

# Hello Typst!

## Table of contents

1 PARTIE I. Accompagnement au changement .....	1
1.1 1. Besoins (Johnny Hélène) .....	1
1.2 Background .....	1
1.3 Description cas d'usage .....	1
1.3.a Classifier des accords d'entreprise .....	2

## 1 PARTIE I. Accompagnement au changement

### 1.1 1. Besoins (Johnny Hélène)

Vision high level de l'intérêt des LLMs

### 1.2 Background

In the case of glaciers, fluid dynamics principles can be used to understand how the movement and behavior of the ice is influenced by factors such as temperature, pressure, and the presence of other fluids (such as water).

### 1.3 Description cas d'usage

1. Utilisation des SLM pour la recherche thématique simple en français (en cours, Zhanna) Malgré la disponibilité et l'attractivité des « grands » modèles langages comme GPT et Mixtral, l'utilisation des petits modèles classiques est parfois plus avantageuse, surtout quand les ressources techniques ou l'accès aux données sont restreints.

C'est vrai dans le cas d'utilisation d'un SLM basé sur un modèle devenu classique, BERT qui donne la naissance à milliers de modèles spécialisés comme CamemBERT un modèle en français ou encore sBERT ou sentenceTransformers permettant un entraînement spécialisé pour une recherche

sémantique. \*\*ici plus d'information sur les avantages des SLM (données, environnement, spécialisation, travail en local, technique) Nous considérons un exemple d'utilisation de CamemBERT-base et un exemple de sBERT :

2. camembert-bio-base avec ses 111M de paramètres, pour une recherche thématique dans des textes scientifiques biomédicaux. Nous utiliserons les transformers de HuggingFace

```
from transformers import
AutoTokenizer, AutoModelForMaskedLM
biotokenizer = AutoTokenizer.from_pretrained("almanach/
camembert-bio-base")
biomodel = AutoModelForMaskedLM.from_pretrained("almana
camembert-bio-base")
```

2. all-MiniLM-L6-v2

```
import requests

api_url = f"https://api-inference.
huggingface.co/pipeline/feature-
extraction/{model_id}"
headers = {"Authorization": f"Bearer
{hf_token}"}
```

2. Classifier des accords d'entreprise

### **1.3.a Classifier des accords d'entreprise**

Les accords d'entreprise sont publiés sur LégiFrance. Ces accords peuvent concerner plusieurs thématiques (télétravail, compte épargne temps, droit à la déconnexion). Ces thématiques sont déclarés par les entreprises et sont éventuellement corrigées par la Direction Générale du Travail. Le besoin est alors de détecter automatiquement les thématiques à la lecture de l'accord. Un jeu de données est disponible à l'adresse suivante : `accords_publics_xx_to_2022_themes_et_texte.parquet`