

[Formation] Traitement automatique du langage naturel : NLP et LLMs

OBJECTIFS

- Présenter les outils de traitement de langue, qu'ils soient basés sur des méthodes statistiques ou sur de méthodes symboliques
- Décrire le fonctionnement et identifier les atouts et les faiblesses des grands modèles de langage LLM comme GPT-4
- Évaluer les techniques et les adapter à chaque type de problème
- Comparer et combiner les approches : exploration d'outils statistiques (approches fréquentistes, similarité sémantique, plongements) et formels (langages formels, logiques de premier ordre et de description, lambda-calcul, ontologies)?

PROGRAMME

Introduction à la linguistique

Approche neuronales

Approches statistiques

- Désambiguïsation de mot
- Classification supervisée de textes
- Similarité et parenté sémantiques
- Pré-traitement du texte
- Modèles fréquentistes : Représentation Bag-of-words, modèles de langue n-gram, et dérivés.
- Deep learning et modèles de langue neuronaux
- Plongements et applications
- Modèles séquentiels et mécanisme d'attention
- Transformers
- Représentations contextuelles
- Apprentissage par transfert et Large Language Models
- Utilisation de SentiWordNet pour la classification des critiques
- Utilisation de réseaux de neurones sur le même corpus de textes, comparaison des résultats ; possibilité d'approche hybride (plongement d'arbres



DATES ET LIEUX

Du 11/05/2026 au 13/05/2026 à Paris

Du 02/11/2026 au 04/11/2026 à Paris

PUBLIC / PREREQUIS

Ingénieurs, chefs de projets, data scientists devant traiter, générer ou intégrer des fonctionnalités avec des données textuelles et du langage naturel.

Des connaissances en langage Python sont nécessaires afin de tirer pleinement profit de cette formation.

COORDINATEURS

Fabian SUCHANEK

Enseignant-chercheur à Télécom Paris. Il a fait ses recherches à l'Institut Max Planck en Allemagne, chez Microsoft Research Cambridge/UK, chez Microsoft Research Silicon Valley/USA, et à l'INRIA Saclay. Il est l'auteur principal de YAGO, une des plus grandes bases de connaissances publiques dans le monde.

Matthieu LABEAU

Enseignant-chercheur à Télécom Paris. Son activité de recherche en traitement automatique du langage, concerne principalement l'apprentissage de

syntaxiques)

Travaux pratiques

Approches symboliques

- Langages formels Graphes conceptuels, ontologies, bases de connaissances
- Extraction d'informations
- Désambiguïsation
- Détection d'entités

Travaux pratiques

Synthèse et conclusion

représentations et la modélisation du langage.

MODALITES PEDAGOGIQUES

La formation comprend des travaux pratiques qui permettent d'appliquer les notions théoriques abordées.

Appelez le 01 75 31 95 90
International : +33 (0)1 75 31 95 90

contact.exed@telecom-paris.fr / executive-education.telecom-paris.fr