

Comment créer du langage avec l’I.A. ?

Terme français : **Analyse de régression**

Domaine : [Informatique](#) - [Mathématiques](#)

Définition : Type de modèle mathématique qui s’appuie sur les valeurs précédentes pour prédire les suivantes.

Équivalent étranger : Autoregressive model

Terme français : **Apprentissage auto-supervisé (AAS)**

Domaine : [Informatique](#) - [Intelligence artificielle](#)

Définition : Catégorie de l’apprentissage automatique dans laquelle les modèles sont formés à une tâche en utilisant les données elles-mêmes pour générer des signaux de supervision, plutôt que de s’appuyer sur des étiquettes externes fournies par des humains.

Note : L’AAS peut être utilisé pour créer des modèles généralistes plus intelligents, capables d’effectuer plusieurs tâches et d’acquérir de nouvelles compétences sans recourir à une quantité massive de données étiquetées.

Équivalent étranger : Self supervised learning (SSL)

Terme français : **Apprentissage de bout en bout (A2B)**

Domaine : [Informatique](#) - [Intelligence artificielle](#)

Définition : Technique où la machine apprend toutes les étapes d’un processus à partir des données d’entrée brutes, sans aucune extraction manuelle des caractéristiques de ces données.

Note : L’apprentissage de bout en bout relève de l’apprentissage automatique.

Équivalent étranger : End to end learning (E2E learning)

Terme français : **Apprentissage de défense contre les perturbations**

Synonyme : [Formation contradictoire](#)

Domaine : [Informatique](#) - [Intelligence artificielle](#)

Définition : Technique mise en place par une personne qui tente de modifier un modèle

d’apprentissage automatique en lui proposant des petites perturbations, généralement des exemples contradictoires, pour que ce modèle d’apprentissage développe une capacité à se défendre et à anticiper d’autres perturbations.

Équivalent étranger : Adversarial training

Terme français : **Apprentissage ponctuel**

Domaine : [Informatique](#) - [Intelligence artificielle](#)

Définition : Approche d’apprentissage automatique dans laquelle un modèle est capable d’apprendre à partir d’un seul exemple ou d’un très petit nombre d’exemples d’une classe particulière.

Équivalent étranger : One shot learning, few-shot learning

Terme français : **Consigne**

Domaine : [Informatique](#) - [Intelligence artificielle](#)

Définition : Instruction donnée en langage naturel, qui est destinée à une intelligence artificielle et contient des contextes et potentiellement des exemples afin d’obtenir une réponse de la part de cette intelligence artificielle.

Note : Il existe des personnes dont le métier est de concevoir des consignes les plus précises possible, avec plus de contextes, afin d’obtenir une réponse plus pertinente : on les appellera des **ingénieurs consigneurs** (« prompts engineers »).

Équivalent étranger : Prompt

Terme français : **Désambiguïsation du sens des mots**

Domaine : [Informatique](#) - [Intelligence artificielle](#)

Définition : Capacité d’une intelligence artificielle à identifier le sens des mots dans leur contexte, qui permet d’éclaircir des mots comportant plusieurs sens.

Équivalent étranger : Word sense disambiguation (WSD)

Terme français : **Données ininterprétables**

Domaine : [Informatique](#) - [Intelligence artificielle](#)

Définition : Données corrompues, distordues ou incompréhensibles, qui ne sont pas interprétables par une machine ou un système informatique.

Équivalent étranger : Noisy data

Terme français : **Étiquetage morphosyntaxique**

Domaine : [Informatique](#) - [Intelligence artificielle](#)

Définition : Processus permettant à un algorithme de lire une phrase et de catégoriser la syntaxe et la grammaire, c’est-à-dire les verbes, les noms, les sujets, etc. en analysant cette phrase plus précisément et en étiquetant certains de ces éléments.

Équivalent étranger : Part of speech tagging (POS tagging)

Terme français : **Hiérarchiseur**

Domaine : [Informatique](#) - [Intelligence artificielle](#)

Définition : Composant de l’algorithme qui attribue un niveau d’importance aux éléments d’une image ou d’un texte pour les comprendre et discerner les plus pertinents.

Équivalent étranger : Attention model

Terme français : **Intervalle de Levenshtein**

Domaine : [Informatique](#) - [Mathématiques](#)

Définition : Processus mathématique permettant de déterminer la distance entre deux chaînes de caractères en comptant le nombre de substitutions, d’insertions et de suppressions qu’il faut effectuer pour passer de l’une à l’autre.

Équivalent étranger : Levenshtein distance

Terme français : **Majoration de l’espace des données (MED)**

Synonyme : [Augmentation des données](#)

Domaine : [Informatique](#) - [Intelligence artificielle](#)

Définition : Technique qui permet d’accroître artificiellement la quantité de données utilisée par les outils d’apprentissage.

Équivalent étranger : Data augmentation

Terme français : **Modèle de très grande envergure**

Domaine : [Informatique](#) - [Intelligence artificielle](#)

Définition : Modèle d’apprentissage automatique qui permet d’appréhender des données très nombreuses ayant trait au langage naturel en quelques millisecondes.

Équivalent étranger : Large Language model (LLM)

Terme français : **Peaufinage de données**

Domaine : [Informatique](#) - [Intelligence artificielle](#)

Définition : Méthode servant à compléter l’entraînement d’une intelligence artificielle grâce à l’importation de données d’un corpus spécifique.

Note : Le peaufinage de données et l’ingénierie de consigne (« prompt engineering ») sont presque équivalents.

Équivalent étranger : Fine tuning

Terme français : **Reconnaissance de mots clés (RMC)**

Domaine : [Informatique](#) - [Intelligence artificielle](#)

Définition : Processus d’analyse d’un texte par une intelligence artificielle qui, entraînée grâce à l’apprentissage profond, en définit les mots clés.

Notes : 1. La RMC utilise la technologie de traitement du langage naturel (« Natural language process »), qui est un processus de reconnaissance de texte pour analyser le texte soumis.

2. Les mots clés (« named entities ») sont des lieux, des dates, des noms, des thèmes, des adjectifs, etc.

3. La RMC peut être utilisée en entreprise, par exemple pour gérer les informations à trier lorsqu'elles sont trop nombreuses.

Équivalent étranger : Named entity recognition (NER)

Terme français : **Réseau antagoniste génératif (RAG)**

Domaine : Informatique - Intelligence artificielle

Définition : Système d'apprentissage informatique qui utilise deux réseaux de neurones dont l'un cherche à tromper l'autre afin de générer de nouvelles données réalistes.

Note : Un réseau antagoniste génératif peut par exemple créer, à partir de vraies photos, une photo nouvelle.

Équivalent étranger : Generative adversarial network (GAN)

Terme français : **Résoloférence**

Synonyme : **Résolution de coréférence**

Domaine : Informatique - Intelligence artificielle

Définition : Tâche qui consiste à trouver toutes les expressions qui font référence à la même entité dans un texte écrit dans un langage naturel.

Note : La résoloférence est utilisée dans la traduction automatique.

Équivalent étranger : Coreference resolution

Terme français : **Technologie d'interprétation du langage humain (TILH)**

Synonyme : **Technologie de traitement du langage naturel (TLN)**

Domaine : Informatique - Intelligence artificielle

Définition : Technologie qui permet aux machines d'interpréter dans son véritable contexte le langage humain et de le traiter sous différentes formes.

Note : La TILH est utilisée dans la traduction, la reconnaissance vocale, la classification de textes, la génération de résumés, la réponse automatique aux consignes, la détection de sentiments et l'analyse d'opinion.

Équivalent étranger : Natural language processing (NLP)

Terme français : **Tromperie**

Domaine : Informatique - Intelligence artificielle

Définition : Transmission de fausses informations par des moteurs d'intelligence artificielle.

Note : Un dialogueur (« Chatbot ») qui invente un chiffre d'affaires pour une entreprise sans avoir de données à ce sujet est un exemple de tromperie.

Équivalent étranger : Hallucination

Terme français : **Unité de texte**

Domaine : Informatique - Intelligence artificielle

Définition : Partie de texte permettant à une intelligence artificielle de comprendre et de construire du langage.

Note : Ce langage peut être complexe et contenir des synonymes et différents types de constructions de phrase.

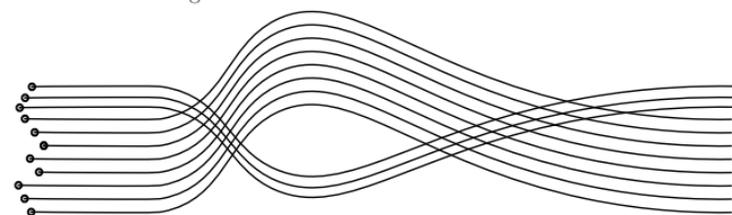
Équivalent étranger : Text token

Terme français : **Vectorisation**

Domaine : Informatique - Intelligence artificielle

Définition : Méthode d'encodage qui n'utilise pas d'intégrations pré-entraînées pour représenter les mots ou les phrases d'un texte mais qui crée ses propres vecteurs représentant ces mots ou ces phrases.

Équivalent étranger : Word embedding



L'intelligence artificielle et le langage naturel... car telle est désormais l'appellation du langage humain, si l'on adopte la perspective de la machine, de l'algorithme. Où l'on verra, grâce à cette liste de termes et de concepts élaborés par les élèves du lycée Raspail, que la machine suit des apprentissages de plus en plus élaborés, seule ou sous la supervision d'ingénieurs consigneurs, qu'elle s'adonne à des tâches proprement linguistiques que n'aurait pas dédaignées Ferdinand de Saussure (étiquetage morpho-syntaxique, hiérarchisation d'éléments textuels, désambiguïsation du sens, résolution de coréférence, etc.), et que pour accomplir cela elle brasse des quantités immenses de données jusqu'à un ultime travail de peaufinage. Parfois même, elle se trompe. Mais l'erreur n'est-elle pas... humaine ?

Bravo à :

Adnan, Alvin, Amar, Amine, Antonin, Chloé, Ethan, Evans, Gabriel, Hamza, Islam, Jawed, Jérémy, Jovan, Nassim, Noa, Omar, Rahim, Rayan, Shanel, Sofia, Yanis

MM. Mazière et Mouzai, Mmes Mosconi et Darry

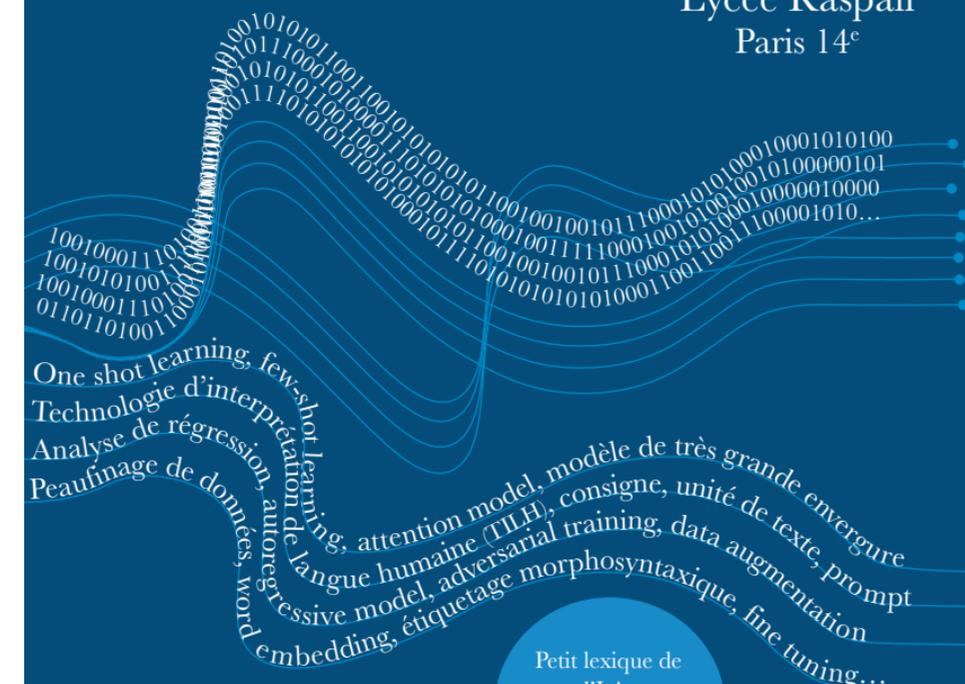
<https://www.dictionnaire-academie.fr>

<https://www.ldmraspail.fr>



Académie française

Lycée Raspail
Paris 14^e



One shot learning, few-shot learning, Technologie d'interprétation de langue humaine (TILH), consigne, unité de texte, prompt, Analyse de régression, autoregressive model, adversarial training, data augmentation, Peaufinage de données, word embedding, étiquetage morphosyntaxique, fine tuning...

Petit lexique de l'IA, conversationnelle

Classe 1^{re} STI2D1

2023 - 2024