

Formation LLM et RAG

3 jours (21 heures)



Présentation

En participant à cette formation, vous vous engagerez dans une exploration approfondie de la technologie des LLM (Large Language Models). Les LLM concentrent en un unique modèle plusieurs capacités habituellement séparées dans des modèles de traitement automatique du langage naturel (TAL/NLP).

Nous rentrerons dans le détail de la théorie derrière les LLM et l'intelligence artificielle. Cette formation vous permettra d'appréhender des articles récents dans le domaine, de remettre en question des choix techniques et de concevoir vos propres modèles. Nous reviendrons ainsi sur les problématiques de l'apprentissage automatique en général, puis sur les premiers modèles développés pour du TAL et les limites qui ont conduit à l'établissement des transformers - actuellement à l'état de l'art.

À travers une approche interactive et pratique, vous apprendrez à concevoir, à entraîner et à déployer des modèles de LLM pour résoudre une gamme étendue de tâches, en utilisant des techniques modernes et des outils de pointe. Nous verrons également le modèle RAG (Retrieval Augmented Generation) et son intégration dans des scénarios réels.

Cette formation vous fournira les outils et les connaissances nécessaires pour exploiter pleinement le potentiel des LLM et du RAG dans votre organisation, en ouvrant la voie à de nouvelles perspectives et opportunités dans le domaine du traitement automatique du langage naturel.

Objectifs

- Comprendre l'importance des LLM dans le traitement du langage naturel
- Maîtriser les architectures et les principes théoriques des LLM
- Constituer un réseau de neurones pour de la régression, de la classification et du traitement du langage naturel
- Concevoir et fine-tuner un modèle à l'état de l'art
- Explorer les cas d'utilisation et les avantages dans divers domaines linguistiques
- Etre en mesure de suivre l'actualité liée aux LLM et aux modèles génératifs
- Apprendre les techniques avancées telles que l'adaptation sur des données spécifiques et la gestion de la mémoire
- Créer des chatbots personnalisés avec différentes technos LLM
- Découvrir le modèle RAG et son intégration dans des scénarios réels



Public visé

- Profils non techniques : Chefs de projet, Managers
- Profils techniques : Data Analyst, Data Scientist, BI Engineer, MLOps, DevOps, Data Engineer, Software Engineer, etc.

Pré-requis

- Familiarité avec les concepts de machine learning
- Expérience de programmation en Python
- Une compréhension basique des réseaux de neurones est un plus

Programme détaillé

Jour 1

L'apprentissage automatique

- Les réseaux de neurones
- Les fondements théoriques des réseaux de neurones
- L'expressivité d'un réseau de neurones
- Le processus d'apprentissage : la 'backpropagation'
- Le dilemme biais-variance, la régularisation
- La méthodologie scientifique d'évaluation d'un modèle

Mise en application de l'apprentissage automatique

TP 1 : Apprentissage supervisé, les tâches de régression et de classification

- Illustration des problèmes régulièrement rencontrés lors de l'apprentissage
- Interpolation vs extrapolation, exemples adversariaux sur LLM
- Comment traiter le langage naturel ?

Traiter les données séquentielles

- Les modèles élémentaires : RNN, LSTM, GRU
- Apprentissage auto-supervisé
- Le mécanisme d'attention : 'apprendre à analyser'
- Les transformers, architecture empirique à l'état de l'art pour les LLM
- Répondre à une tâche particulière : le fine-tuning
- Aux origines de ChatGPT / GPT-4

Jour 2

Mise en application du traitement automatique du langage

- Fine-tuner un modèle
- Les limites des transformers

TP : Apprentissage auto-supervisé, création d'un transformers



Les LLM en pratique

- Cas d'utilisation et avantages dans divers domaines linguistiques
- Impact des LLM sur la société et l'industrie
- Exemples d'applications réussies de LLM
- Introduction des technologies pour utiliser les LLM

Techniques avancées

- Adaptation sur des données/tâches spécifiques (few-shot learning, zero-shot learning, fine-tuning)
- Gestion de la mémoire et du contexte
- Réalisation de tâches plus complexes
- Construire son chatbot avec LangChain, LangFlow, H2O-chatgpt, VLLM, chainlit, etc.

TP : Création d'un chatbot personnalisé avec LangFlow et ChatGPT

Jour 3

Modèle RAG (Retrieval-Augmented Generation)

- Compréhension approfondie du modèle RAG
- Intégration de la recherche dans le processus génératif
- Applications pratiques et avantages

TP : Utilisation du modèle RAG dans des scénarios réels

Les bases de données vectorielles

- Introduction aux bases de données vectorielles
- Modélisation des données vectorielles
- Fonctionnalités et opérations (recherche sémantique, algorithmes ANN, etc.)
- Bases vectorielles VS graphes de connaissances VS contexte des LLM

TP : Utilisation d'une base vectorielle dans pour le RAG

Déploiement et intégration des LLM

- Options et outils de déploiement des LLM
- Intégration dans des applications et des services
- Gestion des modèles déployés et suivi des performances

Perspectives et conclusion

- Tendances et avancées dans le domaine des LLM : gestion du contexte, modèle multimodal, quantization, etc.
- Les MCMC, les modèles de diffusion (type Stable Diffusion)
- Analyse d'un article sur une alternative éventuelle au réseau de neurones : Les Kolmogorov-Arnold networks (KAN)
- Analyse d'un article sur les modèles de Google Deepmind : FunSearch, AlphaGeometry
- Réponses aux questions et ressources supplémentaires

Organisation

- La formation alterne les apports théoriques du formateur soutenus par des exemples et des séances de réflexions et de cas pratiques.
- Support de la formation distribué au format numérique à tous les participants.



Prix : 1590€ HT / participant

Tarif dégressif appliqué à partir de 5 participants, dans la limite de 25 participants maximum par session de formation.

Contact : +33 (0)7 45 41 66 05 / contact@ai-training.fr

