**Títol del Treball**  
Anàlisi i optimització d'embeddings en LLMs: Comparació entre models genèrics i ajustats

**Noms de l’alumne i tutor**   
Martí Mussons Marin  
Antonio Lozano Bagen

**Objectius i plantejament del Treball**

L'avenç dels sistemes d'intel·ligència artificial ha portat a una millora significativa en la manera com els models generen i recuperen informació. Els Large Language Models (LLMs), com GPT-4 o BERT, utilitzen embeddings per representar el llenguatge en un espai matemàtic i, d’aquesta manera, capturar relacions semàntiques entre paraules i frases. Tot i això, els embeddings genèrics no sempre són la millor opció per a aplicacions específiques.

Aquest treball té com a objectiu analitzar el procés de selecció d’informació en sistemes d’IA, centrant-se en l’impacte dels embeddings en la qualitat de les respostes. En concret, s'explorarà el comportament d’un model d’embeddings entrenat en un corpus documental d’un domini específic i es compararà amb un model genèric per determinar si la personalització millora la recuperació d’informació en aquest context. Per fer-ho, s’utilitzaran tècniques com el fine-tuning i el contrastive learning per ajustar el model a les dades específiques del domini.

Els aspectes principals que es tractaran inclouen:

* Revisió dels fonaments teòrics dels embeddings, transformers i mecanismes d’atenció (*attention mechanisms*).
* Construcció, preprocessament o importació d’un corpus documental específic.
* Entrenament i *fine-tuning* d’un model d’embeddings en aquest corpus.
* Comparació entre el model ajustat i un model genèric mitjançant mètriques d’avaluació adequades.
* Anàlisi de resultats i discussió sobre les millores o limitacions observades.

A més, es s’utilitzaran eines com Hugging Face i Nomic AI, que permeten experimentar amb models d’embeddings de manera eficient. Els experiments realitzats permetran obtenir conclusions sobre la capacitat dels embeddings especialitzats per millorar la selecció d’informació en sistemes basats en IA, optimitzant-ne la precisió i rellevància en aplicacions concretes.

**Bibliografia**

**Articles i recursos web**

1. *Vaswani et al. (2017) - "Attention is All You Need"*  
   Introducció de l'arquitectura Transformer, basada exclusivament en mecanismes d'atenció i sense la necessitat de recurrència ni convolucions.  
   [PDF](https://arxiv.org/pdf/1706.03762.pdf)
2. *Ghodsi (2023) - "Attention Mechanism, Transformers, BERT, and GPT: Tutorial and Survey"*Explicació detallada i tutorial sobre els mecanismes d’atenció, transformers i models com BERT i GPT.  
   [PDF](https://hal.science/hal-04637647/document)
3. *Abnar & Zuidema (2020) - "Quantifying Attention Flow in Transformers"*  
   Proposta de mètodes per quantificar i analitzar el flux d’atenció en els transformers per entendre com distribueixen la informació.  
   [PDF](https://arxiv.org/pdf/2005.00928.pdf)
4. *Yeh et al. (2023) - "AttentionViz: A Global View of Transformer Attention"*Presentació d’una tècnica de visualització per analitzar patrons globals en mecanismes d'autoatenció dels transformers.  
   [PDF](https://arxiv.org/pdf/2305.03210.pdf)
5. *Abdaoui & Dutta (2023) - "Attention over pre-trained Sentence Embeddings for Long Document Classification"*Exploració de com els embeddings preentrenats i les capes d’atenció milloren la classificació de documents llargs.  
   [PDF](https://arxiv.org/pdf/2307.09084.pdf)
6. *IBM* - Explicació sobre els mecanismes d'atenció en IA.  
   [Enllaç](https://www.ibm.com/think/topics/attention-mechanism)
7. *Medium* - *Fine-Tuned Embedding with Contrastive Learning*  
   [Enllaç](https://medium.com/@roytalman_7701/fine-tuned-embedding-with-contrastive-learning-71da7b57164d)
8. *Dagshub* - Guia per entrenar un model d'embeddings personalitzat.  
   [Enllaç](https://dagshub.com/blog/how-to-train-a-custom-llm-embedding-model/)
9. *Nomic AI* - Investigació i eines per embeddings i models de llenguatge.  
   [Enllaç](https://www.nomic.ai/research)
10. *Hugging Face* - Plataforma per a models de NLP, embeddings i transformers.  
    [Enllaç](https://huggingface.co/)