

Best

## IA Générative - Les modèles de langages massifs (LLMs)

Développer de nouvelles formes de contenus grâce aux LLM

 4,5/5 (63 avis)

 Présentiel ou en classe à distance



2 jours (14 h)

Prix inter : 2.050,00 € HT

Forfait intra : 6.090,00 € HT

Réf.: BI108

L'avènement des modèles de langages massifs, Large Language Models ou LLMs, tels ChatGPT, constitue une étape clé pour l'IA. Nous sommes au stade où des machines désormais peuvent appréhender une quantité extrême de contenu et d'en développer des formes nouvelles de services aujourd'hui efficents. Pour les entreprises et les organisations il s'agit de se positionner face à ces évolutions majeures en terme de création de valeur. Il devient dès lors nécessaire pour elles de : comprendre les mécanismes sous-jacents à ces modèles, savoir choisir le bon LLM, connaître les écosystèmes incontournables du domaine, de comprendre la personnalisation dans l'usage via le Prompt Engineering, de mieux cerner les capacités d'enrichissement de ces modèles (via des sources internes ou des outils), d'aborder sereinement les processus de réglage fin sur des cas d'usage précis et enfin d'identifier les problématiques de mise en œuvre de tels modèles. À travers cette formation, les participants pourront apprécier, sur la base d'explications et d'ateliers sur des cas concrets, chacune des composantes significatives de ces modèles disruptifs.

### A qui s'adresse cette formation ?



#### Pour qui

- Chefs de projets
- Développeurs
- Data scientists



#### Prérequis

- Une connaissance de base des principes de Machine Learning et de Deep Learning
- La maîtrise d'une langue de script type Python est recommandée
- **Disposez-vous des connaissances nécessaires pour suivre cette formation ? Testez-vous !**

### Programme

#### 1 - Introduction

- Un changement de paradigme
- Qu'est ce qui change ?
- Une publication fondatrice
- Une loi d'échelle pour les modèles de langage
- Évolution temporelle des LLMs
- De nouveaux écosystèmes
- L'ère du Post Deep Learning
- Personnalisation par Prompts
- Personnalisation par enrichissement
- Personnalisation par réglage fin

## 2 - Cas d'usage

- Agents conversationnels et assistants virtuels
- Génération de code et debuggage
- Analyse de sentiments / opinions
- Classification de texte et clusterisation
- Synthèse de texte ou corpus
- Traduction
- Génération de contenu
- Autres cas d'usages significatifs
- LAB : Proof of concept sur cas concrets

## 3 - Fondations

- Le traitement du langage naturel (TAL)
- L'architecture disruptive des Transformers
- La tokenisation des textes
- L'encodeur d'un Transformer
- La couche d'embedding
- L'encodage de positionnement
- Vecteur de positionnement
- Le mécanisme d'attention multi-têtes
- Points essentiels du mécanisme d'attention
- La "spécialisation" des têtes d'attention
- Calcul des scores d'attention
- Addition et Normalisation des sorties
- Le Décodeur d'un Tranformer
- L'entraînement d'un Transformer
- La couche d'auto-attention masquée
- La couche d'attention du décodeur
- Les couches supérieures du décodeur

## 4 - En pratique

- Choisir un LLM
- Critères de choix
- Trois classes de modèles
- Modèles à encodeur simple
- Focus modèles BERTs
- Modèles à décodeur simple
- Focus modèles GPTs
- Un foisonnement de modèles dérivés
- La bataille des LLMs
- La course vers des LLMs légers
- L'exemple de LLaMa
- Trois approches de réduction
- Écosystèmes clés
- APIs de modèles Fermés
- HuggingFace et les modèles ouverts
- Écosystèmes applicatifs type LangChain
- LL Mops et MLFlow
- Atelier

Prise en main des écosystèmes LLMs clés

## 5 - Mise en oeuvre

- Choix service / in house / hybrid
- In house workflow
- Service workflow
- Écosystèmes d'entraînement et d'inférence
- L'entraînement d'un modèle massif
- L'étape d'évaluation des modèles
- Le réglage des hyperparamètres
- Déploiement d'un modèle
- Model fine-tuning
- Prompt engineering
- MLOps d'un LLMs
- LAB : Environnement de déploiement d'un LLM

## 6 - Le Prompt Engineering

- Configuration des paramètres des LLMs
  - Qu'est ce qu'un token ?
  - Notion de distribution des probabilités des LLMs
  - Les échantillonnages Top-K et top-p
  - La température du modèle
  - Le réglage des paramètres en pratique
  - Les composantes d'un prompt
  - Règles générales
  - L'approche Few-Shot Learning
  - Zero, one to Few-shot learning
  - L'approche Chain of thoughts
  - L'incitation par chaînes de pensées
  - Des approches plus avancées
  - ReAct Prompting
  - Méthode ReAct
- Atelier

Prompt Engineering sur cas concrets

## 7 - LLMs augmentés

- Au delà du prompt, l'enrichissement des LLMs
  - Ajout de capacité mémorielle
  - Mémoire tampon (Buffer Memory)
  - Plusieurs mécanismes de mémorisation
  - Les mémoires de l'écosystème LangChain
  - Élargissement des connaissances
  - Retrieval Augmented Generation (RAG)
  - Le partitionnement des textes externes
  - Projection sémantique des documents (Embeddings)
  - Les bases de données vectorielles
  - Les algorithmes du search dans les bases vectorielles
  - Une galaxie d'outils possibles !
- Atelier

Mise en oeuvre d'un agent conversationnel

## 8 - Déploiement de LLMs

- Quand le prompt engineering ne suffit plus
  - Qu'est ce que le réglage fin
  - Trois techniques classiques
  - Reinforcement learning by Human feedback (RLHF)
  - Détails d'un réglage fin Supervisé
  - Les trois options pour l'ajustement des paramètres
  - Les approches PEFT (Parameter Efficient Tuning)
  - La méthode LoRA (Low Rank Adaptation)
  - Une variante efficiente : QLoRA
  - Qu'est ce que la mise en service d'un LLM
  - Journaliser le modèle dans le registre des modèles
  - Création d'un point de terminaison vers le modèle
  - Interroger le point de terminaison
- Atelier

Réglage fin et mise en service d'un LLM



## Les objectifs de la formation

- Comprendre l'utilité des LLM
- Maîtriser l'usage des LLM et de l'IA générative
- Développer de nouveaux projets avec l'aide des LLM



## Evaluation

- Pendant la formation, le formateur évalue la progression pédagogique des participants via des QCM, des mises en situation et des travaux pratiques. Les participants passent un test de positionnement avant et après la formation pour valider leurs compétences acquises.



## Les points forts de la formation

- Une pédagogie basée sur l'alternance de phases théoriques, d'ateliers de mise en pratique, de retours d'expériences et de séances d'échanges.
- Le partage de bonnes pratiques de la part de consultants experts en Intelligence Artificielle.
- 82% des participants à cette formation se sont déclarés satisfaits ou très satisfaits au cours des 12 derniers mois.



## Dates et villes 2026 - Référence BI108



Dernières places disponibles



Session garantie

### A distance

du 15 janv. au 16 janv.

du 5 févr. au 6 févr.

du 12 mars au 13 mars ☈

du 9 avr. au 10 avr.

du 12 mai au 13 mai

du 11 juin au 12 juin

du 23 juil. au 24 juil. ☈

du 10 sept. au 11 sept.

du 12 oct. au 13 oct.

du 5 nov. au 6 nov. ☈

du 10 déc. au 11 déc.

### Marseille

du 15 janv. au 16 janv.

du 12 mars au 13 mars

du 12 mai au 13 mai

du 23 juil. au 24 juil.

du 12 oct. au 13 oct.

du 10 déc. au 11 déc.

### Toulouse

du 15 janv. au 16 janv.

du 12 mai au 13 mai

du 23 juil. au 24 juil.

du 12 oct. au 13 oct.

du 10 déc. au 11 déc.

### Lille

du 15 janv. au 16 janv.

du 12 mars au 13 mars

du 12 mai au 13 mai

du 23 juil. au 24 juil.

du 12 oct. au 13 oct.

du 10 déc. au 11 déc.

### Strasbourg

du 15 janv. au 16 janv.

du 12 mars au 13 mars

du 11 juin au 12 juin

du 10 sept. au 11 sept.

du 5 nov. au 6 nov.

## **Sophia Antipolis**

du 15 janv. au 16 janv.  
du 12 mars au 13 mars

du 11 juin au 12 juin  
du 10 sept. au 11 sept.

du 5 nov. au 6 nov.

## **Rouen**

du 15 janv. au 16 janv.  
du 12 mars au 13 mars

du 11 juin au 12 juin  
du 10 sept. au 11 sept.

du 5 nov. au 6 nov.

## **Aix-en-Provence**

du 15 janv. au 16 janv.  
du 12 mars au 13 mars

du 12 mai au 13 mai  
du 23 juil. au 24 juil.

du 12 oct. au 13 oct.  
du 10 déc. au 11 déc.

## **Nantes**

du 5 févr. au 6 févr.  
du 12 mai au 13 mai

du 23 juil. au 24 juil.  
du 12 oct. au 13 oct.

du 10 déc. au 11 déc.

## **Paris**

du 5 févr. au 6 févr.  
du 12 mars au 13 mars ☺  
du 9 avr. au 10 avr.  
du 12 mai au 13 mai

du 11 juin au 12 juin  
du 23 juil. au 24 juil. ☺  
du 10 sept. au 11 sept.  
du 12 oct. au 13 oct.

du 5 nov. au 6 nov. ☺  
du 10 déc. au 11 déc.

## **Bordeaux**

du 5 févr. au 6 févr.  
du 9 avr. au 10 avr.

du 11 juin au 12 juin  
du 10 sept. au 11 sept.

du 5 nov. au 6 nov.

## Rennes

du 5 févr. au 6 févr.  
du 12 mai au 13 mai

du 23 juil. au 24 juil.  
du 12 oct. au 13 oct.

du 10 déc. au 11 déc.

## Lyon

du 5 févr. au 6 févr.  
du 9 avr. au 10 avr.

du 11 juin au 12 juin  
du 10 sept. au 11 sept.

du 5 nov. au 6 nov.

## Montpellier

du 12 mai au 13 mai

du 10 déc. au 11 déc.

## Metz

du 12 mai au 13 mai

du 10 déc. au 11 déc.

## Nancy

du 12 mai au 13 mai

du 10 déc. au 11 déc.

## Angers

du 12 mai au 13 mai

du 10 déc. au 11 déc.

## Grenoble

du 11 juin au 12 juin

du 5 nov. au 6 nov.

## Toulon

du 11 juin au 12 juin

du 5 nov. au 6 nov.

## Saint-Etienne

du 11 juin au 12 juin

du 5 nov. au 6 nov.

## Tours

du 11 juin au 12 juin

du 5 nov. au 6 nov.