

Machine Learning, Transformer et NLP

ECTS : 2

Volume horaire : 18

Description du contenu de l'enseignement :

1. Introduction au Traitement Automatique du Langage Naturel (NLP)

- Définition et enjeux du NLP
- Applications du NLP dans l'industrie et la recherche
- Évolution des approches NLP : des modèles classiques aux Transformers

2. Modélisation des Textes

- Représentation des textes : Bag of Words, TF-IDF
- Limites des représentations classiques
- Introduction aux Word Embeddings

3. Pré-traitement du Texte

- Tokenization : règles et techniques
- Utilisation des expressions régulières (REGEX) pour le nettoyage des textes
- Stemming et lemmatisation : différences et usages
- Suppression des stopwords et normalisation des textes

4. Techniques de Word Embeddings

- Introduction aux embeddings contextuels et statiques
- Présentation des modèles GloVe et Word2Vec
- Comparaison et utilisation des embeddings dans le NLP

5. Sentiment Analysis

- Utilisation des lexiques de sentiments
- Visualisation des sentiments avec des graphiques
- Approches non supervisées pour l'analyse des sentiments
- Classification Naïve Bayes appliquée à l'analyse de sentiments

6. Introduction aux Transformers

- Architecture des Transformers : concepts clés
- Attention et auto-attention : mécanismes et visualisation
- Applications générales des Transformers (traduction, résumé, génération de texte)
- Présentation de l'environnement Hugging Face

7. Utilisation des Transformers

- Panorama des modèles de Transformers : BERT, GPT, T5, etc.
- Visualisation des mécanismes d'attention
- Entraînement et fine-tuning d'un Transformer
- Pipelines NLP avec Hugging Face : classification, résumé, traduction

8. Application aux Résumés de Texte

- Types de résumé automatique : extractif vs. abstractive
- Comparaison des modèles de résumé
- Métriques d'évaluation : ROUGE, BLEU, METEOR
- Implémentation pratique d'un modèle de résumé avec Transformers

9. Projet Final et Évaluation

- Mise en œuvre d'un projet appliqué utilisant les Transformers
- Présentation des résultats et discussion
- Évaluation finale du cours

Evaluation

- Examen en salle informatique (100%)

Compétence à acquérir :

Comprendre des NLPs et les Transformers

Mode de contrôle des connaissances :

Examen en salle informatique 100%

Document susceptible de mise à jour - 01/04/2026

Université Paris Dauphine - PSL - Place du Maréchal de Lattre de Tassigny - 75775 PARIS Cedex 16