

# **Traitement automatique du langage par deep learning from scratch.**

## **Durée**

4 jours.

## **Participants**

Développeur voulant produire des fonctionnalités modernes de traitement automatique du langage naturel.

## **Prérequis**

Notions de Python, produit scalaire, matrices, moyenne, variance, dérivation

## **Description**

Formation spécialisée dans le traitement automatique du langage naturel.

## **Objectifs pédagogiques**

Acquérir les fondamentaux de l'état de l'art en traitement automatique du langage naturel.

## **Travaux pratiques**

Multiples cas d'utilisation avec Keras.

## **Programme**

### **Données, Intelligence Artificielle et Machine Learning**

- Introduction à l'intelligence artificielle ;
- De l'importance des données et des prétraitements ;
- Les différents cadres : non-supervisé, supervisé, renforcement ;

### **Concepts fondamentaux d'un réseau de neurones**

- Réseau de neurones : formalisme, fonctions d'activations ;
- Apprentissage d'un réseau de neurones : fonctions de coût, SGD, Adam ;
- Initialisation et régularisation : orthogonalité à l'initialisation, régularisations L1/L2, politiques de batchs, dropout.

## **Traitement automatique du langage naturel**

- Représentation sac de mots, TFIDF ;
- Prétraitements usuels (étiquetage morphosyntaxique, lemmatisation, etc) ;
- Dictionnaire de mots et n-gram ;
- Représentation one-hot et word embeddings.
- Utilisation du texte dans le cadre de signaux hétérogènes.

## **Réseaux récurrents (RNNs)**

- Présentation des RNNs : principes fondamentaux et applications ;
- Variantes modernes : LSTMs, GRUs, réseaux bi-directionnels ;
- Éléments techniques de gestion des RNNs : padding, taille des séquences ;
- Mécanisme de décodage pour produire une séquence.

## **Modèles à attention**

- Attention simple intégrée à un RNN ;
- Modèles transformeurs ;
- Architectures encodeur-décodeur ou à base de masque ;
- Variantes modernes : Reformer, Performer, XLNet, etc.

## **Cas d'utilisation du traitement du langage moderne**

- Classification ;
- Question/réponse, sélection de texte pertinent et compréhension de langage ;
- Génération conditionnelle ;
- Transfert & multitâche (SuperGLUE) ;
- Agents conversationnels avec DialogFlow ;
- Détection de thématiques avec BigARTM.

## **Questions éthiques**

- Perpétuation des biais des données ;
- Représentation des sous-populations dans les données ;
- Intoxication des conversations publiques avec des données générées ;
- Utilisation des données utilisateurs (RGPD & CNIL).