

Plan: Erasmus Mundus Master in Language and Communication Technologies (LTC)

Subject: Practicum (Internship i.e. prácticas obligatorias)

DETAILS OF THE COMPANY

Nombre de la empresa:	HiTZ Basque Center for Language Technology
Persona de contacto	
Email de contacto	emlct.internship@ehu.es
Teléfono de contacto	
País	
Provincia	Gipuzkoa
Localidad	Donostia

CONTACT DETAILS OF THE TUTOR: the supervisor within the university

Given name	Gorka
Family name	Azkune
Email	gorka.azcune@ehu.eus

DETAILS OF THE INTERNSHIP

Title	Analysing and improving current Large Language Models
Goal	<ul style="list-style-type: none">● Collaborate with the CohereAI researcher Jon Ander Campos.● Analyse different types of LLMs, such as LLMs trained with Reward Models (RM), and Retrieval Augmented Language Models (RAG).● Understand the implications of using those paradigms of learning for language generation.● Propose new solutions to mitigate detected errors or missbehaviours.
Tasks	<ul style="list-style-type: none">● Task 1. Analyse the state of the art on RMs.● Task 2. Analyse the state of the art on RAGs.● Task 3. Identify important resources to work with RMs and RAGs.● Task 4. Design analysis methodologies for Rms and RAGs● Task 5. Describe and document observed limitations and missbehaviours.● Task 6. Design, implement and experiment possible

	improvements.
Learning outcomes	By the end of this internship, the students <ul style="list-style-type: none"> ● will learn pros and cons of RMs and RAGs. ● will be able to work with RMs and RAGs.
Materials /Resources	Access to a server with GPUs.
Starting date:	01/12/2023
End date:	June (due date for transcript of records in GAUR)
Timetable:	Flexible
Number of hours (10ECTS):	250h
Language	The internship will be developed in English. No other language is required.
Financial support	0€
Intellectual Property %	<ul style="list-style-type: none"> ● The intellectual property will be 50% shared with HiTZ.
Specific requirements (background of the candidate)	Background of the candidate: <ul style="list-style-type: none"> ● Preferably EMLCT Y2 student ● Undergraduate degree: Preferably Computer Scientist ● Specific background: Deep Learning, LLMs ● Good programming skills in Python