

The background of the slide is a teal-colored image showing the silhouettes of several people sitting around a table in a meeting room. Overlaid on this are several white, abstract geometric shapes consisting of interconnected lines and dots, resembling network graphs or data structures. The overall aesthetic is modern and tech-oriented.

# sherpa.ai

## Federated Natural Language Processing Tasks

AI Research & Development

*Master Thesis Project*

## Project Description

Pre-trained neural language models such as BERT or GPT-3 have become the state-of-the-art end-to-end models for natural language understanding. These pre-trained language models allow one to devote the efforts to fine tune these models for specific tasks.

Federated Learning is a new paradigm in machine learning aiming at enabling to learn from private, probably distributed, data providers. Federated Learning has been successfully applied to deep learning methods.

**The project consists of exploring deep learning end-to-end language models, such as BERT or GPT-3, and their fine tuning for diverse tasks using private data in a federated learning scenario.**

## Goals

1. Understand the problems raised when fine-tuning BERT models using private data.
2. Study and reproduce state-of-the-art approaches.
3. Identify current difficulties and communicate conclusions.
4. Develop the experimental setting using Sherpa's Data Privacy API.

## Materials

- Sherpa's Data Privacy API.
- Computer.

## Working Plan & Expected Results

1. Study a few selected relevant papers and reproduce some of their experiments.
2. Agree with Sherpa in a particular experiment of interest for Sherpa and the student.
3. Report.

## Academic and Industrial Mentoring

- Dr. Miguel A. Veganzones (Sherpa AI Director)
- Prof. Eneko Agirre (IXA Team, EHU-UPV)

## Candidate Profile

Basic knowledge of:

- Python
- Natural Language Processing

Interest on:

- End-to-end neural language models
- Data Privacy

## Benefits and Practical Information

- Funding: 2600€
- Duration: 3-6 Months
- Location: Aula SHERPA, Fac. Informática San Sebastián

Sherpa Europe, S.L. (Sherpa) accepts no liability for the content of this document, or for the consequences of any actions taken on the basis of the information provided. This document is intended to provide preliminary guidance in anticipation of further discussion and has not been prepared with the level of due diligence and analysis that would be needed to constitute a commitment of Sherpa. Anyone who receives this document are cautioned to consider that its contents are unaudited and that it may contain inaccurate, incomplete or summarized information, which may be change without notice. This document is confidential and its contents may not be totally or partially disclosed or reproduced, without the prior written consent of Sherpa. By allowing you access to the aforementioned document we shall have no liability, duty or obligation of any kind to you.