



# sherpa.ai

## Artificial Journalism

AI Research & Development

*Master Thesis Project*

HAP-LAP 2021 – UPV-EHU

## 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.

In the preface of “New powers, new responsibilities. A global survey of journalism and artificial intelligence”, Prof. Charlie Beckett comments on Artificial Journalism: **“Machines might soon be able to do much routine journalism labour. But the reality and the potential of artificial intelligence (AI), machine learning, and data processing is to give journalists new powers of discovery, creation and connection.** The hope is that journalists will be algorithmically turbo-charged, capable of using their human skills in new and more effective ways. AI could also transform newsrooms from linear production lines into networked information and engagement hubs that give journalists the structures to take the news industry forward into the data-driven age.”

**The project consists of exploring deep learning end-to-end language models, such as BERT or GPT-3, in the context of artificial journalism, particularly in automated article creation.**

## Goals

1. Study and reproduce state-of-the-art approaches in automated article creation.
2. Design and perform experiments using real datasets.
3. Identify current difficulties and communicate conclusions.

## Materials

- 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
- Journalism

## 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.