Schärz Nicolas, Gymnase Bienne Jura bernois

25J

ChatGPT

[Sous-titre du document]

Table des matières

[Préface 2](#_Toc169177015)

[Introduction 2](#_Toc169177016)

[Fonctionnement 3](#_Toc169177017)

[Introduction 3](#_Toc169177018)

[LLM 3](#_Toc169177019)

[Structure des grands modèles de langages 5](#_Toc169177020)

[NPL 5](#_Toc169177021)

[Références 6](#_Toc169177022)

# Préface

# Introduction

L’intelligence artificielle a fait d’énormes progrès ces dernières années, et l’une des avancées les plus remarquables est la création de modèles de langage tels que ceux développés par OpenAI. Dans le cadre de mon travail de maturité, je vais répondre à la question suivante : Comment ChatGPT est-il utiliser par les étudiants gymnasiens, et quel est la vision de ChatGPT par des enseignants gymnasiens sur le future de l’éducation ainsi que comment ChatGPT fonctionne ?

Pour cela, je vais en premier lieu décrire les nouveaux modèles de langages qui a révolutionné le d’écrire des textes. Pour cela, je vais expliquer comment cette machine ou intelligence artificielle fonctionne en précisant comment elle traite les informations pour pouvoir renvoyer une réponse naturelle et cohérente. De plus, le travail sera plus précis sur comment **entraime** Cela me permettra d’enchainer en énumérant ces limites actuelles qu’elles soient éthiques ou techniques. Ce travail ne portera seulement par les modèles de langages GPT3.5 GPT4, et GPT4o d’OpenAI. Autrement dit, les intelligences artificielles pouvant lire ou créer des images ne seront pas étudier, de même pour le son

Ensuite, le discussion sera portés sur l’utilisation de ChatGPT par les gymnasiens. L’analyse sera portés premièrement sur les cas de tricherie dans le gymnase de Bienne. L’étude permettra de categoriser le phénomène et prouver qui sera plus favorable à utiliser ChatGPT sans en mentionner la source. Comment se déroule les procédures du gymnase s’il y a un cas de tricherie. Ainsi nous verrons comment cela se fait détecter et par quels programmes arrive à trouve les cas de plagiats. Nous partagerons les procédures mis en place par l’établissement. Deuxièmement, nous regarderons comment l’intelligence artificielle de ChatGPT permet d’aider les étudiants gymnasiens. Quels sont les méthodes utilisées et la démarche qu’ils font pour pouvoir obtenir de l’aide. Finalement, la question de l’efficacité réelle de ces solutions reste ouverte et nécessite une évaluation continue.

Enfin, je vais discuter l’intégration de ChatGPT pendant les leçons. Pour cela, il faudra distinguer l’utilisation de l’intelligence artificielle dans des tests, des devoirs… Nous allons voir si les enseignants sont formés pour intégrer ChatGPT dans les leçons pour aider les élèves à comprendre comment utiliser cette ressource pour enrichir leur apprentissage plutôt que pour faciliter le plagiat. Ont-ils des difficultés à incorporer Chatgpt. Mais encore, on pourrait se demander s’il faudrait éduquer et enseigner les gymnasiens sur comment utiliser ChatGPT éthiquement pour défavoriser les cas de plagiat. Dernièrement, faudrait-il l’utilisation de Chatgpt lors de travaux écrits et de quelles manières est-ce que cela pourrait se faire

Afin de remplir mes objectifs je vais utiliser des questionnaires pour découvrir le nombre de cas de tricherie au sein de l’établissement du gymnase. Cette méthode me permet de créer une nouvelle étude sur la tricherie qui n’a pas encore était réaliser. En plus, l’avantages est que j’obtiendrais des réponses sur de plus vaste sujet comme l’intégration de ChatGPT durant les leçons. Je vais aussi interviewer un expert sur la sécurité informatique au gymnase. En discutant directement avec l'expert, je pourrai poser des questions spécifiques sur la manière dont le gymnase détecte les situations de plagiat, en particulier en utilisant des outils comme ChatGPT. Je pourrai ainsi comprendre les techniques et les technologies utilisées pour identifier les cas de tricherie, ainsi que les défis rencontrés dans ce processus. En combinant les données des questionnaires et les informations de l'interview, je pourrai fournir une vue d'ensemble complète et approfondie sur la tricherie au sein du gymnase, les impacts de l'utilisation de ChatGPT, et les mesures de sécurité informatique.

# Fonctionnement

## Introduction

Les intelligences artificielles ont évoluées de plus en plus rapidement. Ils commencent à créer des textes ressemblant à comment un humain écrirait.

## LLM

Les « LLM » ou grands modèles de langages sont des intelligences artificielles ayant des capacités qui ne se spécifie pas à une catégorie, mais au contraire excellent à réaliser plusieurs tâches. Il y a par exemple, le plus connu, l’écriture de texte avec des ChatBot **idée**. Cela est du LLM qui a pour but de prédire des tokens. Les tokens sont des mots, parties ou mêmes groupes de mots que l’intelligence artificielle met en relations pour prédire le future Tokens

## Tokenisation

Une image contenant texte, capture d’écran, logiciel, Logiciel multimédia

Description générée automatiquement

Figure 1

On remarque ici, que l’IA sépare en différent Tokens le mots long avec leurs suffixes et les majuscules. Un Token vaut généralement 4 charactères donc trois quarts d’un mot. Ces tokens ont pour avantages de ne pas avoir une langue précise et permettent une compréhension de plusieurs langues

C’est un réseau neuronal une entrée une séquence de mots qui sera, ensuite, dispersé dans le réseau neuronal. Le mot avec la probabilité la plus élevée est choisi comme prédiction, en tenant compte de la cohérence et du contexte appris pendant l'entraînement. Le modèle compare le contexte actuel avec ses connaissances pour garantir que le mot prédit est logique et cohérent, ressemblant à des réponses déjà existantes et gardant un sens logique.

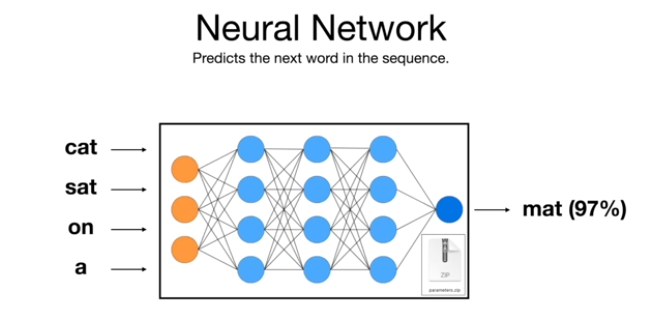


Figure 2

https://youtu.be/zjkBMFhNj\_g

### Pre-training

Pour pouvoir prédire le mot le plus cohérent il doit été entraîné sur une grande quantité d’informations. C’est ce qu’on appelle le **Pre-Training**. Des experts rassemblent un corpus diversifiés de données variées de nombreuses sources comme des livres, des sites internet, des réseaux sociaux et des articles… Certaines sources auront plus de poids que d’autres pendant l’entraînement. Ensuite, ces données sont nettoyés pour garantir leur qualité et leur pertinence. D’abord il y aura une suppression des doublons pour éviter d’avoir plusieurs textes redondants qui disent la même chose et donc prennent inutilement de l’espace. Les textes sont ensuite filtrés pour retirer des contenues inappropriés comme des publicités. On appelle cela des « bruits ». Ces données sont ensuite convertir en Tokens puis utilisées pour entrainer les modèles neuronaux. C’est un processus très couté et qui prend énormément de temps. C’est pourquoi le dernier fait par OpenAI date de de 2023 et donc ChatGPT d’OpenAI n’est pas mis à jour depuis 2023

### Fine-tuning

Elle consiste à affiner les performances d’un modèle pré-entraîné sur une nouvelle tâche, tout en préservant les connaissances acquises précédemment. Le principe clé est la capacité du modèle à s’adapter à de nouvelles situations sans perdre sa spécialisation initiale. Cette technique ajuste les paramètres du modèle pour mieux répondre aux exigences spécifiques de la tâche ciblée, sans changer significativement les apprentissages passés.

Après une apprentissage ou on montre à la machine une bases de données, il peut reconnaitre des objets ou des voix dans des contextes particuliers. Voici un exemple avec un assistant vocal intelligent. D’abord, il est entraîné pour comprendre et répondre à des commandes vocales dans une langue spécifique avec un accent général. Le modèle pourrait être fine-tuné pour reconnaître de nouvelles langues et des nouveaux accents. De cette manière, le Fine-Tuning permet au modèle de rester performant dans sa tâche originale tout en s’adaptant à de nouvelles exigences linguistiques et culturelles.

**Etape** https://datascientest.com/fine-tuning-tout-savoir

## Structure des grands modèles de langages

Une image contenant texte, capture d’écran, cercle, Police

Description générée automatiquement

Figure 3 https://inwedo.com/blog/llm-ml-models-in-business/

Pour comprendre plus précisément les grands modèles de langages, il faut distinguer de quoi il en est construit. Les grands modèles de langages résultent de la combinaison de l’apprentissage automatisé « Machine Learning, ou ML », apprentissage profond « Deep Learning, DL » et du traitement du langage naturel « Natural Language Processing, NLP ». Pour mieux comprendre les LLM nous allons expliquer chacun des composants.

### NPL

Le traitement des langages naturels doit permettre aux ordinateurs d’interpréter de comprendre et de générer des textes en langages humains. Son but principal est lier la compréhension des ordinateurs avec les interactions humaines, ce qui permet aux machines d’analyse et de répondre en conséquence dans un langage naturel. (Amaratunga, 2023)

Pour réaliser cette liaison les NPL ont été assigner à quelques tâches

* **Reconnaissance vocale** : Convertir la parole en texte écrit.
* **Synthèse vocale (Text to Speech)** : Générer des paroles à partir de texte écrit
* **Traduction automatique**
* **Résumé automatique**
* **Génération de texte**
* **Reconnaissance d’entités nommées (NER)**
* **Classification de texte :** assigner une étiquette à un morceau de texte, comme pour la détection de spam, analyser le sentiment d’un texte
* **Trouver des informations :**

Pour réaliser c'est tâche le NLP doit passer d'abord en une phase de prétraitement avant de passer à l'analyse ou au traitement linguistique avec les tâches précédemment énuméré. Cette phase consiste à préparer les données pour qu'elles soient dans un format approprié et facilement compréhensible par les traitement des langues naturels. Voici les étapes :

* **Tokenisation**
* **Nettoyage de données inutile**
* **Retirer les mots vides (avec peu de sens)**
* **Stemming et lemmatisation**
* **Analyse morphologique**

Qu’est-ce qui différencie les grands modèles de langages et le traitement du langage naturel ? Premièrement les grands modèles de langages sont comme expliqué précédemment spécialisé sur une multitude de tâches. Tandis que le traitement du langage naturel est spécialisé pour un nombre de tâche restreint. Par exemple, Un LLM peut traduire un texte de l’anglais vers le français, résumer un article, et répondre à des questions sur ce texte, alors qu’un système spécialisé en analyse sentimentale qui évalue les émotions exprimées dans les avis des utilisateurs sur un produit spécifique. De plus, les LLM sont entraînés sur d’énormes ensembles de données textuelles et prennent plus de temps en entraînement que les NPL. En résumé c’est un domaine plus large qui englobe toutes les techniques permettant aux ordinateurs de comprendre et de manipuler le langage humain, y compris les LLM.

### Machine Learning

Comme montré sur l’image le machine learning est un type d’intelligence artificielle. Le machine Learning « Apprentissage automatisé, ML » a pour but de permettre aux ordinateurs d’apprendre sans qu’il y soit programmer précisément. Ainsi la machine qui sera confronté à des nouveaux cas pourra sans s’adapter de manière autonome et effectuer des tâches de plus en plus complexe. L’apprentissage automatisé permet donc de s’ajuster seul, à chaque nouveaux cas avec précision même s’il n’a pas été programmé pour ce cas spécifique. Il utilise des méthodes algorithmiques et des modèles basés sur la statistique pour identifier des motifs et des corrélations. Après l’analyse des données, il tente de faire des prédictions sur les résultats possibles. Le concept central repose sur l’exploitation des données issues d’expériences antérieures afin de prédire des résultats ou des valeurs inconnus. Il existe trois grands moyen d’entraînement

#### Apprentissage supervisé

#### Apprentissage non supervisé

### Deep learning

# Références

Amaratunga, T. (2023). *Understanding Large Language Models.* Nugegoda.

Une image contenant texte, capture d’écran, Police

Description générée automatiquementCapture :