

retengr

# NLP

---

Durée : 4 jours



# Méthode pédagogique

La formation se passe en mode présentiel ou distanciel, et se compose de 70% de travaux pratiques (Mises en situation, débats, exercices). Le formateur alterne entre théorie et pratique ; chaque participant dispose d'un ordinateur avec un accès direct à une Machine Virtuelle préparée par nos soins, sur lequel il réalisera les exercices proposés.

Une évaluation quotidienne de l'acquisition des connaissances de la veille est effectuée.

Une synthèse est proposée en fin de formation. Une évaluation à chaud sera réalisée par le stagiaire à la fin du cours.

Un support de cours (version électronique) sera remis à chaque participant comprenant les slides sur la théorie, les exercices.

## Présentation

Le traitement naturel du langage est un domaine à part entière de l'intelligence artificielle à l'intersection de l'informatique, des mathématiques, de la linguistique et des sciences cognitives. L'objectif est de construire des applications capables d'analyser, de modéliser, de comprendre et d'imiter le langage humain.

Les transformers introduits en 2017 et intégrant des capacités d'attention ont révolutionné ce domaine en offrant une nouvelle méthode pour la compréhension du langage naturel.

D'une part, nous pouvons désormais réutiliser de nombreux modèles pré-entraînés extrêmement performants et d'autre part, ils ont contribué à l'IA générative permettant de générer du contenu original à partir de données existantes.

Vous connaissez d'ores et déjà les principes fondamentaux de l'apprentissage automatique et vous avez déjà mis en oeuvre différentes architectures de réseaux de neurones, mais vous souhaitez en savoir plus sur les opportunités proposées par le Machine et le Deep Learning dans le



domaine du traitement naturel du langage, cette formation est faite pour vous.

Vous serez capables de pré-traiter des données textuelles et vous serez en mesure d'implémenter des modèles permettant de les valoriser en TensorFlow ou PyTorch : classification de texte, recherche d'informations, analyse de sentiments, extraction d'informations, traduction, résumé de texte, etc...

## Objectifs

- Mettre en place un prétraitement efficace d'un jeu de données textuelles
- Maîtriser les architectures de réseaux de neurones récurrents et des transformers
- Comprendre les bases théoriques et pratiques des LLM
- Réutiliser des modèles existants avec du transfer learning
- Comprendre le principe du RAG
- Mesurer la pertinence des modèles mis en œuvre
- Mettre en œuvre des cas concrets pour chaque architecture/solution

## Audience

Data Analysts, Data Scientists, Développeurs

## Le formateur

Le formateur est un expert du domaine qui intervient sur le sujet depuis plusieurs années en formation mais aussi en conseil.



Doté d'une grande qualité d'écoute, sa pédagogie et sa compétence technique vous permettront d'acquérir les compétences sur le sujet.

Il saura alterner entre théorie, pratique, et retours d'expérience.

# Pré-requis

Les stagiaires doivent avoir une expérience en Deep Learning ou avoir suivi la formation « Maitriser le Deep Learning »

Il est aussi nécessaire de maîtriser des concepts du Machine Learning et d'avoir mis en œuvre différentes architectures de réseaux de neurones : convolutionnels, récurrents, ...)

La connaissance du langage Python, des bibliothèques scientifiques (scikit-learn, pandas, numpy) et du framework Deep Learning TensorFlow ou PyTorch est nécessaire.

Afin de valider les compétences et les prérequis de chaque participant, en amont de la formation, le formateur organise un entretien téléphonique. Il confirme alors que le participant a le niveau nécessaire et que le contenu répond bien à ses attentes.

# Programme

## Introduction [3.5h]

- Les bases de la linguistique
- Introduction au Text Mining / Fouille de texte (Data Mining pour le texte)
- Extraction d'informations
- Recherche d'informations
- Catégorisation du texte
- Résumé du texte



## Text Normalisation / Normalisation du texte [3.5h]

- n-grams
- Tokenization
- Stop Word
- Stemming
- Part-Of-Speech (POS) tagging
- Lemmatization

## Vectorisation du texte [3.5h]

- Analyse de la fréquence des termes (Counter, TF-IDF, Word vectors)
- Bag of word
- Word Embedding : Word2vec, GloVe, FastText, etc.
- Création d'une pipeline de preprocessing

## Les réseaux de neurones récurrents [3.5h]

- Principes fondamentaux des RNN
- LSTM et GRU
- Approches Encodeur-Decodeur
- Utilisation de connexions résiduelles (skip connections)

## L'architecture transformers [3.5h]

- Seq2Seq
- Attention is All You Need
- Architecture des transformers
- Mécanisme d'attention
- Tokens & Embeddings

## Les LLM [3.5h]

- Les familles de modèles : encoder, decoder & encoder-decoder
- Panorama de quelques modèles
- Méthodes d'apprentissage et d'optimisation
- Closed model APIs vs OpenModel weights vs Fully open model



- Coût et empreinte écologique des LLM
- Des LLM vers les SLM

## HuggingFace & Keras Hub [3.5h]

- HuggingFace et ses pipelines
- Keras Hub et Kaggle

## RAG – Retrieval-Augmented Generation [3.5h]

- Fondements de Retrieval-Augmented Generation
- Les bases vectorielles
- Premiers pas avec LangChain
- Du RAG avancé vers L'Agentic AI

# Modalités et délais d'accès à la formation

Les inscriptions sont possibles jusqu'à 48 heures ouvrées avant le début de la formation, en interentreprises, dans la limite des places disponibles.

Pour les formations organisées en intra entreprise, la liste des participants peut être modifiée jusqu'à 24h ouvrées avant le début de la formation.

## Accessibilité

RETENGR facilite l'accessibilité de ses formations.

Cette formation est accessible aux personnes en situation de handicap.

Si vous avez un besoin d'accès spécifique, contactez Céline BOURREIL (celine.bourreil@retengr.com) qui étudiera avec Handifiel's (notre référent handicap) votre demande et vous proposera les meilleures solutions

The background is a vibrant, abstract composition. It features large, overlapping organic shapes in shades of purple, yellow, and red. A prominent yellow shape on the left transitions into a red shape on the right. In the upper right, there is a pattern of small yellow dots arranged in a grid-like structure. At the bottom, there are more organic shapes in light pink, teal, and purple. A dashed blue line with a circular loop is visible in the lower-left quadrant.

**Vous allez nous adorer si  
comme nous vous pensez que...**

# Une formation doit être au service de la performance du collaborateur et de l'entreprise

Ceci nécessite une quête constante d'excellence de la part de l'organisme formateur avec une adaptation systématique aux enjeux de l'entreprise, la mise à jour régulière des supports de cours et une veille technologique indispensables pour toujours être à la pointe du domaine.



# L'expertise technique est aussi importante que les qualités pédagogiques



Nos formateurs sont tous des experts de leur domaine. Mais qu'ont-ils de plus que les autres ? Nous les sélectionnons en plus pour leurs qualités de pédagogue et leurs méthodes d'enseignements. Nous plaçons les qualités pédagogiques au même niveau que l'expertise afin que nos stagiaires tirent le meilleur de leurs formations.



re'engr

## L'excellence naît de l'excellence

Beaucoup de nos clients se classent parmi les leaders de leurs industries respectives ou parmi les start-ups les plus prometteuses. Nous savons que former les collaborateurs de telles entreprises nécessite de prêter attention à chaque détail en prodiguant un accompagnement à la hauteur de l'ambition de nos stagiaires. C'est pourquoi nous savons faire des leaders d'aujourd'hui les champions de demain !





retengr

**Faire du leader  
d'aujourd'hui, le champion  
de demain**