

# Bruno Gomes Coelho

Last updated December 2024

bruno.coelho@nyu.edu

brunogomescoelho.github.io

in/bruno-coelho-277519129/

## EDUCATION

---

- **New York University (NYU)** NY, USA  
*PhD in Computer Science focused on Machine Learning* September 2020 – Expected May 2025
- **New York University (NYU)** NY, USA  
*Master's in Computer Science* September. 2020 – May 2022
- **University of São Paulo (USP), ICMC** SP, Brasil  
*Bachelor's Degree in Computer Science; Emphasis in Data Science; GPA: 8.8/10.0* February 2016 – January 2020

## INDUSTRY EXPERIENCE

---

- **Kensho - A company of S&P Global** NY, USA  
*Returning Machine Learning Intern (NLP Team)* May 2024 - August 2023
  - **Project:** Research on LLM Guardrails against attacks and toxicity; Evaluation of LLMs for code generation.*Machine Learning Intern (NLP Team)* June 2023 - August 2023
  - **Project:** Zero-shot text classification through different approaches (NLI, semantic similarity, in-context learning), and LLM fine-tuning regimes (PEFT, LoRA, accelerate).
  - **Impact:** Increased performance of production model on zero-shot classification
- **Data Science MBA (University of São Paulo)** Remote  
*TA/Tutor* February 2020 - current
  - **Responsibilities:** Teaching, Preparing and revising materials, grading, interacting with students.
  - **Classes:** Neural Networks and Deep Learning, Machine Learning, Introduction and Programming for Data Science.
- **The AI Academy** Remote  
*Data Scientist Consultant* May 2020 - October 2020
  - **Continuous Development for DS:** Prototyped and evaluated new DS processes and tools.
  - **DevOps:** Integrated H2O Driverless AI into automated CD pipelines for one of Brazil's biggest bank.
- **Imaflora Origens** Brazil  
*Data Scientist* February 2020 - July 2020
  - **ETL:** Data wrangling Geo-spatial data from different sources for a interactive visualisation.
  - **Visualisation:** Worked alongside business owners to build a interactive platform for exploring data-driven insights.
- **University of São Paulo, ICMC** Brazil  
*Research Assistant - International scholarship by USAID* March 2017 - December 2019
  - **Processing:** Transformed raw audio data (mosquito's wingbeats) into structured data using time series analysis.
  - **Classifying:** Building pipelines and evaluating models to classify different insect species.
  - **Undergraduate thesis:** Applying CNNs to raw audio data for insect classification; Analysing covariate shift in the context of sensor generalization.
- **Getúlio Vargas Fundation (FGV) and Imaflora** Brazil  
*Data Scientist* August 2019 - December 2019
  - **Data Processing:** Integration of multiple government systems related to timber exploration data.
  - **Illegal Exploration:** Developing a django/D3 visualisation tool to facilitate the discovery of anomalous data.

- **Data Acquisition and Mining:** Acquired more than 1 million e-commerce products through crawlers, building a reliable back-end infrastructure; Extraction of useful features from the unstructured textual data of the products.
- **Machine Learning:** Building hierarchical models for product-category classification
- **Production:** Deployed the models as micro-services processing 10 million+ daily products.

## TECHNICAL SKILLS

---

**Languages:** Python, SQL, C, JavaScript, Java

**Topics:** NLP, Computer Vision, Deep Learning, MLOps, Big data, Internet Measurement

**DS/ML:** Pandas, Scikit-Learn, PyTorch, TensorFlow, huggingface, DVC, Seaborn, Plotly

**Others:** BeautifulSoup, Scrapy, Django, Git, Docker, Airflow, Driverless AI, GCP

## MAJOR PROJECTS

---

- **Winner of the Dotz NLP Competition** May 2020 - July 2020
  - **Competition:** Hierarchical classification of e-commerce products from their description.
  - **Solution:** Building a integrated pipeline using a mixture of portuguese Bert, Character CNNs, traditional sparse BOW model and hierarchical specific models.
- **Finalist of the Data Science Games 2018** May 2018 - September 2018
  - **Qualification Phase:** Analysed the bond market with data provided by BNP Paribas; Predicted customer interest in a bond using diverse data; Final result in the top 30 worldwide from 220 participating universities.
  - **Final:** Used e-commerce data to help predict customer interest in online shopping.
- **Co-founder of DATA - Extracurricular Data Science Group** September 2018 - January 2020
  - **Co-Founder:** Creating and running a extracurricular data science group recognised by the university.
  - **Teaching:** Weekly classes teaching basic Machine Learning concepts and frameworks to a class of 30 students.
- **Computer Science and Data Science Teaching/Volunteering**
  - **NYU's Introduction to CS:** Python programming general grading, Fall 2021.
  - **Invited DS Course Lead:** Heading the DS part of a 3 month course at a startup incubator, Summer 2019.
  - **Python for DS course:** Teaching industry members at the Universidade de São Paulo, Fall 2019.
  - **Mentoring:** High school student, helping with SATs and career questions, 2018.
  - **Programming Tutor:** Taught fundamentals of computer programming to underprivileged communities, 2017.

## SELECTED PAPERS

---

- *Propaganda Política Pagada: Exploring U.S. Political Facebook Ads en Español* - **Bruno Coelho**, Tobias Lauinger, Laura Edelson, Ian Goldstein, Damon McCoy - ACM Web Conference 2023 (WWW 2023) / doi
- *Optimal Team Economic Decisions in Counter-Strike* - Peter Xenopoulos, **Bruno Coelho**, Claudio Silva - IJCAI 2021

## AWARDS

---

- Winner of the NDD Tech Challenge - First place building ML models to classify fiscal documents
- Finalist of the Data Science Games 2018 - top 20, from 400 participating teams across 220 universities and 46 countries.
- Silver medal on Elo's Merchant Category Recommendation on Kaggle. Over 4000 teams of 1-4 people took part.
- 19th place at Data Mining Cup, from 149 Teams from 114 universities in 28 countries - Best team from Brazil.
- Honorable mention at USP's International Symposium of Undergrad Research. 600 mentions given from 3200 applicants.
- Recipient of Academic Featured Student (2016), for students with a GPA above 9 / 10. Awarded to 7 / 120 students.

## INVITED EVENTS

---

- **PyData 2023:** Invited participant and scholarship winner.
- **NYU:** Invited speaker at NYU's PhD prospective students and NYU's undergraduate panel in academia (2022, 2023).
- **Khipu 2019:** Latin American Meeting In Artificial Intelligence, invited participant with full merit-based scholarship.
- **Data Day:** Head of the discussion panel "Present and Future of AI" with invited professors and industry members.
- **Mathematics Solutions for Industry Problems Workshop (2020):** Analysing timber transport data for an NGO.