



# UMBERTO SELVA

Linguistics PhD, NLP / AI Engineer  
Python, Deep Learning, Data Science

✉ [u.o.selva@gmail.com](mailto:u.o.selva@gmail.com)

📞 +39 333 5218014

🌐 [linkedin.com/in/umberto-selva](https://linkedin.com/in/umberto-selva)

🌐 [github.com/umbertoselva](https://github.com/umbertoselva)

## PROFILE

I am a former post-doctoral researcher and lecturer with a Phd in Linguistics from Leiden University in the Netherlands. After a career in academia I trained and freelanced as an NLP specialist acquiring a broad expertise across Machine Learning and Data Science with a solid Software Development foundation. I am currently working as an AI Engineer and I specialize in designing LLM-powered solutions and prompt engineering for a web platform in the pharmaceutical sector. I continue to offer freelancing services.

## EXPERIENCE

### ● AI Engineer

[Pharmazie.com](#)

[2023–Today](#)

Pharmazie.com is a web platform that gathers data from over 25 databases to provide reliable information on pharmaceutical products to healthcare professionals.

My task is to develop all Generative AI and NLP projects, such as LLM-powered conversational assistants (chatbots), APIs, web applications, and pipelines, designing all steps from text and table data ingestion, vectorization / embedding, retrieval augmented generation (RAG), and prompt engineering, all the way to server management and production deployment on AWS.

**Skills:** Python, LLMs, LangChain, OpenAI, Huggingface, Prompt Engineering, Chatbot Design, Retrieval Augmented Generation (RAG), Back-end Web Development, AWS Cloud

### ● NLP / AI / ML / Data Science Consultant

[Freelance](#)

[2020–Today](#)

Between 2020 and 2022 I transitioned from academia to the private sector, training in Software Development with a focus on NLP, AI, Machine Learning and Data Science, working on various personal projects, as well as offering freelancing services.

**Skills:** Python, NLP, Machine Learning, Data Science, Statistics, LLMs, Prompt Engineering, Software Development, Back-end Web/API Development, SQL, bash/zsh, Linux

### ● Academic Researcher and Lecturer (Linguistics)

[LUCL / IIAS, Leiden, NL](#)

[2014–2020](#)

Between 2014 and 2020, I was research assistant, then PhD researcher, post-doctoral researcher and lecturer at the Leiden University Centre for Linguistics (2014-2019) and at the International Institute for Asian Studies (2019-2020) in Leiden, Netherlands.

**Skills:** Linguistics (General, Historical, Computational), Academic Research, Academic Writing, Teaching, Research Project Management



# PORTFOLIO

github/umbertoselva

## GPT-3.5 Museum FAQ Chatbot (2023) [repo]

A chatbot built with Langchain, OpenAI GPT-3.5 and a Streamlit UI that can answer questions based on the Stedelijk Museum Amsterdam's FAQ list.

**Skills:** Python, Langchain, OpenAI, Streamlit, Retrieval Augmented Generation, Chatbot design, Prompt Engineering

## GPT-3 Unbiased News App (2023) [repo]

A Streamlit app built with OpenAI GPT-3 that rewrites news articles by removing bias and highlights logical fallacies.

**Skills:** Python, OpenAI, Streamlit, Prompt Engineering

## Fine-tuning BERT with TensorFlow for NER-based Sentiment Analysis (2022) [repo]

A Data Science project in which I combine NER-extraction with spaCy and fine-tuning a BERT model with TensorFlow to perform sentiment analysis on a real-world movie review dataset extracted from the popular "I Just Watched" subreddit with the goal of finding out which actors are most popular among the subreddit users (spoiler: it was Tom Hanks!).

**Skills:** Python, TensorFlow, Keras, spaCy, Pandas, Huggingface

## Exploratory Data Analysis and House Price Prediction (2022) [repo]

A step-by-step example of how to apply Data Science to a classic regression task. I conduct exploratory data analysis, data preprocessing, feature selection, model design, training, evaluation and model explainability.

**Skills:** Python, Pandas, NumPy, Sci-kit Learn, TensorFlow, Keras, Scipy.stats, Statsmodels, Matplotlib, Seaborn, Lime

## The Sanskrit Meterscan (2020) [repo]

A digital humanities web app, built with Python's back-end framework Flask, that automates the metrical scansion of Sanskrit and Vedic texts. Used by academic researchers in the fields of Linguistics and Philology

**Skills:** Python, Flask, HTML, CSS, JavaScript, Back-end Web Development

# CERTIFICATIONS

## Natural Language Processing Specialization

Deeplearning.ai – Jan 2022 (four-month program)

- In-depth theoretical intro to probabilistic, sequence and deep learning models, from N-Grams to vector space, word embeddings (Word2Vec, etc.), Neural Networks, RNNs, LSTM, Attention, Transformer models, etc.

## NLP with Transformers in Python

Udemy (by J. Briggs) – Jul 2022

- Topics: Transformer architecture, Attention, pre-training, fine-tuning, applied to tasks such as text classification, sentiment analysis, NER, reader-retriever QA, preprocessing, metrics, etc., with Huggingface, TensorFlow, SpaCy, FAISS, etc.

## Machine Learning

Stanford Online – Aug 2020

- Andrew Ng's classic introduction to ML

## Deep Learning Specialization

Deeplearning.ai – Nov 2022

- Courses: (1) Neural Networks and Deep Learning, (2) Hyperparameter Tuning, Regularization and Optimization, (3) Structuring Machine Learning Projects

## Statistics for Data Science with Python

IBM – Oct 2022

- Descriptive statistics, data visualization, probability distributions, hypothesis testing, regression analysis, etc.

## Practical Data Science on the AWS Cloud

AWS & Deeplearning.ai – Dec 2022

- Build, train and deploy ML pipelines using BERT with Amazon SageMaker.

# SKILLS

- Coding languages: Python, SQL(MySQL)
- AI / NLP (Generative AI, Prompt Engineering, Retrieval Augmented Generation (RAG), OpenAI, LangChain, Huggingface Transformers, LlamaIndex, spaCy, NLTK, BERT, GPT-x, LLMs)
- ML / DL (Scikit-learn, Tensorflow, Keras)
- MLOps (AWS Cloud, Docker)
- Data Science (Pandas, NumPy, Matplotlib, Seaborn)
- Web App/API Dev (Flask, FastAPI, Gradio, Streamlit)
- Other (git, Linux, bash/zsh, Jupyter Notebook)
- Languages: native level English, Italian