



# Andrés Cabero Busto

Stuttgart, Germany

+4917684048732 | cabero96@protonmail.com | cbr9 | andres-cabero-busto

## Skills

<b>Data Science, AI &amp; ML</b>	Pandas, Plotly, Matplotlib, SLURM, W&B, PyTorch, PyTorch Lightning, HuggingFace, Captum
<b>NLP</b>	LangChain, SpaCy, NLTK
<b>Data Analysis</b>	PowerBI, Tableau, Excel, SQL
<b>Programming</b>	Rust, Python, JavaScript, TypeScript, HTML, CSS, LaTeX
<b>Languages</b>	Spanish, English, Italian, German (B1)
<b>Other</b>	Docker, Nushell, Linux, NixOS, Nix, Bash, CLI, Git

## Experience

### Solventum

Munich, Germany

#### SPEECH DEVELOPER

Jan. 2024 - Present

- Management and improvement of NLU pipeline for a commercial medical ASR software package
- Assistance with debugging of E2E ASR models (Kaldi-based)
- Created high-performance text processing pipelines using Rust

### Sony Europe B.V. (AI, Speech & Sound Group)

Stuttgart, Germany

#### MASTER THESIS

Apr. 2023 - Oct. 2023

- Topic: Multi-domain adaptation of self-supervised speech transformers via continual curriculum learning
- Performed domain-adaptation on an open-source, self-supervised (SSL) speech recognition model.
- Applied regularization-based and replay-based continual learning methods to avoid incurring into catastrophic forgetting on the old Librispeech ASR task.
- Implemented several research papers from scratch (Elastic Weight Consolidation; Curriculum Learning by Dynamic Instance Hardness)

### Sony Europe B.V. (AI, Speech & Sound Group)

Stuttgart, Germany

#### WORKING STUDENT

Jul. 2022 - Mar. 2023

- Developed a natural language grammar based on weighted finite-state transducers (WFSTs) and the Thrax framework, similar to NVIDIA NeMo grammars, that enabled a 10% WER improvement for the pre-existing Italian ASR system.
- Time-aligned and transcribed speech utterances using Label Studio for use in text-to-speech (TTS) model training.

### Institut für Maschinelle Sprachverarbeitung (IMS)

Stuttgart, Germany

#### RESEARCH ASSISTANT

Jul. 2022 - Dec. 2022

- Developed a toolkit for natural language processing research in lexical semantic change detection (LSCD)
- Set up the project using Hydra for easy experimentation and customizability
- Data validation done using Pydantic

### Institut für Maschinelle Sprachverarbeitung (IMS)

Stuttgart, Germany

#### LINUX SYSTEM ADMINISTRATOR

Jul. 2022 - Mar. 2022

- Advising students and employees at IMS' IT management office
- Management and installation of personal computers and GPU servers
- Management of users, network & websites
- Technical support and assistance for professors and students

### Institut für Maschinelle Sprachverarbeitung (IMS)

Stuttgart, Germany

#### DATA ANNOTATOR

Nov. 2021 - Jan. 2022

- Data annotator for the first Shared Task of Lexical Semantic Change Discovery in Spanish (<https://fdzr.github.io/lscdiscovery/>).
- The process involved annotating the degree of difference between uses of the same word based on the surrounding context, using a Likert scale.
- Identified practical difficulties in the annotation platform, and developed a JavaScript solution to enhance the efficiency for the project and the team.

# Projects

---

## Model Compression using Explainable AI (XAI)

Stuttgart, Germany

INSTITUT FÜR MASCHINELLE SPRACHVERARBEITUNG (UNIVERSITY OF STUTTGART)

- Compressed a Bi-LSTM trained for a speech emotion recognition task.
- Using XAI methods, we realized that more than half of the input features were completely irrelevant, thus allowing us to reduce the number of parameters of the network.
- Because we were able to reduce the parameters in this way, this also allowed us to experiment with other hyperparameters more efficiently.
- The final model preserved the exact same accuracy of the original model, but with more than half of its parameters.

## Bachelor Thesis Project

Cádiz, Spain

UNIVERSITY OF CÁDIZ

2020

- Topic: Automatic Keyword Extraction via Latent Semantic Analysis (LSA)
- Automatic extraction of keywords from BBC news articles using LSA embeddings.

# Education

---

## M.Sc. Computational Linguistics & Natural Language Processing

Apr. 2021 - Oct. 2023

INSTITUT FÜR MASCHINELLE SPRACHVERARBEITUNG (UNIVERSITY OF STUTTGART)

Stuttgart, Germany

## B.A. Linguistics

Sept. 2016 - Dec. 2020

UNIVERSITY OF CÁDIZ

Cádiz, Spain

# Certifications

---

## MySQL Bootcamp

May 2025

[HTTPS://WWW.UDEMY.COM/CERTIFICATE/UC-B2DA505C-5A37-4FF3-AE4A-B00EF7FE41E8/](https://www.udemy.com/certificate/UC-B2DA505C-5A37-4FF3-AE4A-B00EF7FE41E8/)

Online

## PyTorch For Deep Learning

Apr. 2022

[HTTPS://WWW.UDEMY.COM/CERTIFICATE/UC-C8142260-A945-4F41-91F5-D213671B8493/](https://www.udemy.com/certificate/UC-C8142260-A945-4F41-91F5-D213671B8493/)

Online

## Mathematics For Machine Learning

Mar. 2022

[HTTPS://WWW.COURSERA.ORG/ACCOUNT/ACCOMPLISHMENTS/SPECIALIZATION/CERTIFICATE/DJPXWT9X9KVS](https://www.coursera.org/account/accomplishments/specialization/certificate/DJPXWT9X9KVS)

Online

## Neural Networks and Deep Learning

Dec. 2021

[HTTPS://WWW.COURSERA.ORG/ACCOUNT/ACCOMPLISHMENTS/CERTIFICATE/JXLQNF3H9S](https://www.coursera.org/account/accomplishments/certificate/JXLQNF3H9S)

Online

## Linear Algebra

Mar. 2020

[HTTPS://WWW.UDEMY.COM/CERTIFICATE/UC-0D00A3F9-2A0C-4881-A50E-4341FA30B055/](https://www.udemy.com/certificate/UC-0D00A3F9-2A0C-4881-A50E-4341FA30B055/)

Online

## Probability & Statistics

Mar. 2020

[HTTPS://WWW.UDEMY.COM/CERTIFICATE/UC-BHG909SJ/](https://www.udemy.com/certificate/UC-BHG909SJ/)

Online