Caterina Bonan

PhD in Linguistics | Junior NLP Engineer

OrcID 🐙 CaterinaBi 💼 LinkedIn ✉ caterina.bonan@outlook.com 📞 07983818663

SELECTED EXPERIENCE

AICORE | DATA AND AI TRAINEE
May 2022 – Present | London, UK
Training in AI & Data Engineering using industry-standard tools. Working on multiple industry projects throughout.
- → Skills: Software engineering (Git & GitHub, advanced Python, algorithms & data structures); Data engineering (SQL, data lakes, data warehousing, web scraping); Cloud Engineering (cloud computing, designing/building APIs, Docker, Apache Airflow, AWS Serverless Stack).

UNIVERSITY OF CAMBRIDGE | POSTDOCTORAL RESEARCHER
September 2019 – Current | Cambridge, UK
Postdoctoral research on comparative Romance syntax. Worked on 30+ standard and non-standard varieties throughout. Published 8 peer-reviewed research articles, 2 edited volumes and 1 research monograph. Awarded a SNSF 'PostDoc.mobility' scholarship twice.
- → Tools: Excel for data analysis and visualisation, Python, MatPlotLib, HF Transformers, SQL, LaTeX, VS code, Microsoft Word and Power Point.
- → Links to scientific writing samples: **Cadernos**, **IOS Press**, **Glossa**, **Isogloss**.

UNIVERSITÉ DE GENÈVE | RESEARCH ASSISTANT
September 2015 – August 2019 | Geneva, CH
Fully-funded doctoral research as part of the SNSF-funded 'Wh-in situ in French interrogatives: Prosody and Syntax' research group. Worked on standard and non-standard French, and 50+ northern Italian varieties. Published 3 research articles, 1 co-edited volume, and a PhD dissertation.
- → Tools: Excel for data analysis and visualisation, Python, Praat, LaTeX, Microsoft Word and Power Point.

ELSE: CONSEIL, CRÉATION, ÉDITION | TRANSLATOR (IT>FR)
March 2014 – May 2014 | Paris, FR
Translation of book 'Grands écrivains: Les auteurs célèbres vus par de grands photographes' (Éditions du Chêne, 522 pp.; ISBN 978-2-8123-1085-0).

EXPERIENCE THROUGH PROJECTS

USING NLP TO PREDICT SELF-HARM (OMDENA LONDON) | PYTHON
Repo: udacity-ai-programming
- → Cooperative project intended to asses whether the topic of eating disorders and depression can be used to predict self-harm in London's 10-24yo females.
- → Project team envisage a set of time-series sentiment scores by topic mapped to Central and Northwest London ambulance call-out data for self-harm/injury.

WEB SCRAPING PIPELINE | PYTHON, BEAUTIFULSOUP, SELENIUM, AWS, PROMETHEUS, GRAFANA.
Repo: aicore-web-scraping-pipeline
- → Developed a module that scrapes data to upsert information to a database;
- → Currently pushing an image of the containerised application in a CI/CD pipeline that performs unit testing and deploys the image to an EC2 instance.
- → Will monitor the container using Prometheus and create dashboards to visualise metrics in Grafana.

IMAGE CLASSIFIER FOR FLOWER RECOGNITION | PYTHON, TENSORFLOW.
Repo: udacity-image-classifier
- → Python image classification application that runs from the command line.
- → The application trains a deep learning model on the 102 Category Flower Dataset, and uses it to classify images of flowers.

COMPUTER VISION APPLICATION | PYTHON, TEACHABLE MACHINE, OPENCV
Repo: aicore-computer-vision
- → Trained a computer vision model and used Tensorflow to detect whether Rock, Paper or Scissors is shown to the camera in real-time and with a high accuracy
- → Used the OpenCV library to access the webcam and play Rock Paper Scissors with the computer using the image from the camera

SKILLS

PROGRAMMING

Intermediate:
Python (>2y) • LaTeX (>5y)

Familiar:
Java (>1y) • HTML (<1y)

LIBRARIES/FRAMEWORKS

OpenCV • TensorFlow • Selenium
PyTorch • pandas • MatPlotLib
NLTK • SpeechRecognition •
PyDub • spaCy • gensim

TOOLS/PLATFORMS

Git • GitHub • Docker • SQL

LINGUISTIC RESEARCH

Syntax, Documentation, Data modelling, IPA, Praat.

EDUCATION

PHD IN LINGUISTICS
UNIVERSITÉ DE GENÈVE
December 2015 - May 2019

MA IN SCIENCE OF LANGUAGE
UNIVERSITÀ CA' FOSCARI VENEZIA
October 2011 - June 2013

BA IN SCIENCE OF LANGUAGE
UNIVERSITÀ CA' FOSCARI
September 2008 - October 2011

CERTIFICATES

UDACITY
March 2022 - October 2022
Programming in Python for AI.

DATACAMP
October 2022 - Present
NLP in Python Spoken | Spoken Language Processing in Python.

AICORE
May 2022 - September 2022
Python Programmer Certificate.

UNIVERSITÉ DE GENÈVE
December 2015 - December 2017
Specialisation in Linguistics.

CURIOSITIES

I'm a retired pro volleyball player, and since 2015 I have attracted research funds for more than 400.000chf (about £375.000).