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	Aitor
Family name	Soroa
Email	a.soroa@ehu.eus

DETAILS OF THE INTERNSHIP

Title	Data-to-text models for meteorological reports
Goal	The goal of this project is to use multilingual generative language models to produce meteorological reports given numerical data represented in tables. The student will use data-to-text techniques to produce meteorological reports from numerical input in various geographical areas and time frames. The work will be carried out within the DeepR3 project.
Tasks	The aforementioned goals require fulfilling the following tasks: ● Task 1. Study and select generative language models that are already pre-trained in the languages of interest. ● Task 2. Select available datasets with meteorological data and textual reports in various languages. ● Task 3: Devise the best strategy to leverage the information in various languages.

	● Task 4. Evaluate the models, including regional biases, as well as hallucinations that the model can perform.
Learning outcomes	By the end of this internship, the students will ● Learn to use and adapt generative language models to new tasks. ● Use machine translation systems to translate datasets. ● Learn data-to-text methods to generate textual descriptions given numerical data. ● Learn to evaluate generative language models.
Materials /Resources	The student will use resources from the HiTZ center, including computing power (GPU), etc.
Starting date:	
End date:	June (due date for transcript of records in GAUR)
Timetable:	Flexible timetable and work schedule.
Number of hours (10ECTS):	250h
Language	The internship will be developed in English. Speaking other languages (specially Spanish) is recommended but not necessary.
Financial support	0€
Intellectual Property %	● The work done through the project will be publicly accessible.
Specific requirements (background of the candidate)	Background of the candidate: ● Preferably EMLCT Y2 student ● Preferable Engineer or Computer Scientist ● Background on Deep Learning and LLMs ● Good programming skills in Python