



INTELLIGENCE ARTIFICIELLE POUR LA SCIENCE

UNE FORMATION AU COEUR D'UN ÉCOSYSTÈME RICHE

Établissement composante de l'Université PSL, l'ESPCI Paris – PSL est à la fois une Grande École d'ingénieurs originale, un centre de recherche de renommée internationale et un puissant générateur d'innovation pour l'industrie. Elle recrute sur un concours sélectif une centaine d'étudiants pour les former à devenir les ingénieurs et les scientifiques de demain. Sa formation est résolument interdisciplinaire, au croisement de la Physique, de la Chimie et de la Biologie, et vise à donner à ses élèves-ingénieurs un socle unique et solide de connaissances et de compétences scientifiques dans une approche holistique et originale. L'ADN de la formation à l'école réside dans la mise en situation et les travaux pratiques en mode projet.

Dans le prolongement de cette mission, la formation continue de l'ESPCI Paris – PSL s'appuie sur une expertise scientifique et une approche interdisciplinaire, pratique et innovante. Elle propose aux professionnels des modules de perfectionnement et des parcours sur mesure, couvrant un large éventail de domaines en pleine évolution.

APPLICATION DE L'IA EN SCIENCE

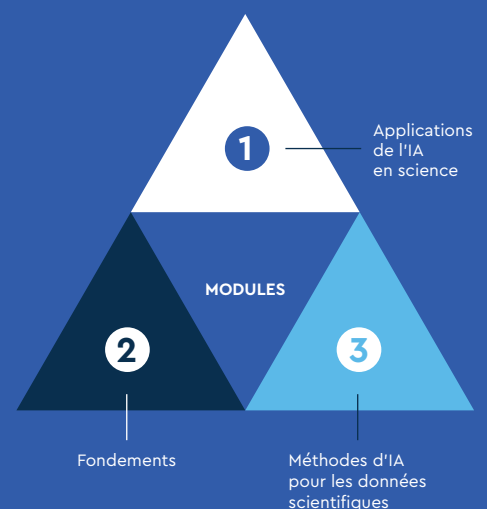
Cette formation présente les capacités et les limites pratiques de l'intelligence artificielle (IA) en recherche scientifique, avec un accent particulier sur la chimie, les matériaux et la biochimie. Les participants explorent les outils d'IA existants, acquièrent une compréhension conceptuelle de leurs principes sous-jacents et apprennent à appliquer des méthodes d'IA à de véritables données expérimentales, y compris la manière de travailler avec des jeux de données limités et imparfaits, et d'évaluer les résultats avec honnêteté.

Les travaux pratiques s'appuient sur de véritables jeux de données expérimentales, et les méthodes abordées vont des modèles statistiques fondamentaux et de la régression linéaire jusqu'aux réseaux de neurones modernes, en insistant autant sur le moment où une méthode est pertinente que sur la manière de l'employer.

UNE FORMATION EN 3 MODULES

La formation est conçue en 3 modules qui alternent cours magistraux et travaux pratiques dans des salles équipées pour le travail informatique, progressant des fondements statistiques jusqu'à l'apprentissage automatique en pratique.

Un module dédié aux plans d'expérience permet en outre aux participants d'optimiser la planification de leurs expériences et ainsi l'obtention de données expérimentales adéquates.



APPLICATIONS DE L'IA EN SCIENCE

🕒 1 JOUR

Une journée de panorama sur ce que l'intelligence artificielle peut et ne peut pas apporter à la R&D. Aucune expérience préalable de l'IA ni de la programmation n'est requise. La journée alterne présentations et démonstrations commentées ; il n'est pas demandé aux participants d'écrire du code.

À l'issue de la journée, les participants disposeront d'une vision claire et sans survente des outils d'IA utilisés en recherche : ce qu'ils font réellement, là où ils apportent de la valeur, et comment distinguer une application prometteuse d'une impasse (coûteuse).

✎ DÉTAILS DES COURS

Matinée : le panorama

- Qu'est-ce que l'IA ? Démystification en langage simple : apprendre des régularités à partir de données passées.
- Un tour d'horizon par domaine, à travers des exemples concrets : analyse d'images (microscopie, imagerie cellulaire), modélisation et conception moléculaire, prédiction de la structure des protéines (de type AlphaFold), et prédiction de propriétés à partir de mesures.
- Ce que l'IA générative et les grands modèles de langage (LLM) changent et ne changent pas pour la R&D : usages réels, limites, « hallucinations » et confidentialité des données.

Après-midi : bien décider (atelier)

L'après-midi passe de ce que l'IA peut faire à la question de savoir si un projet donné vaut la peine d'être mené. D'abord un court temps d'enseignement, puis un exercice pratique en comité.

À L'ISSUE DE CE MODULE

Les participants développeront des compétences clés :

- Identifier les outils d'IA utilisés en recherche et comprendre leur fonctionnement.
- Appliquer ces outils, ou en commanditer l'usage, de manière efficace dans des contextes scientifiques concrets.
- Évaluer si un projet d'IA mérite d'être mené, au regard de ses données, de son chemin de validation et de sa valeur attendue.

FONDEMENTS

🕒 2 JOURS

La maîtrise des méthodes d'intelligence artificielle passe d'abord par une solide compréhension des statistiques. En effet, l'analyse des incertitudes liées à une série d'expériences repose sur des concepts bien antérieurs à l'IA. Une première journée est dédiée aux notions clés de statistiques et à leurs applications dans des cas concrets.

Les plans d'expérience sont un outil pour planifier, organiser ses expériences de manière optimale afin de déterminer quels paramètres ont une influence significative. Ils offrent un accès aux interactions entre paramètres, permettent la modélisation des réponses et l'optimisation des procédés. Cette approche, applicable à de nombreux domaines de la R&D, permet un gain significatif de temps et de coûts ainsi que de générer des données pour des algorithmes d'IA. Une deuxième journée est dédiée à cet outil.

✎ DÉTAILS DES COURS

- Notions clés de statistiques : les intervalles de confiance, le test du χ^2 , l'ACP (Analyse en Composantes Principales) et les méthodes de classification
- Applications concrètes de ces notions clés
- Plans de criblage
- Plans pour surface de réponse
- Significativité statistique des conclusions

À L'ISSUE DE CE MODULE

Les participants développeront des compétences clés :

- Rappel des principaux fondements en probabilité.
- Choisir le type de plan le plus adapté à leur problématique
- Définir les expériences constitutives de ce plan (matrice des expériences)
- Exploiter les résultats en termes de facteurs significatifs et de prédiction de l'optimum de fonctionnement.

MÉTHODES D'IA POUR LES DONNÉES SCIENTIFIQUES

🕒 3 JOURS

Ce module apprend à transformer des données expérimentales en décisions. En trois jours, les participants passent des modèles fondamentaux à une compréhension éclairée des méthodes modernes, en construisant et en évaluant honnêtement des modèles sur de données de laboratoire. L'accent est mis autant sur le choix de la bonne méthode et sur le moment où il ne faut pas en utiliser, que sur sa mise en œuvre : le modèle le plus sophistiqué est rarement le bon. La dernière journée prend la forme d'une journée projet (hackathon) sur un jeu de données réel.

✎ DÉTAILS DES COURS

- Prédire une propriété : régression linéaire, arbres de décision, forêts aléatoires et boosting.
- Faire confiance à ses résultats : validation croisée, fuite de données, métriques (RMSE, MAE, R^2).
- Données réelles et limitées : prétraitement, valeurs manquantes et aberrantes, grandes dimensions.
- Interprétabilité et incertitude : importance des variables, domaine d'applicabilité.
- Choisir la prochaine expérience : apprentissage actif, optimisation multi-objectifs.
- Sensibilisation à l'apprentissage profond et à l'IA générative : quand cela en vaut la peine, et leurs limites.
- Journée projet (hackathon) sur un jeu de données réel.

À L'ISSUE DE CE MODULE

Les participants développeront des compétences clés :

- **Construire un modèle prédictif à partir d'un tableau de mesures expérimentales.**
- Évaluer honnêtement un modèle et reconnaître des résultats trompeusement bons.
- **Travailler avec des jeux de données restreints, incomplets et bruités.**
- Utiliser un modèle pour choisir la prochaine expérience à mener, et reconnaître son domaine d'applicabilité.
- Situer l'apport et les limites de l'apprentissage profond et de l'IA générative en R&D.

ADMISSION

- **Le module 1** s'adresse aux techniciens expérimentés, ingénieurs et cadres, en leur offrant les bases nécessaires pour identifier et comprendre les outils utilisés en IA. Il requiert des connaissances de base en mathématiques de niveau L2.
- **Le module 2** s'adresse aux techniciens expérimentés, ingénieurs et cadres, en leur offrant les bases en concepts fondamentaux en probabilité et statistiques et des outils pour planifier leurs expériences. Il requiert des connaissances de base en mathématiques et statistiques de niveau L2.
- **Le module 3** s'adresse aux techniciens expérimentés et aux ingénieurs qui souhaitent acquérir les bases nécessaires pour déployer des outils d'IA dans leur propre travail. Il requiert des connaissances de base en mathématiques de niveau L2. Aucune expérience préalable de l'intelligence artificielle ou de la programmation n'est requise ; les séances pratiques sont structurées de manière que les participants appliquent les méthodes au moyen de notebooks jupyter guidés plutôt qu'en écrivant du code de zéro.



contact-fc@espci.fr

COÛT

Notre programme de formation continue est structuré en 3 modules, conçus pour offrir une montée en compétences progressive, approfondie et directement applicable dans votre environnement professionnel.

- Module 1 : 1 000 €
- Module 2 : 2 000 €
- Module 3 : 3 000 €

Chaque module peut être suivi indépendamment selon les besoins. Afin de compléter au mieux leur formation, les participants peuvent donc choisir de suivre l'ensemble des modules pour un total de 6 000 €, ou bien un ou plusieurs modules de leur choix (montant variable en fonction du nombre de modules).

A noter qu'une seule des deux journées du module 2 peut être suivie. Le tarif reste de 1 000 € la journée.



CALENDRIER

Une première session est organisée sur la période juin-juillet et une seconde sur la période novembre-décembre. Veuillez nous contacter pour connaître les dates exactes de chaque module.

MODULE 1	Applications de l'IA en science 1 jour Qu'est-ce que l'IA ? Démystification en langage simple. Un tour d'horizon par domaine, à travers des exemples concrets. IA générative et les grands modèles de langage (LLM) Atelier « Comité d'investissement IA »		
	Fondements 2 jours		
MODULE 2	JOUR 1 Statistiques et analyse descriptive Théorie : statistiques fondamentales, probabilités, tests d'hypothèses, ACP/UMAP. Pratique : application à des jeux de données en chimie.	JOUR 2 Plans d'expériences (Données limitées) Théorie : principes des plans d'expériences ; limites de l'IA avec des données limitées. Pratique : élaboration de plans d'expériences simples ; simulation de données limitées.	
	Méthodes d'IA pour les données scientifiques 3 jours		
MODULE 3	JOUR 1 Prédire une propriété et faire confiance à ses résultats Théorie : régression linéaire ; arbres de décision, forêts aléatoires et boosting ; validation, fuite de données et métriques (RMSE, MAE, R^2). Pratique : construction d'un modèle de référence, comparaison de modèles et détection d'une fuite de données.	JOUR 2 Données limitées et choix de la prochaine expérience Théorie : prétraitement, valeurs manquantes et aberrantes, grandes dimensions ; interprétabilité, incertitude et domaine d'applicabilité ; apprentissage actif. Sensibilisation à l'apprentissage profond et à l'IA générative (CNN, GNN, transformeurs, LLM) : quand cela en vaut la peine, et leurs limites. Pratique : nettoyage et modélisation d'un jeu de données réaliste ; boucle d'apprentissage actif.	JOUR 3 Journée projet (hackathon) Théorie : Application de bout en bout de l'ensemble des outils, en équipe, sur un jeu de données réel (apporté par les participants ou fourni). Pratique : exploration, modélisation, validation honnête et présentation des résultats.

ESPCI PARIS

 UNIVERSITÉ PARIS SCIENCES & LETTRES

 10 rue Vauquelin, 75005 Paris

 contact-fc@espci.fr

