

Julien PLU

## Research and development engineer

10 years of experience in various projects related to artificial intelligence on Natural Language Processing, in particular on the development of models and their deployment in production. Autonomous on my position while enjoying teamwork with on-site or remote collaborators.

### Work experience

#### Research and Development Engineer in AI

**Buster.Ai** Paris / Since October 2022

My role at Buster is to bring AI from research to production in a end-to-end manner. This involves as much as conducting research as making this research available to improve our product.

I'm mostly working with LLMs for doing search and fact checking, also optimize those LLMs in order to have the best ratio low size / low latency in a constrained production environment. The full stack is in Python and the cloud provider is GCP.

#### Senior Machine Learning Engineer

**Beam App** Paris / From May 2021 to October 2022

My role at Beam was to work on two different projects:

- 1) The first project was to generate paragraphs when a user was taking some notes into our Web browser, this generation was taking into account the whole context of the notes, including the videos in it, no LLMs were used to do that only simple text generation models.
- 2) The second project was to cluster opened tabs in the Web browser by topics and to give them a name.

The models of both projects needed to be run on-premise directly on the user machine and as fast as possible. To do that a lot of model optimization were used and the integration was made by creating C++ libs that were used directly in Swift as the Web browser was for MacOS/iOS only.

#### Machine Learning Engineer

**Hugging Face** Paris / From September 2020 to March 2021

My role was to bring TensorFlow into the Hugging Face transformers and datasets frameworks by implementing more and more models / tools. Also I was helping the NLP community to better support these Tensorflow models and features.

#### Research and Development Engineer in AI

**Leboncoin** Paris / From April 2018 to August 2020

My role at Leboncoin was to investigate new IA approaches in order to improve the experience of our users.

I worked on multiples projects but the biggest project was to train and deploy a language model for French in order to classify ads titles. This brought a lot of challenges such as:

- curating a huge set of data (~100M examples), a lot of cleaning was needed.
- not much of GPUs was available, then I had to deal with a very small possibility of compute to train a language model.
- I had to implement and integrate this model into the Leboncoin stack in Go.
- the infra was not made to welcome such new service and then, with the SRE team we had to define an entire pipeline only for ML services.



- ✉ plu.julien@gmail.com
- ☎ +33 6 70 14 19 68
- 📅 Date of birth 19/06/1987
- 🏠 Bagnole (93170)
- 🔗 <https://jplu.github.io>
- 🇫🇷 French
- 💻 Open to remote work
- 📄 Driving licence (B)
- 🚴 Cyclist
- 📍 Île-de-France
- 👤 Engaged

### Languages

English

Japanese

### Skills

Machine/Deep Learning

Natural Language Processing

Semantic Web

Computer Science

Python

Java

Go

Miscellaneous

Kubernetes

Google Cloud Platform

Interests

Video Game

Scuba Diving

Guitare

Football

## Education

- **PhD**

Eurecom Sophia Antipolis / From July 2014 to December 2017

My PhD topic is to extract lexical units (entities) from textual documents to disambiguate them against Web resources contained in a knowledge base and in an adaptive manner. This adaptivity is on the language of the processed texts, the used knowledge base and the textual content that can be either formal text (journalistic style) or informal text (microposts).

- **Computer Science Master, Specialized in AI**


Université Montpellier 2 Montpellier / From September 2010 to September 2012

- **Computer Science Bachelor**

Université Montpellier 2 Montpellier / From September 2007 to June 2010

## Social networks

 @julienplu

 @jplu

 @julienplu