Caterina Bonan

Theoretical Linguistics | NLP Engineering

🏠 OrcID   ⬡ CaterinaBi   in LinkedIn   ✉ caterina.bonan@outlook.com   📞 07983818663

SELECTED WORK EXPERIENCE

UNIVERSITY OF CAMBRIDGE | POSTDOCTORAL RESEARCHER IN THEORETICAL LINGUISTICS
September 2019 – Current | Cambridge, UK

Research in comparative syntax (data collection, visualization, modelling). Published 8 research articles, 2 edited volumes, 1 research monograph. Awarded a SNSF grant twice.

→ Tools: Excel, Python, MatPlotLib, Selenium, LaTeX, VS code, HF Transformers.
→ Links to scientific writing samples: **IOS Press**, **Datasets (pre-print)**, **Cadernos**, **Glossa**.

UNIVERSITÉ DE GENÈVE | RESEARCH ASSISTANT IN THEORETICAL LINGUISTICS
September 2015 – August 2019 | Geneva, CH

Worked on comparative Romance syntax (data collection, visualization, mmodelling). Published 3 research articles, 1 co-edited volume, and a PhD dissertation.

→ Tools: Excel, Python, Praat, LaTeX, Word and Power Point.

RELEVANT TRAININGS

AICORE | MACHINE LEARNING AND DATA ENGINEERING TRAINEE
May 2022 – Present | London, UK

Training in AI & Data using industry-standard tools. Working on industry projects throughout.

→ Skills & tools: Web scraping; Data cleaning & visualization; Machine learning (Regression, random forest, SVMs, PCA, t-SNE); Deep learning (Pytorch, neural networks, CNNs, RNNs, Autoencoders, GANs), NLP (LSTMs, transformers, attention mechanism, BERT, HuggingFace); Training in the cloud (AWS EC2); Deployment (Python Flask, Cron, API creation, AWS Lambda, Docker, ONNX, TensorFlow.JS, GCP AI services).

OPEN-SOURCE COLLABORATIONS

POLYGLOT @ ELEUTHERAI | NLP ENGINEER & LANGUAGE EXPERT.
Repo: eleutherAI-polyglot

→ Creation of multilingual models that show better non-English performance than mBERT, BLOOM, and XGLM.
→ Currently working towards the establishment of a Romance dataset (3T/language target).

OMDENA | NLP ENGINEER & CHAPTER LEAD.
Webpage: Global impact with Artificial intelligence

→ Worked on 4 NLP-powered 'AI for good' cooperative projects (two conversational AI assistants, and two classifiers to prevent online hate speech and child grooming);
→ Currently preparing the first challenge ('standalone chrome extension for fake news detection') for the Cambridge Local Chapter, which I lead.

ADDITIONAL AI PROJECT EXPERIENCE

FB MARKETPLACE RECOMMENDATION RANKING SYSTEM | PYTHON, PYTORCH, AWS, EKS.
Repo: aicore-recommendation-ranking-system

→ Currently training a model that works with images and text to make recommendations to potential buyers based on their demographic information;
→ Will host the app on EKS to orchestrate the containers that maintain/retrain the model.

WEB SCRAPING PIPELINE | PYTHON, SELENIUM, DOCKER, GITHUB ACTIONS.
Repo: aicore-web-scraping-pipeline

→ Developed a module that scrapes data to upsert information to a database;
→ Pushed an image of the containerised application in a CI/CD pipeline that performs unit testing and deploys the image to an EC2 instance.

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 • spaCy • seaborn • gensim

TOOLS/PLATFORMS

Git • GitHub • Docker

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

AiCore
September 2022 - November 2022
Software Engineer Certificate.

AiCore
May 2022 - September 2022
Python Programmer Certificate.

Université de Genève
September 2015 - December 2017
Certificat de spécialisation en linguistique.

LANGUAGES

Italian	native
French	near-native
English	proficient
Spanish	conversational

CURIOSITIES

I'm a retired pro volleyball player.