de formation : QCM d'évaluation.

ACCESSIBILITÉ

- La formation se déroulera à distance dans le cadre d'une formation inter-entreprise
- Dans le cas d'une formation intra-entreprise, la formation peut se dérouler dans les locaux de celle-ci.
- Les formations sont pensées pour être les plus inclusives possible permettant au plus grand nombre d'en tirer profit.

PLAN DE FORMATION :

Pipeline de données : Comment passer d'une image contenant du texte à une information utilisable numériquement par des juristes ? : (2h)

- Théorie (1h) : Base de données structurée, principe de la technologie d'OCR, collecte de documents
- Pratique (1h) : Utilisation d'un logiciel libre d'OCR (Tesseract)

Représentation du langage - Les Word embeddings : (4h30)

- Théorie (2h) : Notion de tokens et de lemmatisation, espace et champs de vecteurs, distance et mesures, notion de représentation et de représentation optimale, algorithme de type TF-IDF (Term Frequency-Inverse Document Frequency), word embedding
- Pratique (2h30) : Réalisation d'un algorithme de TF-IDF pour la recherche d'informations dans un document, réalisation d'un word embedding à partir d'un corpus de textes juridiques par une méthode de bag-of-words

Natural Language Processing - Concepts et applications : (4h30)

- Théorie (2h) : Classification et régression, champs aléatoires conditionnels, réseaux de neurones séquentiels (LSTM, CNN 1D)
- Pratique (2h30) : Réalisation d'un NER (Named Entity Recognition) tagger, permettant l'identification d'entités nommées, par exemple pour anonymiser ou pseudonymiser un document

Déploiement et mise en production : (5h)

- Théorie (1h) : Environnement virtuel, prise en main de git, Docker, calcul sur GPU, différences avec des CPU, Cython, notions de Cloud et d'API (Application Programming Interface)
- Pratique (4h) : Mise en production d'un projet de machine learning miniature : création d'environnement virtuel, utilisation de git, déploiement sur GPU, en local (sur serveurs Magic LEMP) et dans le Cloud (compte AWS), cythonisation, utilisation de Flask pour API

USERS: LAST 7 DAYS USING MEDIAN ▾

EXPLOITATION ET VALORISATION DE DONNEES PAR INTELLIGENCE ARTIFICIELLE



Des données brutes à la création de valeur.

Cette formation complète est découpée en modules indépendants (mais s'enchaînant de manière logique). Il est donc possible d'adapter la composition globale de la formation sur mesure pour chaque employé intéressé, en fonction de son niveau de maîtrise des objets et concepts de science des données présentés.

Des ajustements de tarifs et de contenus peuvent par ailleurs être envisagés pour la réalisation de la formation complète sur devis ultérieur.

MODULES DE LA FORMATION

- 1] PYTHON POUR LES DONNEES (1.5 jours – 1000 € HT / apprenant)
- 2] INGENIERIE DES DONNEES (2 jours – 1300 € HT / apprenant)
- 3] ANALYSE ET VISUALISATION DES DONNEES (2 jours – 1800 € HT / apprenant)
- 4] MACHINE LEARNING/ INTELLIGENCE ARTIFICIELLE (4 jours – 4000 € HT / apprenant)
- 5] DU PoC A LA PROD (passer d'un prototype à un produit rentable) (2 jours – 2300 € HT / apprenant)

MODALITÉS ET DÉLAIS D'ACCÈS

- Positionnement et validation des pré-requis et du profil du ou des apprenants (par téléphone ou par mail), confirmation avec un questionnaire de positionnement numérique, validation de l'inscription par bon de commande et convention de formation.
- Compter un délai de deux semaines pour l'organisation d'une formation pour un minimum de 5 apprenants (en-dessous de 5 personnes, sur devis personnalisé).

MÉTHODES MOBILISÉES

- Cours magistral sur support numérique
- Mise en pratique, exercices guidés de programmation
- Mise en situation sur un problème représentatif.

MODALITÉS D'ÉVALUATION

- Evaluation continue par les formateurs.
- En sortie de formation : QCM d'évaluation.

ACCESSIBILITÉ

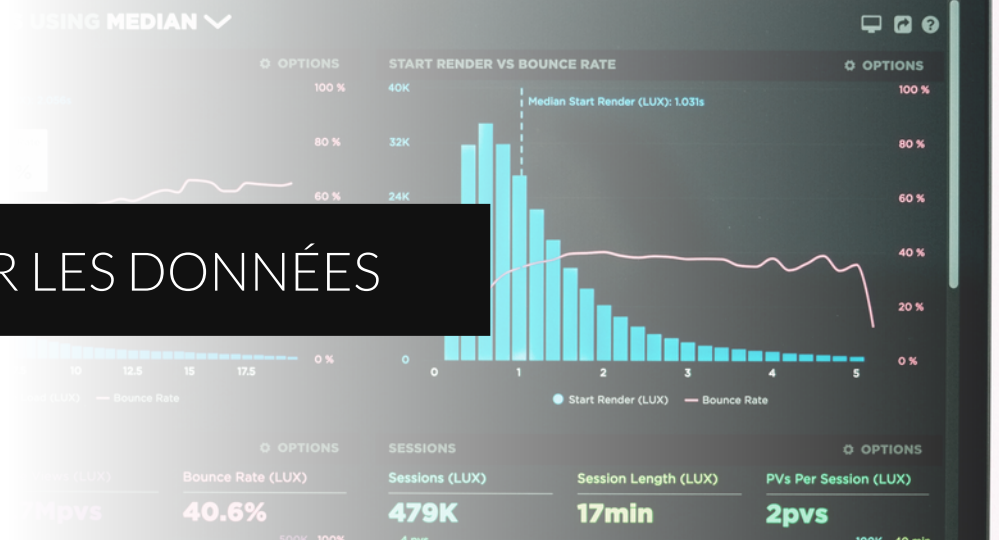
- La formation se déroulera à distance dans le cadre d'une formation inter-entreprise
- Dans le cas d'une formation intra-entreprise, la formation peut se dérouler dans les locaux de celle-ci.

PUBLIC

- Ingénieurs et spécialistes métier dans les domaines où l'importance du traitement des données est de plus en plus manifeste (actuaire, analystes, développeurs), etc.)

MODULE 1

PYTHON POUR LES DONNÉES



CONTENU

- Rappels de base sur les architectures de base de données (SQL, formats de données)
- Utilisation basique de Python pour générer et lire des bases de données
- Présentation et utilisation de librairies de manipulation de données (Pandas, Numpy)
- Optimisation de base de données

Durée : 1.5 jours – 12 heures

Coût : 1000 € HT / salarié
(au-delà de 10 participants, nous contacter)

PRÉ-REQUIS

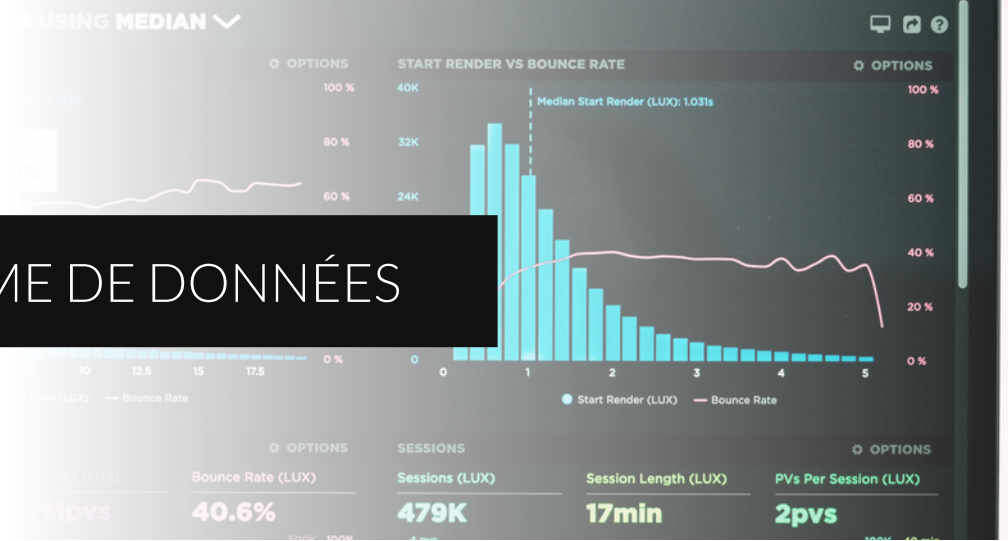
- Principes de base de la programmation (variables, structures de contrôle, fonctions, etc.)
- Formation Python perfectionnement et usages avancés (recommandée mais optionnelle), à défaut bonne connaissance du Python

OBJECTIFS

- Rendre les apprenants autonomes pour manipuler et gérer les données et les entrepôts de données
- Maîtriser quelques librairies adaptées aux besoins des apprenants
- Acquisition des pré-requis nécessaires au module de formation « Mise en forme de données ».

MODULE 2

MISE EN FORME DE DONNÉES



CONTENU

- Gestion, ingestion et récupération de données de tous types (scrapping, formatage de données)
- Normalisation et standardisation de données tabulaires
- Familiarisation avec des librairies adaptées (Scikit-learn, Numpy)
- Prise en charge de données non-structurées (images, textes, vidéos, etc.)

Durée : 2 jours – 16 heures

Coût : 1300 € HT / salarié
(au-delà de 10 participants, nous contacter)

PRÉ-REQUIS

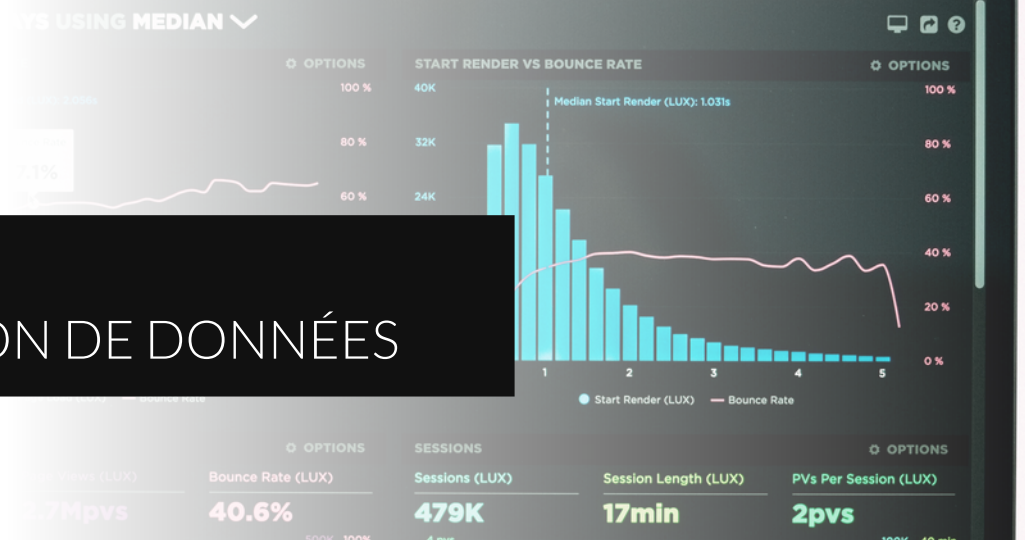
- Formation « Python perfectionnement et usages avancés » (recommandée mais optionnelle), à défaut bonne connaissance du Python,
- Module 1 « Python pour les données » (recommandé mais optionnel), à défaut connaissance du Python pour données et sur l'utilisation et la gestion de bases de données

OBJECTIFS

- Rendre les apprenants capables de constituer des bases de données utilisables pour des approches d'analyse de données et d'intelligence artificielle
- Maîtriser quelques librairies adaptées aux besoins des apprenants
- Acquisition des pré-requis nécessaires au module « Analyse et visualisation de données »

MODULE 3

ANALYSE ET VISUALISATION DE DONNÉES



CONTENU

- Rappels de statistiques et probabilités
- Utilisation de notebook Jupyter
- Présentation de méthodes d'analyse de données multidimensionnelles (analyse en composantes principales, analyse factorielle des correspondances, classification hiérarchique, etc.)
- Présentation d'outils de visualisation simples d'utilisation et puissants (Matplotlib, Plotly, Bokeh, Seaborn)
- Evaluer la pertinence des données à visualiser
- Éléments de *features engineering*

Durée : 2 jours – 16 heures

Coût : 1800 € HT / salarié
(au-delà de 10 participants,
nous contacter)

PRÉ-REQUIS

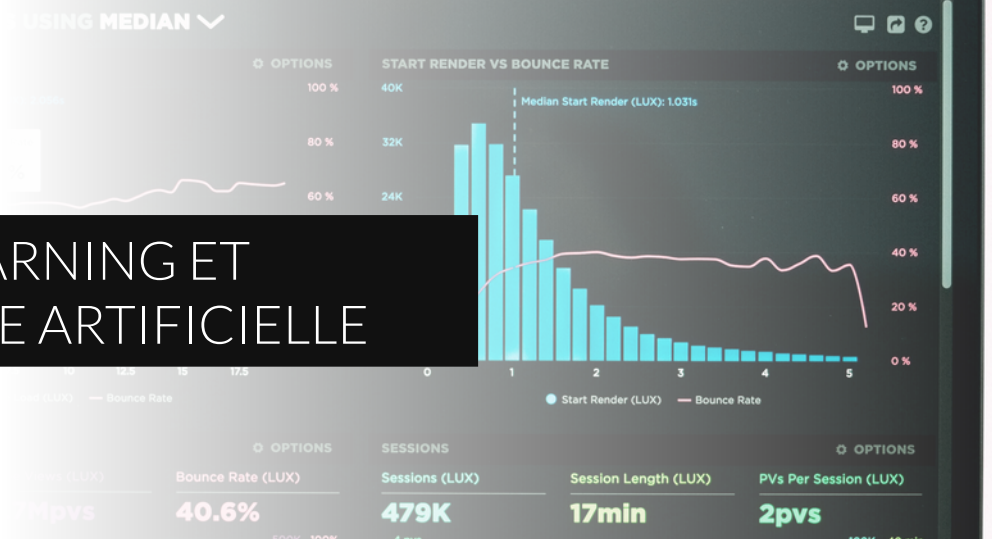
- Formation « Python perfectionnement et usages avancés » (recommandée mais optionnelle), à défaut bonne connaissance du Python,
- Module 1 « Python pour les données » (recommandé mais optionnel), à défaut connaissance du Python pour données et sur l'utilisation et la gestion de bases de données
- Module 2 « Mise en forme de données » (recommandé mais optionnel), à défaut capacité à récupérer, stocker et formater les données d'intérêt
- Notions basiques de statistiques

OBJECTIFS

- Rendre les apprenants capables d'analyser différents types de données
- Rendre les apprenants capables de visualiser et d'identifier les variables pertinentes d'un jeu de données
- Acquisition des pré-requis nécessaires à la conception d'algorithmes de machine learning/intelligence artificielle

MODULE 4

MACHINE LEARNING ET INTELLIGENCE ARTIFICIELLE



CONTENU

- Introduction aux méthodes de machine learning : définitions, présentation de quelques algorithmes classiques
- Présentation des différents membres de la famille du machine learning (bagging, random forest, deep learning)
- Rappels d'analyse fonctionnelle et d'algèbre
- Problèmes classiques de régression et de classification
- Deep learning (LSTM, GRU, ConvNet, etc.)
- Familiarisation avec les librairies classiques (Scikit-learn, Xgboost, Keras, Pytorch)
- *Application à un problème spécifique à l'entreprise allant jusqu'à la réalisation d'un PoC (sur devis)*

Durée : 4 jours – 32 heures

Coût : 4000 € HT / salarié
(au-delà de 10 participants,
nous contacter)

PRÉ-REQUIS

- Formation « Python perfectionnement et usages avancés » (recommandée mais optionnelle), à défaut bonne connaissance du Python,
- Module 1 « Python pour les données » (recommandé mais optionnel), à défaut connaissance du Python pour données et sur l'utilisation et la gestion de bases de données
- Module 2 « Mise en forme de données » (recommandé mais optionnel), à défaut capacité à récupérer, stocker et formater les données d'intérêt
- Module 3 « Analyse et visualisation de données » (recommandé mais optionnel), à défaut capacité à évaluer, à pré-traiter les données et à en faire des visualisations pertinentes
- Notions basiques de statistiques et de probabilités

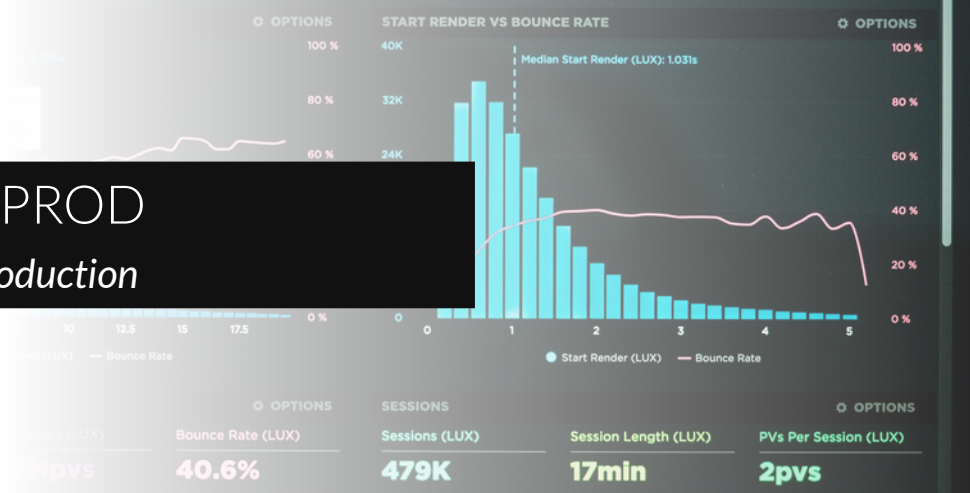
OBJECTIFS

- Donner des outils concrets d'intelligence artificielle applicables aux besoins de l'entreprise
- Démystifier l'intelligence artificielle et donner des outils mathématiques simples mais efficaces permettant la compréhension de la portée et des limites des approches de machine learning enseignées
- Acquisition des compétences nécessaires à la réalisation d'un prototype (PoC) adapté aux besoins métier

MODULE 5

DU POC À LA PROD

Du prototype à la production



CONTENU

- Versionnement et bonnes pratiques (tests unitaires, robustesse)
- Déploiement d'un environnement de production (d'un modèle à une API)
- Mise en place d'un pipeline pour récupérer et traiter le flux de données réelles
- Choix de l'infrastructure adaptée aux besoins de l'entreprise (AWS, OVH, serveur local, etc.)

Durée : 2 jours – 16 heures

Coût : 2300 € HT / salarié
(au-delà de 10 participants, nous contacter)

PRÉ-REQUIS

- Formation « Python perfectionnement et usages avancés » (recommandée mais optionnelle), à défaut bonne connaissance du Python,
- Module 1 « Python pour les données » (recommandé mais optionnel), à défaut connaissance du Python pour données et sur l'utilisation et la gestion de bases de données
- Module 2 « Mise en forme de données » (recommandé mais optionnel), à défaut capacité à récupérer, stocker et formater les données d'intérêt
- Module 3 « Analyse et visualisation de données » (recommandé mais optionnel), à défaut capacité à évaluer, à pré-traiter les données et à en faire des visualisations pertinentes
- Module 4 « Machine learning et intelligence artificielle » (recommandé mais optionnel), à défaut connaître les techniques de machine learning adaptées aux données et au problème, et savoir développer un modèle pertinent de machine learning
- Notions sur les systèmes de versionnement
- Notions d'architecture système

OBJECTIFS

- Permettre le passage d'un modèle de machine learning à un service utilisable par l'entreprise
- Générer des retours sur investissement concrets quant à l'utilisation d'intelligence artificielle par l'entreprise

- Former les salariés aux bonnes pratiques de mise en production d'un service complexe

FORMATEURS

- Antonin Penon, spécialiste de machine learning, directeur technique de Magic LEMP, très expérimenté dans plusieurs langages de programmation et en Python en particulier.
- Marie Chupeau, agrégée et docteur en physique, enseignante expérimentée (préparation à l'agrégation et université), chef de projet R&D en Python
- Raphael-David Lasserri, docteur en physique, enseignant à l'université, chercheur en intelligence artificielle
- Julien Ripoché, docteur en physique, ancien enseignant à l'Université Paris-Saclay, machine learning engineer, spécialiste de la programmation orientée objet et du déploiement de solutions numériques en production.



SIRET : 84439862800025 -NAF : 7112B -N° TVA : FR59844398628 MAGIC LEMP SAS au capital social de 1000 €

Organisme de formation enregistré sous le numéro 11910881491
<https://magic-lemp.com>

Ce document est la propriété de Magic LEMP SAS, il ne peut être communiqué et/ou reproduit sans une autorisation préalable.