
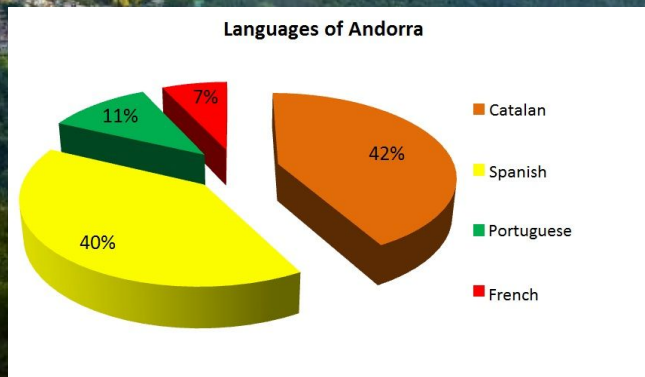




Catalan Customer Service

Presented by Eric Risco





ANDORRA

POPULATION: 77,142
LAND AREA: 470 KM²
DENSITY: 164 P/KM²
CAPITAL: ANDORRA LA VELLA
CURRENCY: EURO
OFFICIAL LANGUAGES: CATALAN



Catalan?

- Spoken in parts of Spain and France, but holds a unique status in Andorra.
- In Andorra, Catalan is central to national identity and is used in all official domains.
- Many LLMs struggle with Catalan due to its minority status. They often make spelling mistakes or mix it with other languages, hindering effective communication.

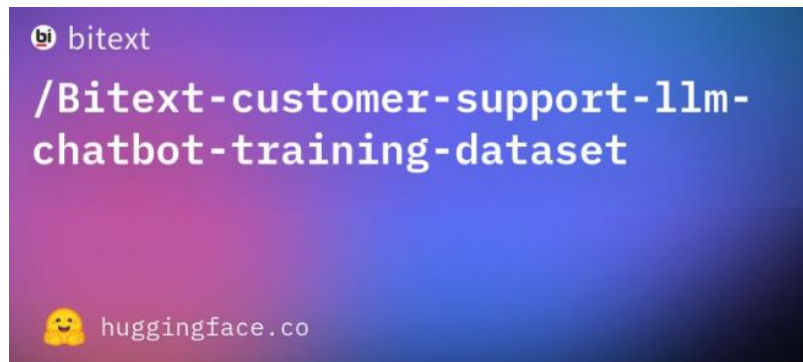


Customer service

- Our project aims to develop a LLM specifically trained to excel in responding to customer service inquiries.
- Our goal is to empower businesses and organizations in Andorra to communicate effectively with their customers in their own language, enhancing overall customer satisfaction.
- Beyond improving customer service, this project contributes to the preservation and promotion of the Catalan language, reinforcing its importance in the digital age.

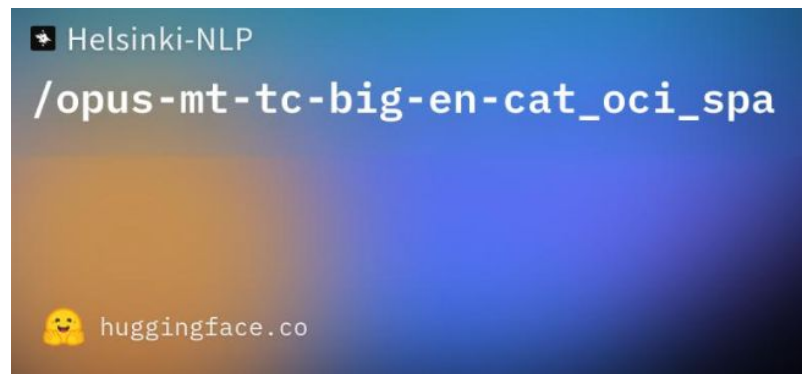
The dataset

- A dataset for training Large Language Models (LLMs) for Intent Detection in Customer Service.
- 26,872 question/answer pairs with 30 entity types and 12 language generation tags.
- Verticals include Automotive, Banking, Healthcare, and more.



The dataset curation

- This model is a neural machine translation (NMT) model designed for translating English (en) to Catalan, Occitan, and Spanish (cat+oci+spa).
- Used to translate the Bitext/Bitext-customer-support-llm-chatbot-training-dataset to Catalan.



Projecte Aina

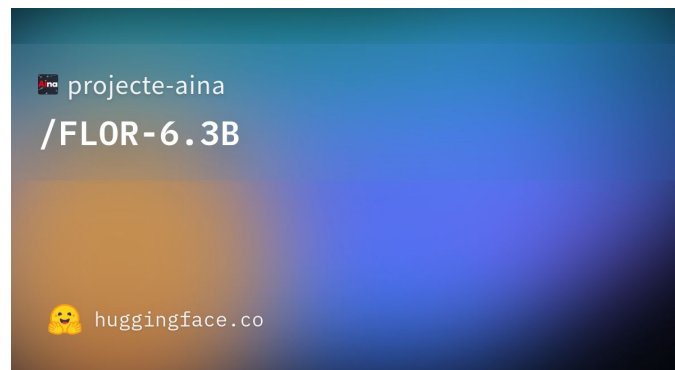
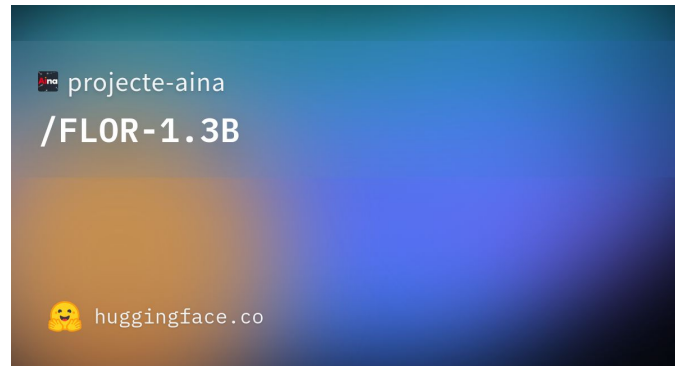


- Launched in 2020 with the Barcelona Supercomputing Center (BSC).
- AINA Project is a Catalan-focused artificial intelligence initiative.
- Aims to generate language corpora and computer models for the Catalan language.
- Facilitates the development of AI-based applications such as voice assistants, search engines, translators, and conversational agents in Catalan.
- Models developed within the AINA project are open-source and available for use by businesses and organizations.

The FLOR model

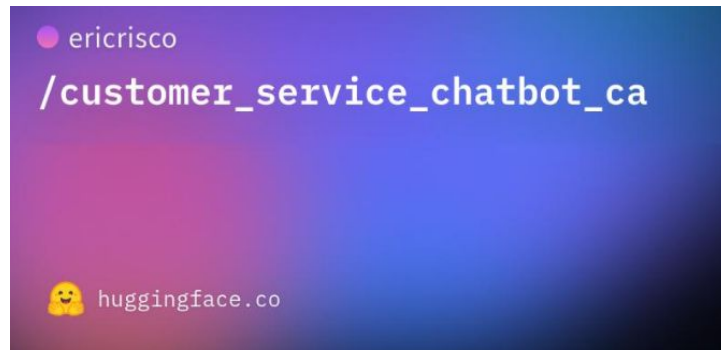


- AINA Project trained the FLOR-6.3B and the FLOR-1.3B
- Both supports Catalan, Spanish and English
- Derived from larger models (BLOOM) through language adaptation techniques
- Suitable for casual language modeling, text generation, and fine-tuning for specific applications



Fine-tuning FLOR-1.3B

- Fine-tuned from *projecte-aina/FLOR-1.3B* for a Catalan customer service chatbot.
- Uses the translated dataset version of bitext4 customer service
- Aims to provide contextually relevant responses in Catalan.





 **BONUS TRACK!**



FLOR-1.3B-customerservice + RAG Langchain

- Using the fine-tuned model
FLOR-1.3B-customerservice
- Langchain is a framework designed to enhance the capabilities of language models by integrating them with external knowledge sources
- Simulating a company document loading to get LLM external information



Notebooks!



Future work

- Real-world implementation of the model in Andorran and Catalan-based companies.
- Dataset Enhancement. Better translation and more use cases!
- Training optimization, exploring advanced techniques to improve the model's efficiency.
- Promotion and Adaptation of the model to different datasets



**THANK YOU
LLME Cohort 2!!**

