Intelligence Artificielle
Projet AI SPOTTER

Étude des outils permettant l'analyse et la création d'un détecteur de textes générés par IA et utilisation des RCR et du ML pour répondre à une partie de la problématique.

Année universitaire 2023-2024

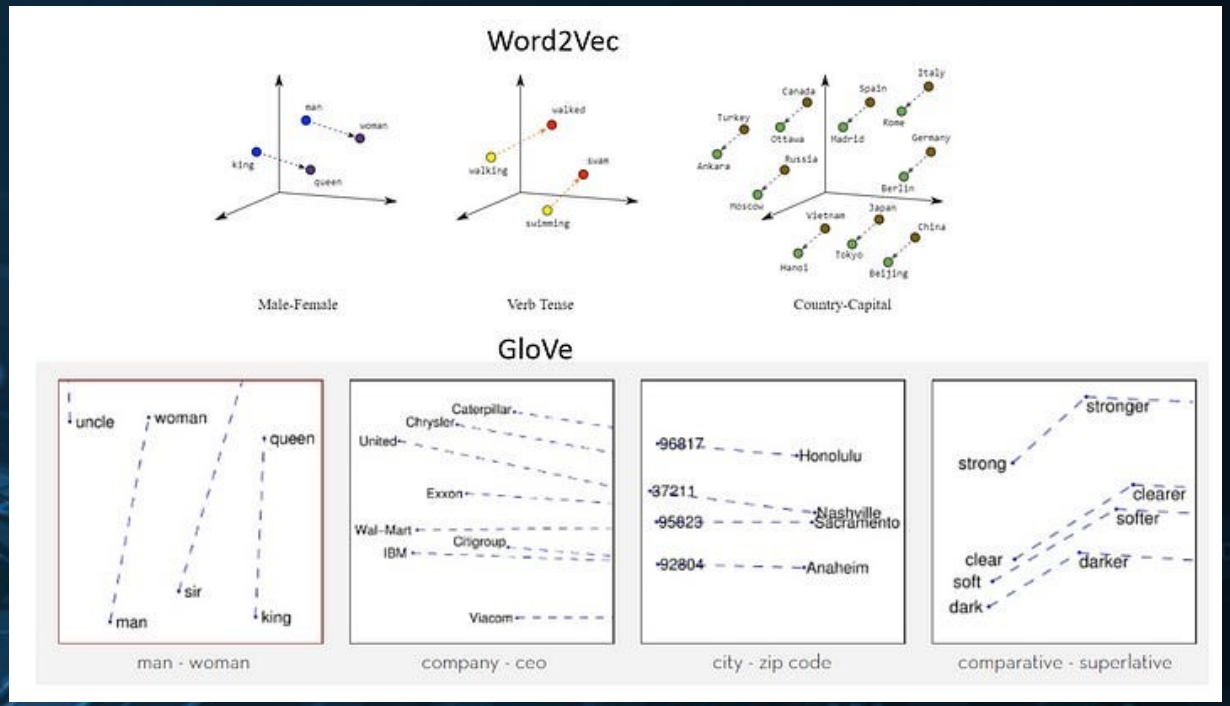
Lorenzo BARBEY
Axel FEVEZ

L’analyse distributionnelle comme outil de développement du plongement lexical

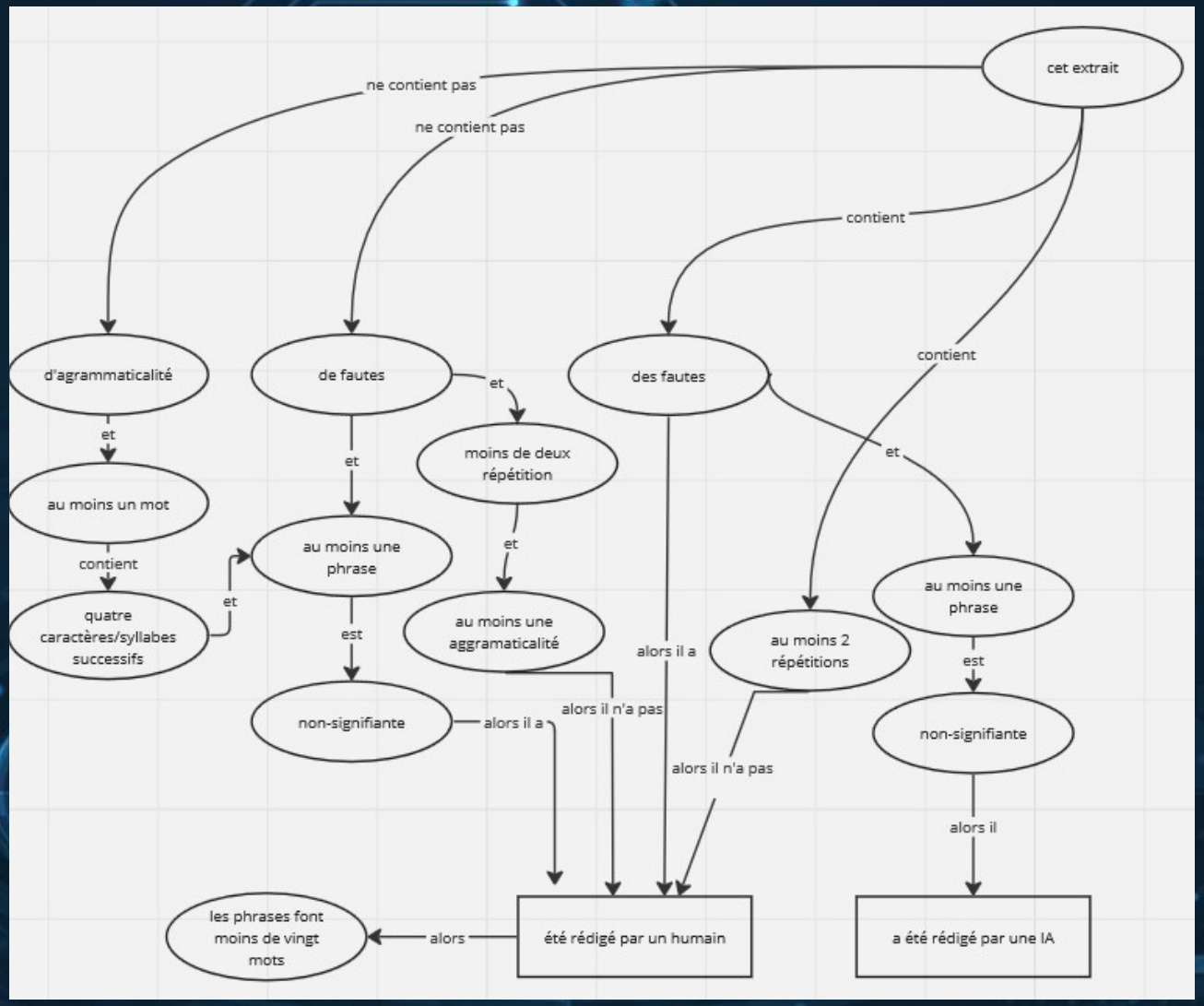
L’image à droite présente la représentation graphique de Word2Vec, il exprime la proximité entre deux mot par la proximité et la direction qui relie deux mots différents.

Sur les différents exemples, on volt que l’association se fait par catégorie et que dépendamment d’elle, les associations se font différemment.

Le fonctionnement derrière Word2Vec est l’utilisation à double couche, avec deux modèles possibles qui sont le CBOW : « Modèle de sacs de mots continues » et le modèle Skip-Gram qui se base sur un algorithme de matrices.



Représentation des connaissances



Graphe de connaissances du langage présenté dans le rapport.

Comme on peut le voir, on peut représenter notre langage de façon visuel afin de permettre de mieux analyser et comprendre notre problématique.

De manière similaire, il existe d’autres écritures et représentations différentes qui permettent d’être plus ou moins précis, d’être plus ou moins rigoureux (notamment sur le plan mathématique et logique par exemple)

Configuration du Machine Learning

- ## Dataset Utilisé

  Le dataset utilisé dans notre projet provient de Kaggle et se compose d'une variété de textes générés, provenant aussi bien d'intelligences artificielles que d'êtres humains.

  Il comprend 1000 textes en anglais, chaque entrée étant structurée avec un identifiant, un texte associé, et un entier indiquant si le texte a été généré par un être humain (0) ou par une intelligence artificielle (1).

  Voici un exemple d'une donnée:

  1,"There are a variety of emerging applications for NLP, including the following:, voice-controlled computer interfaces (such as in aircraft cockpits), programs that can assist with planning or other tasks ,, more-realistic interactions with computer-controlled game characters, robots that interact with humans in various settings such as hospitals, automatic analysis or summarization of news stories and other text, intelligence and surveillance applications (analysis of communication, etc.), data mining, creating consumer profiles, and other ecommerce applications, search-engine improvements, such as in determining relevancy",0

- ## Paramètres Utilisés

  Nous avons configuré notre modèle en utilisant les paramètres suivants :

  - ❑ Taille du Vocabulaire (vocab_size): 1000 - Limitant le modèle aux 1000 mots les plus fréquents pour se concentrer sur les termes les plus pertinents.
  - ❑ Dimension de l'Espace d'Embedding (embedding_dim): 50 - Définissant la représentation vectorielle de chaque mot dans notre vocabulaire.
  - ❑ Longueur Maximale d'une Séquence (max_sequence_length): 1500 - Fixant la limite de mots examinés par le modèle pour chaque article.

```
vocab_size = 1000
embedding_dim = 50
max_sequence_length = 1500
```

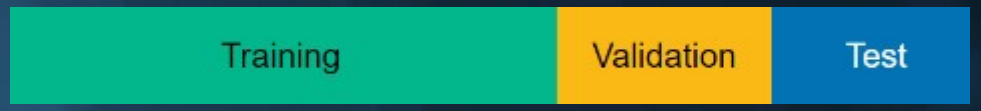
Entraînement du Modèle

- **Séparation de l'Ensemble de Données**

Nous avons utilisé la méthode train_test_split pour diviser nos données en ensembles d'entraînement, de validation et de test.

Répartition des données :

- ✓ Ensemble d'entraînement : 70%
- ✓ Ensemble de validation : 15%
- ✓ Ensemble de test : 15%



Pour maintenir la distribution IA/Humain de manière équilibrée dans chacun de ces ensembles, nous avons opté pour une approche de stratified split. Cette méthodologie assure que le modèle est exposé à une diversité suffisante tout en conservant une représentation juste des deux sources de texte.

- **Résultats de l'Entraînement**

  - Précision sur l'ensemble d'entraînement : ~99.02%
  - Perte sur l'ensemble d'entraînement : ~0.0460
  - Précision sur l'ensemble de validation : ~90.20%
  - Perte sur l'ensemble de validation : ~0.3512

Le modèle a vraiment bien appris avec notre collection de textes, atteignant une précision élevée d'environ 99.02% sur ceux qu'il avait vus auparavant. Cela signifie qu'il a bien compris les détails subtils des textes générés par des humains et des IA grâce aux paramètres spécifiques que nous avons choisis.

Cependant, il faut garder à l'esprit que ces résultats sont basés sur notre groupe de textes particulier. On se demande si le modèle maintiendra cette performance lorsque confronté à une plus grande variété de textes générés par IA. Pour le savoir, nous devons le tester sur différentes sources pour avoir une idée plus précise de sa performance globale.

Architecture du Projet

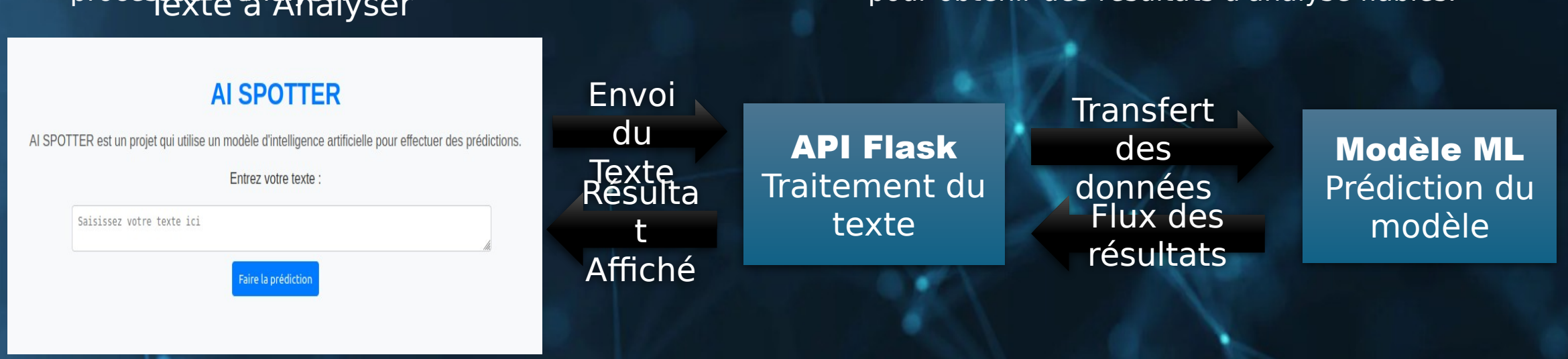
• Interface Utilisateur et API Flask

- ❖ L'interface utilisateur a été soigneusement conçue en utilisant les langages HTML, CSS et JavaScript pour offrir une expérience facile d'interaction.
- ❖ L'intégration d'une API Flask en Python ajoute une couche fonctionnelle à l'interface. Elle permet de tester les textes en les envoyant au modèle de machine learning, facilitant ainsi le processus d'analyse.

• Acheminement des Données

Les données provenant de l'interface utilisateur sont acheminées de manière transparente vers le modèle de machine learning. Cela garantit une interaction fluide entre l'utilisateur et le système d'analyse.

Le processus d'acheminement des données assure une transmission efficace, garantissant que les textes soumis sont traités de manière optimale pour obtenir des résultats d'analyse fiables.

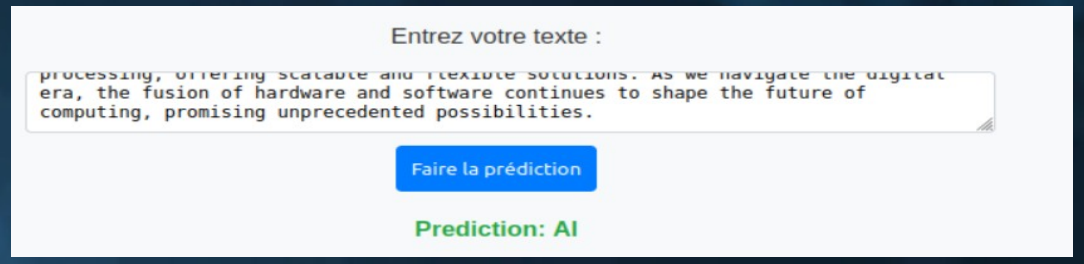


Résultats et perspectives d’amélioration

• Résultats du Modèle

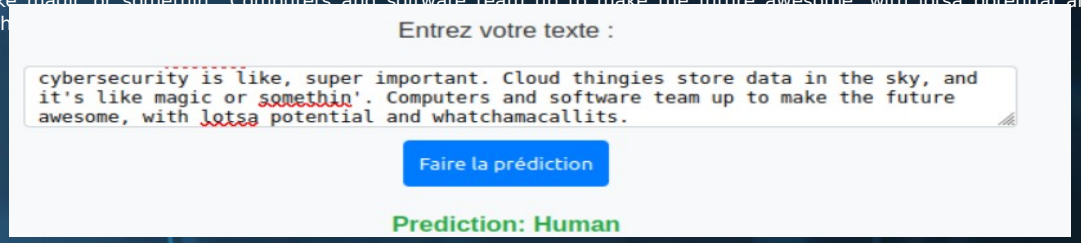
Exemple IA, texte généré par ChatGPT:

“The field of computer science is an ever-evolving landscape, characterized by constant innovation and technological advancements. In this dynamic realm, professionals explore the intricacies of algorithms, data structures, and programming languages to develop efficient and robust software solutions. The pursuit of artificial intelligence and machine learning has gained prominence, with researchers delving into the realms of neural networks and deep learning algorithms. Cybersecurity stands as a critical pillar, safeguarding digital assets from malicious threats. Cloud computing has revolutionized data storage and processing, offering scalable and flexible solutions. As we navigate the digital era, the fusion of hardware and software continues to shape the future of computing, promising unprecedented possibilities.”



Exemple Humain, rédigé par Lorenzo avec insistance sur les répétitions et les fautes linguistiques :

“The computer science stuff is always changin', ya know? Like, there's these thingies called algorithms and data stuff, and we use 'em to make computer thingamajigs work. It's all about programmin' languages and makin' softwares that do cool stuff. People are super into makin' fake smart things with artificial brains and learnin' machines. Gotta watch out for them hacker dudes, 'cause cybersecurity is like, super important. Cloud thingies store data in the sky, and it's like magic or somethin'. Computers and software team up to make the future awesome, with lotsa potential and wh



• Perspectives d'Amélioration

- **Gestion des Fautes**: Explorer des techniques avancées de correction orthographique pour renforcer la précision du modèle face aux fautes linguistiques.
- **Détection de Phrases Grammaticalement Incorrectes**: Intégrer des mécanismes pour améliorer la capacité du modèle à détecter les phrases qui ne respectent pas les règles grammaticales.
- **Expansion de la Base de Données**: Élargir la base de données pour améliorer la capacité du modèle à généraliser sur une plus large gamme de textes générés par IA.
- **Ouverture à d'Autres Langues**: Envisager l'extension du modèle pour inclure d'autres langues.