



**LumenAi**  
Real Time Machine Learning

---

## Formation en Intelligence Artificielle (Pau, France)

### Qui sommes-nous ?

LumenAi est une entreprise qui développe des algorithmes de machine learning (M.L.) pour tous les secteurs, des start-ups aux grandes entreprises. L'idée de proposer nos formations est venue de la volonté de partager notre expérience afin de répondre aux demandes émergentes du marché concernant l'intelligence artificielle.

Nos formations alternent théorie et cas pratiques (réalisés et analysés sur des ensembles de données réelles). Nous proposons aussi des interventions de professionnels de l'industrie et de l'activité de service (défense, santé, numérique...) pour parler du M.L. dans leur secteur et aborder un problème à résoudre dans le cadre d'un data challenge.

### Prérequis

- Une bonne connaissance des mathématiques et/ou informatique (Bac +5).
- L'envie d'apprendre à utiliser les outils que nous enseignons.
- Des bases de machine learning sont utiles mais pas nécessaires.

### Objectifs pédagogiques

#### Partie Générale

- Se familiariser avec les concepts d'I.A. (Machine Learning).
- Identifier les leviers de l'IA, les technologies et compétences nécessaires à leur mise en œuvre.

## Partie Technique

- Savoir implémenter un algorithme d'intelligence artificielle sous forme de POC.
- Maîtriser les connaissances et les applications liées au/à,
  - L'apprentissage statistique et en temps réel.
  - Système de recommandation.
  - Traitement du langage naturel.
  - Deep learning (apprentissage profond).
  - Graph mining.

## Contenu de la formation

### Durée : 68 heures

#### Apprentissage statistique et en temps réel (4 heures)

Dans ce cours, nous introduisons deux paradigmes de l'apprentissage machine : Apprentissage statistique et apprentissage en ligne. L'apprentissage statistique est le cadre conventionnel du Machine learning, dans lequel un modèle ou un algorithme est affiné grâce à un ensemble statique de données d'observations. Cette technique de M.L. est basée sur des hypothèses standard iid et révèle certaines limites des applications actuelles. Nous introduisons donc un cadre plus réaliste appelé apprentissage en ligne, dans lequel chaque observation se fait de façon séquentielle. Nous présentons des algorithmes capables de s'adapter à chaque nouvelle donnée. Nous introduisons des algorithmes standard (ex : la moyenne pondérée exponentielle, les algorithmes de bandit...) et mettons en évidence les performances pratiques sur différents scénarios.

#### Programme :

- Apprentissage statistique : cas supervisés et non supervisés, surajustement, validation croisée et généralisation.
- Apprentissage en ligne : prédiction avec conseils d'experts, régression et clustering en ligne, UCB.
- Applications pour la prévision de séries chronologiques (pétrole et gaz), AB testing (Digital Marketing).

#### Systèmes de recommandation (8 heures)

Les systèmes de recommandation sont bien connus grâce au défi Netflix. Historiquement, le défi visait à explorer des approches permettant de proposer un contenu personnalisé précis en prédisant le classement des films par les utilisateurs. Toutes les informations nécessaires étaient basées sur les évaluations précédentes des utilisateurs de films. Les systèmes de recommandation sont également utilisés pour la recherche d'informations et la découverte de contenu : combinés à l'interrogation et à la navigation, ils permettent aux utilisateurs confrontés à une énorme quantité d'informations de naviguer à travers ces informations de manière efficace et satisfaisante. Enfin, la notion de décision en temps réel combinée à la volonté de proposer un contenu précis mais diversifié a contribué à faire évoluer les approches traditionnelles du système de recommandation telles que le filtrage collaboratif de la machine de factorisation vers les modèles bandits. Cette évolution historique du système de recommandation vers des modèles bandits sera au cœur de ce cours. Nous nous concentrerons sur l'explication des principales approches statistiques de manière à la fois théorique et algorithmique. Les machines de factorisation et les modèles de bandits seront expliqués et testés sur des data set réel.

### Programme :

- Machines de factorisation : théorie et algorithmes.
- Modèles de bandits pour la prédiction en ligne : théorie et algorithmes.
- Applications pour les recommandations sur les supermarchés alimentaires / pour les modèles de bandits en temps réel, la saisie de données sur les médias sociaux.

## Traitement du langage naturel (24 heures)

Cette formation offre un panorama des méthodes traditionnelles, des modèles récents et des algorithmes d'analyse des données textuelles. Vous découvrirez ici les principaux enjeux et leviers du traitement du langage naturel (NLP), la vectorisation des mots ainsi que seq2seq. Vous serez capable de pratiquer ces concepts en utilisant le langage Python sur des travaux pratiques. Une session complète sera consacrée à la création d'un chatbot et prendra en compte tous les concepts appris pendant la formation NLP.

### Programme :

- Prétraitement NLP : tokenisation, lemmatisation, part of speech, stopwords.
- Vectorisation du texte.
- Similitudes, distances entre les textes.
- Regroupement et classification du texte, Latent Dirichlet Allocation.
- Modèles génératifs et réseaux de neurones.
- One-hot encoder, tf-idf.
- Utilisation de word embeddings (word2vec, doc2vec, fasttext, ...).
- Applications à la classification et au regroupement de textes.

## Deep learning (16 heures)

La formation fournira des éléments pour comprendre le deep learning et la façon dont il est mis en œuvre. Il y aura une partie pratique offrant la possibilité de manipuler ces nouveaux concepts, en résolvant des cas impliquant l'analyse de jeux de données réelles.

### Programme :

- ERM, généralisation et régularisation.
- Réduction des biais : des méthodes linéaires aux méthodes de kernel.
- Convolutional Neural Network pour les images : architecture et complexité.
- Généralisation et dimensionnalité réelle d'un réseau profond.
- Extension lexicale non supervisée.
- Apprentissage supervisé et « curse of dimensionality ».
- Réseaux de neurones pour la résolution de grands problèmes.
- Étape préliminaire : architecture supervisée.
- TensorFlow : un framework Python pour un apprentissage approfondi.
- Recurrent Neural Networks avec TensorFlow.

## Graph mining (16 heures)

Cette formation donne un aperçu des méthodes, modèles et algorithmes traditionnels et plus récents pour l'analyse de grands graphes. Vous y découvrirez les principaux enjeux et leviers du graph mining, les modèles de graphes aléatoires et les connaissances théoriques et algorithmiques de l'analyse des

graphes et de la détection communautaire. Vous aurez également l'occasion de pratiquer ces concepts au travers de travaux pratiques en utilisant le langage Python.

#### Programme :

- Modélisation mathématique et modèles graphes aléatoires.
- Théorie spectrale et algorithme de groupement spectral.
- Optimisation de la modularité.
- Génération de graphes aléatoires.
- Décomposition du spectre et détection communautaire.
- Lecture de graphes à grande échelle.
- Algorithme de Louvain.
- Applications à l'analyse de données hétérogènes et aux réseaux sociaux.

### Session d'initiation à l'IA. (8 heures)

Nous proposons aussi des sessions d'une journée ou +, où nous présentons les notions d'IA. et plus précisément de Machine Learning avec les différentes approches existantes (supervisée/non supervisée/renforcée), les technologies qu'elles incluent (deep learning, extraction de graphes, traitement du langage naturel, systèmes de recommandation...) et leurs applications (Maintenance prédictive, ciblage prédictif, détection d'anomalies, classification de documents...). Cette session a pour objectif, de familiariser les équipes avec ces nouvelles notions, identifier des projets auxquels ils pourraient intégrer de l'IA et enfin les technologies et compétences nécessaires à leurs implémentations. Les 5 technologies au-dessus sont évoqués par la suite lors de nos formations techniques (développeurs, data scientists).

### Organisation de la formation

#### Moyens pédagogiques et techniques

- Accueil des participants dans une salle dédiée à la formation.
- Documents supports de formation projetés.
- Exposés théoriques.
- Etude de cas concrets.
- Mise à disposition en ligne de documents supports à la suite de la formation.

#### Dispositif de suivi de l'exécution de l'évaluation des résultats de la formation

- Questions orales ou écrites (QCM).
- Mises en situation.
- Formulaire d'évaluation de la formation.

#### Personne à contacter

Paul Marion | Business Dev. | 06 35 14 69 08 | [pmarion@lumenai.fr](mailto:pmarion@lumenai.fr) | [www.lumenai.fr](http://www.lumenai.fr)