

Mini-Projet 2: **Détection des fausses nouvelles à l'aide des Transformateurs**

Membres de Groupe Réalisateur :

BARBARA Racha
ZERHERI Fatima Zahra

Encadré par :

Mr HAMIDA Soufiane

PLAN

1. Définition et Objectif
2. Collecte et Préparation de données
3. Architecture du modèle
4. Résultats
5. Conclusion

Définition et Objectif

1. Définition:

La détection des fake news est cruciale dans un monde où l'information se propage rapidement en ligne. Les fausses informations peuvent avoir des conséquences graves, telles que la manipulation de l'opinion publique et des décisions erronées. Il est donc vital de développer des outils efficaces pour garantir une information fiable et protéger le public.

2. Objectif:

Créer un modèle capable de Classer les articles en fake ou réel, visant à améliorer la précision de la détection des fausses nouvelles en combinant les atouts de BERT et de GPT. BERT excelle dans la classification de texte, tandis que GPT fournit une étape de vérification supplémentaire grâce à ses capacités de génération de texte.

Collecte et Préparation de données

- **Source des données** : Jeu de données "**fake-and-real-news-dataset**" importé depuis KAGGLE.
- **Organisation des données** : Le dataset comprend 63 678 textes d'actualité réels et faux, avec 21 197 textes pour chaque classe "fake" et "real".
- **Nettoyage des textes** : conversion en minuscules, suppression des URL, balises HTML, ponctuation, chiffres et mots inutiles. Suppression des espaces multiples et des retours à la ligne.
- **Séparation des données** : Division du dataset en ensembles d'entraînement et de test (80% pour l'entraînement, 20% pour le test).
- **Encodage des étiquettes** : Conversion des étiquettes textuelles en indices numériques.
- **Conversion en CSV** : Stockage des textes et étiquettes dans des fichiers CSV "train.csv" et "test.csv", prêts pour l'entraînement et l'évaluation du modèle.
- **Résultat** : Une base de données propre, équilibrée et prête pour l'entraînement du modèle, avec des fichiers CSV "train.csv" et "test.csv".

Collecte et Préparation de données



	title	text	target
0	breaking exclusive black republican fired from...	for just a little over four months wayne bradl...	0
1	trump calls russia reports fake news a total p...	presidentelect donald trump in an apparent ref...	1
2	obama pretends he hasn't started a race war...pu...	it would appear that barack didn t get the mem...	0
3	billionaire gop donors refuse to crown king tr...	in when we were still smarting from the disast...	0
4	hillary clinton and bill blasio's racist 'joke...	new york city s inner circle dinner is typical...	0
...
42389	the ladies of 'the view' call trump exactly wh...	tuesday morning on the view the panel discusse...	0
42390	clinton scorches trump 'all his business books...	hillary clinton has been gearing up for months...	0
42391	nz first leader says foreign ownership to be p...	wellington the party holding the balance of po...	1
42392	obama says not worried about being last democr...	lima us president barack obama on sunday said ...	1
42393	vp pence in the baltics voices support for mut...	tallinn us vice president mike pence on monday...	1

42394 rows × 3 columns

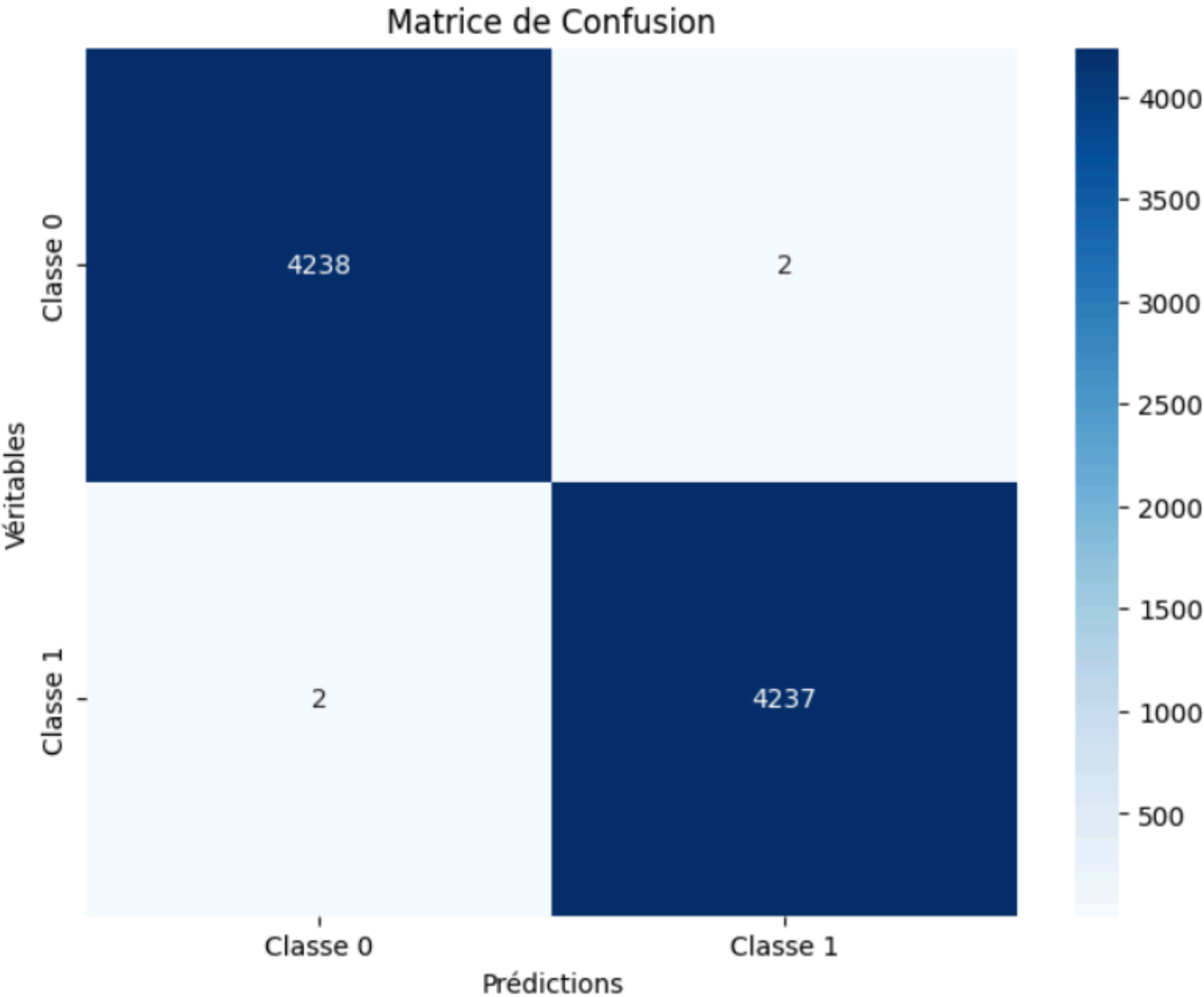
Architecture du modèle BERT

- **Tokenisation et prétraitement des données** : Le modèle utilise un tokenizer BERT pour transformer les textes en séquences de tokens, qui sont ensuite convertis en identifiants numériques pour l'entraînement.
- **Modèle BERT pour la classification** : Un modèle pré-entraîné **AutoModelForSequenceClassification** est utilisé pour la classification binaire des fake news. Le modèle est fine-tuné et utilisé tel quel.
- **Configuration de l'entraînement** : L'entraînement est géré par la classe Trainer de Hugging Face, avec des paramètres comme le nombre d'époques, la taille des lots et l'évaluation après chaque époque. Les calculs sont effectués en précision flottante 16 bits pour accélérer le processus.
- **Calcul des métriques de performance** : Les métriques comme la précision, le rappel, le score F1 et l'exactitude sont calculées durant l'entraînement, et une matrice de confusion est générée pour évaluer les résultats sur les données de test.

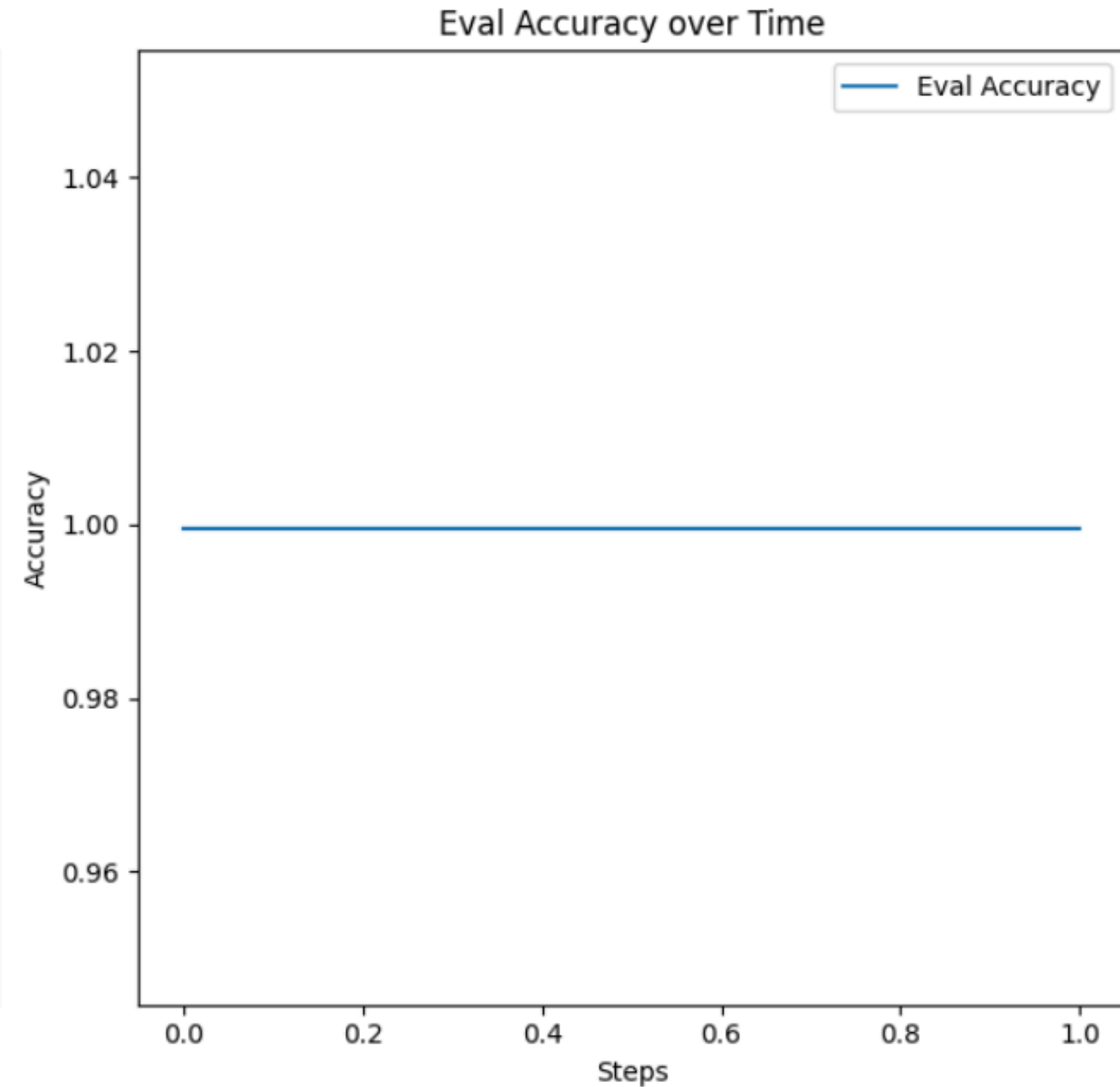
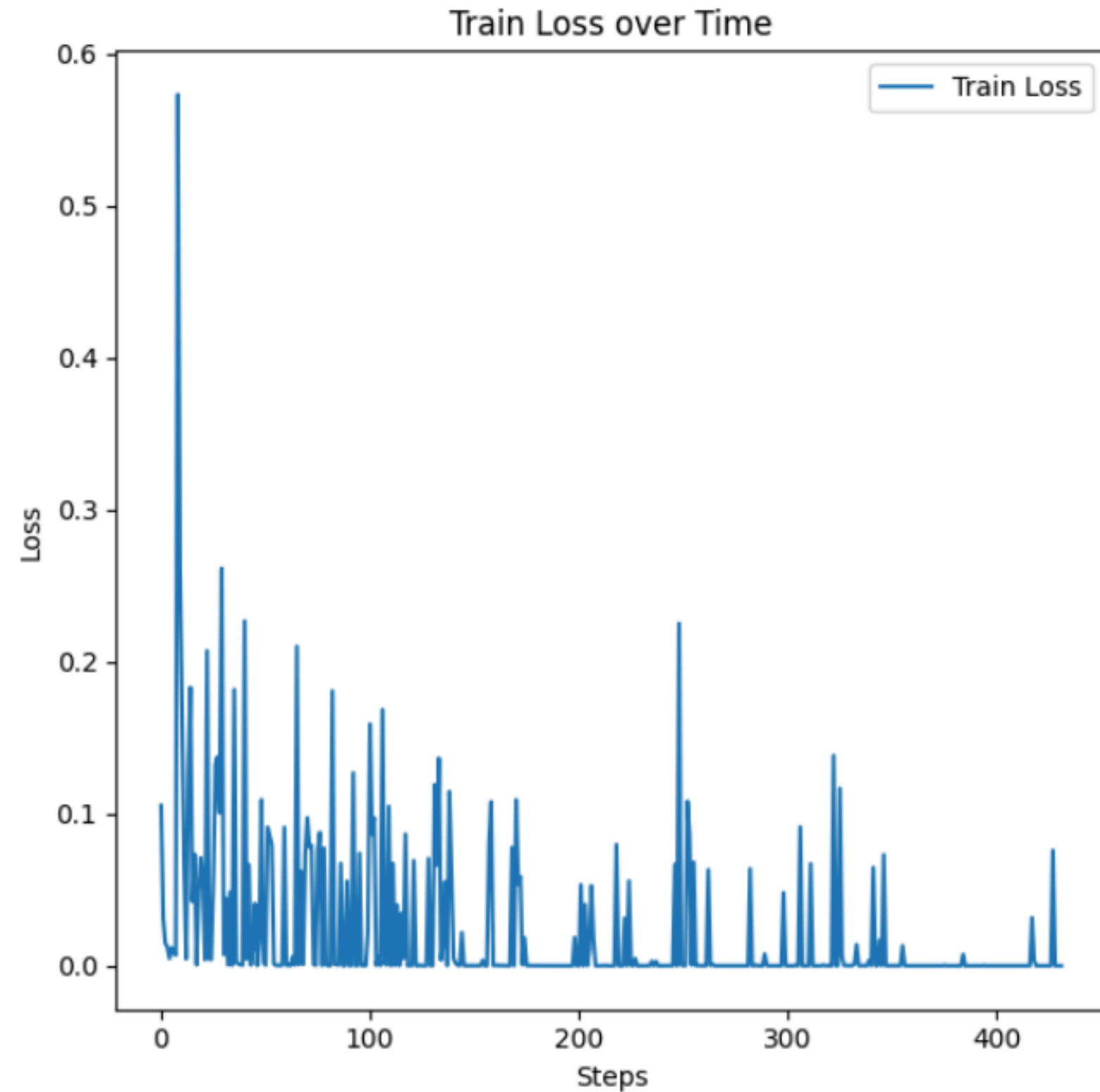
Architecture du modèle BERT

Début de l'entraînement sur 1 époque(s).
[4240/4240 20:09, Epoch 1/1]

Epoch	Training Loss	Validation Loss	Accuracy	F1	Precision	Recall
1	0.000000	0.001992	0.999646	0.999646	0.999646	0.999646



Architecture du modèle BERT



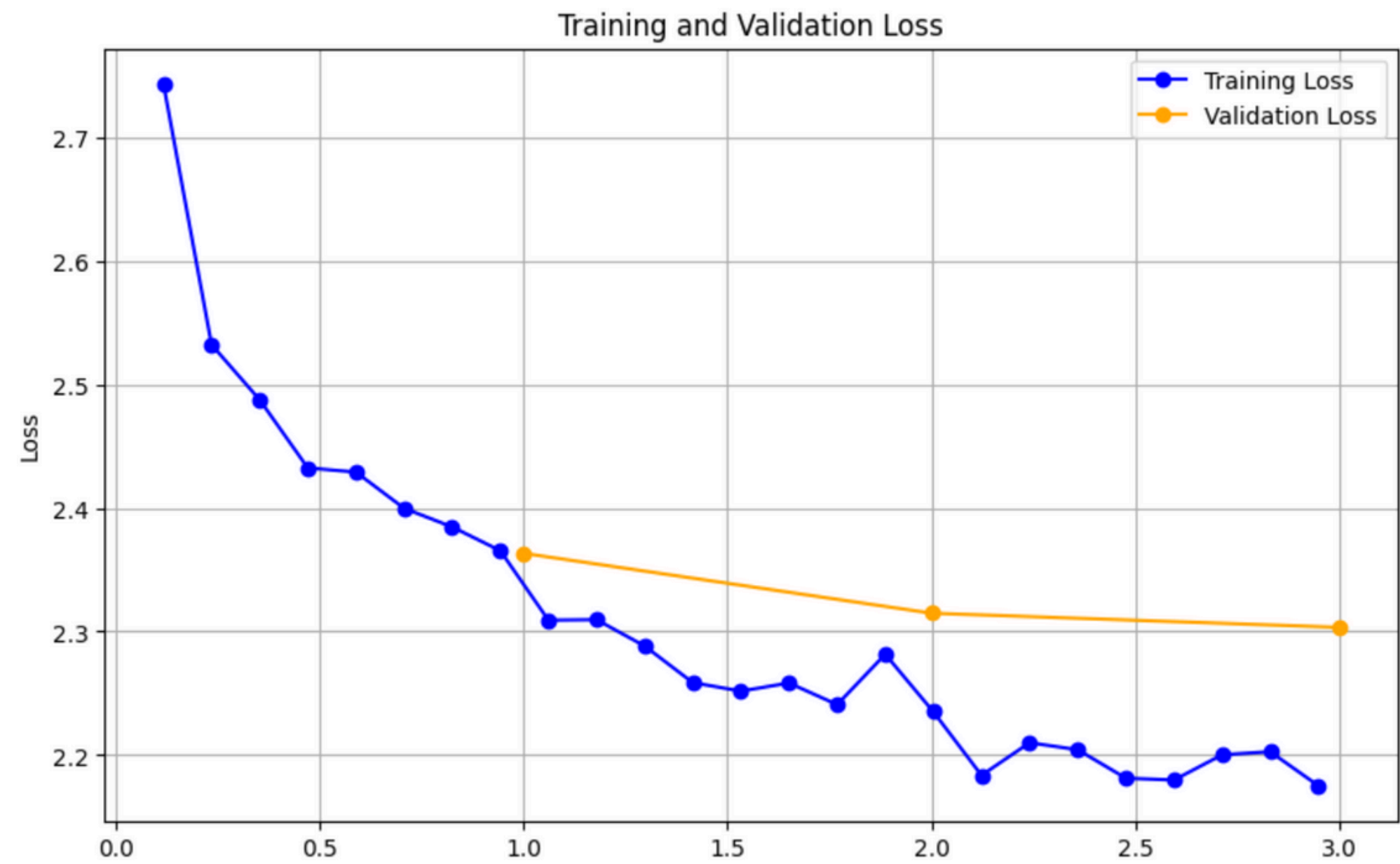
Architecture du modèle GPT

- **Tokenisation et préparation des données** : Le tokenizer GPT-2 découpe les textes en sous-mots compréhensibles par le modèle. Le token de fin de séquence (eos_token) est utilisé comme padding.
- **Modèle GPT pour l'analyse de texte** : AutoModelForCausalLM, pré-entraîné pour la génération de texte causal, est fine-tuné pour prédire des séquences adaptées à la tâche cible.
- **Configuration de l'entraînement** : La classe Trainer de Hugging Face gère l'entraînement avec des paramètres à définir comme les époques, la taille des lots et l'évaluation régulière. Les calculs utilisent la précision flottante 16 bits pour optimiser la vitesse.
- **Entraînement avec Trainer** : Trainer automatise l'entraînement, la validation et le calcul des métriques en utilisant le modèle, les arguments d'entraînement, et les datasets.

Architecture du modèle GPT

[12720/12720 2:14:24, Epoch 3/3]

Epoch	Training Loss	Validation Loss
1	2.365400	2.363275
2	2.281600	2.314851
3	2.175300	2.303364



Combinaison des modèles

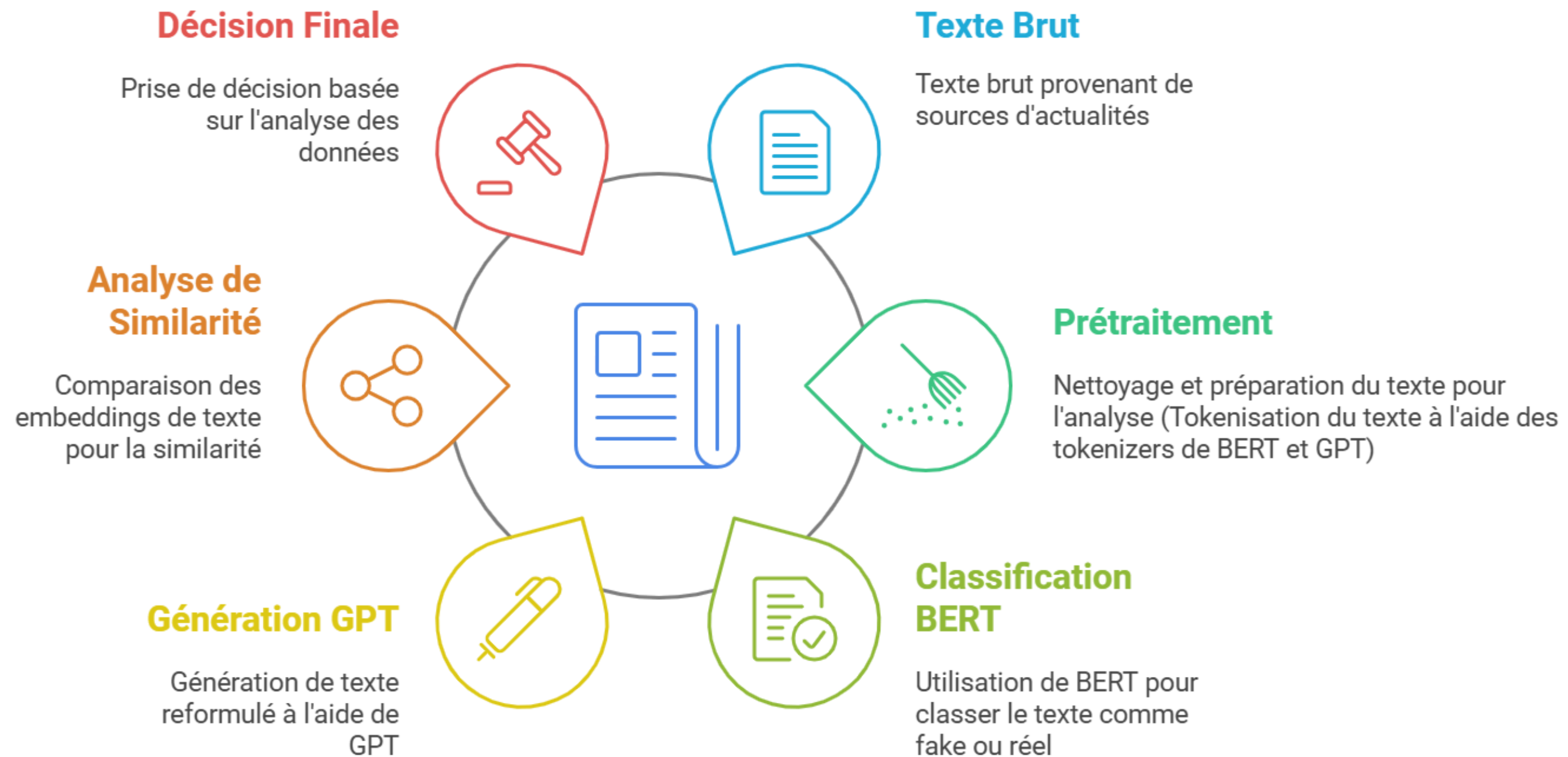
Une fonction **FakeDetection** combine les forces de deux modèles pour détecter les fausses nouvelles :

- **Détection via BERT** : Un texte est identifié comme une fausse nouvelle si BERT le classe comme tel.
- **Analyse de similarité avec GPT** : Si la similarité entre le texte généré par GPT et le texte original est faible, cela renforce l'identification comme fausse nouvelle.

Cette approche hybride permet d'améliorer significativement la précision de la détection en croisant deux perspectives complémentaires.

Combinaison des modèles

Architecture pour la Détection de Fake News



Résultats

News 1 : Real News Detected.
News 2 : Real News Detected.
News 3 : Real News Detected.
News 4 : Real News Detected.
News 5 : Real News Detected.
News 6 : Real News Detected.
News 7 : Real News Detected.
News 8 : Real News Detected.
News 9 : Real News Detected.
News 10 : Real News Detected.
News 11 : Real News Detected.
News 12 : Real News Detected.
News 13 : Real News Detected.
News 14 : Fake News Detected.
News 15 : Fake News Detected.
News 16 : Fake News Detected.
News 17 : Fake News Detected.

News 21 : Fake News Detected.
News 22 : Fake News Detected.
News 23 : Fake News Detected.
News 24 : Fake News Detected.
News 25 : Fake News Detected.
News 26 : Fake News Detected.
News 27 : Fake News Detected.
News 28 : Fake News Detected.
News 29 : Fake News Detected.
News 30 : Fake News Detected.
News 31 : Fake News Detected.
News 32 : Fake News Detected.
News 33 : Fake News Detected.
News 34 : Fake News Detected.
News 35 : Fake News Detected.
News 36 : Fake News Detected.
News 37 : Fake News Detected.

Real Acc count : 13
Fake Acc count : 22

CONCLUSION

- Ce projet a mis en évidence l'efficacité de l'approche hybride combinant BERT et GPT pour détecter les fausses nouvelles. BERT excelle dans la classification précise des textes, tandis que GPT permet d'évaluer la cohérence sémantique. Cette synergie améliore la robustesse globale du système.
- Pour renforcer les performances, il sera intéressant de tester d'autres architectures de modèles, afin de diversifier les approches et d'explorer de nouvelles stratégies. Enrichir le modèle avec des données supplémentaires ou affiner davantage les hyperparamètres pourra également contribuer à des performances accrues.
- Ce projet constitue également une avancée significative dans la lutte contre la désinformation et offre un cadre pour des recherches futures. Il ouvre la voie à des applications concrètes dans des domaines variés, tels que la modération de contenu en ligne ou la vérification automatique des faits.

MERCI POUR VOTRE ATTENTION

 Barbara Racha

 Zerheri Fatima Zahrae