

## LLM Basics

**Durée: 2 Jours**    **Réf de cours: GK840035**    **Méthodes d'apprentissage: Classe à distance**

### Résumé:

Ce cours offre une introduction complète aux modèles de langage de grande taille (LLM), en se concentrant sur ce qu'ils sont, comment les construire en utilisant PyTorch, et comment les utiliser pour l'inférence dans les tâches de traitement du langage. Les participants apprendront l'histoire des LLM, comment ils s'intègrent dans le paysage plus large de l'IA/IA générative, les modèles de langage basés sur les réseaux neuronaux, et comment utiliser les RNN, LSTM et transformers pour les tâches de traitement du langage.  
Mise à jour : 9/12/2024

### Public visé:

Data Scientists, Data Engineers intéressés par l'utilisation des LLM pour la prédiction et le fine-tuning. Les développeurs de logiciels souhaitant acquérir une expérience pratique de base avec les frameworks NLP et les LLM.

### Objectifs pédagogiques:

- En travaillant dans un environnement d'apprentissage pratique et guidés par un expert, les étudiants apprendront les bases des modèles de langage de grande taille (LLM) et comment les utiliser pour faire des prédictions afin de créer des applications alimentées par l'IA.
- Comprendre les bases du traitement du langage naturel (NLP)
- Mettre en œuvre des techniques de prétraitement et de tokenisation du texte en utilisant NLTK
- Expliquer les embeddings de mots et l'évolution des modèles de langage
- Utiliser les RNN et LSTM pour gérer les données séquentielles
- Décrire ce que sont les transformers et utiliser des modèles clés comme BERT et GPT
- Comprendre les risques et les limitations des LLM
- Utiliser des modèles pré-entraînés de Hugging Face pour mettre en œuvre des tâches de NLP
- Comprendre les bases des systèmes de génération augmentée par la récupération (RAG)

### Pré-requis:

Pour assister à cette formation, il est recommandé pour les candidats de:

- Maîtriser la programmation Python
  - Maîtriser l'analyse de données à l'aide de Pandas
- Afin de vous assurer que vous possédez toutes les connaissances requises pour participer à la formation, notre équipe commerciale vous proposera un QCM

## Contenu:

### Introduction au NLP

- Qu'est-ce que le NLP ?
- Bases du NLP : Prétraitement et tokenisation du texte
- Bases du NLP : Embeddings de mots
- Introduction aux bibliothèques NLP traditionnelles
- Brève histoire de la modélisation du langage
- Introduction à PyTorch et HuggingFace pour le prétraitement du texte
- Réseaux neuronaux et données textuelles
- Construction de modèles de langage utilisant des RNN et des LSTM

### Transformers et LLMs

- Introduction aux transformers
- Utilisation des transformers de Hugging Face pour la prédiction
- LLMs et IA générative
- Options actuelles de LLM
- Fine-tuning de GPT
- Alignement des LLMs avec les valeurs humaines
- Systèmes de génération augmentée par la récupération (RAG)

## Méthodes pédagogiques :

Les participants réalisent un test d'évaluation des connaissances en amont et en aval de la formation pour valider les connaissances acquises pendant la formation.

Un support de cours électronique sera remis aux participants.

Suivi de cette formation à distance depuis un site client équipé. Il suffit d'avoir une bonne connexion internet, un casque avec micro et d'être dans un endroit au calme pour en profiter pleinement

Une fiche explicative est adressée en amont aux participants pour leur permettre de vérifier leur installation technique et de se familiariser avec la solution technologique utilisée.

L'accès à l'environnement d'apprentissage, ainsi qu'aux feuilles d'émargement et d'évaluation est assuré.

En savoir plus : <https://www.globalknowledge.com/fr-fr/solutions/methodes-d'apprentissage/classe-a-distance>

## Autres moyens pédagogiques et de suivi:

- Compétence du formateur : Les experts qui animent la formation sont des spécialistes des matières abordées et ont au minimum cinq ans d'expérience d'animation. Nos équipes ont validé à la fois leurs connaissances techniques (certifications le cas échéant) ainsi que leur compétence pédagogique.
- Suivi d'exécution : Une feuille d'émargement par demi-journée de présence est signée par tous les participants et le formateur.
- En fin de formation, le participant est invité à s'auto-évaluer sur l'atteinte des objectifs énoncés, et à répondre à un questionnaire de satisfaction qui sera ensuite étudié par nos équipes pédagogiques en vue de maintenir et d'améliorer la qualité de nos prestations.

Délais d'inscription :

- Vous pouvez vous inscrire sur l'une de nos sessions planifiées en inter-entreprises jusqu'à 5 jours ouvrés avant le début de la formation sous réserve de disponibilité de places et de labs le cas échéant.
- Votre place sera confirmée à la réception d'un devis ou "booking form" signé. Vous recevrez ensuite la convocation et les modalités d'accès en présentiel ou distanciel.
- Attention, si cette formation est éligible au Compte Personnel de Formation, vous devrez respecter un délai minimum et non négociable fixé à 11 jours ouvrés avant le début de la session pour vous inscrire via [moncompteformation.gouv.fr](https://moncompteformation.gouv.fr).

Accueil des bénéficiaires :

- En cas de handicap : plus d'info sur [globalknowledge.fr/handicap](https://globalknowledge.fr/handicap)
- Le Règlement intérieur est disponible sur [globalknowledge.fr/reglement](https://globalknowledge.fr/reglement)