

# JOSÉ MARIA POMBAL

[pombal.josemaria@gmail.com](mailto:pombal.josemaria@gmail.com) | <https://zeppombal.github.io> | [Publications](#)

## WORK EXPERIENCE

### Unbabel

03/23 - Present

#### AI Research Scientist

- Part of Unbabel's AI research team, where my research focuses on all things automatic evaluation and adaptation of multilingual language models.
- Currently working on research projects to better understand bias in [automatic evaluation and model development](#), to develop language model benchmarks fully automatically, and to improve existing methods for multilingual automatic evaluation (preprints of the latter two will be released in April 2025).
- Core developer of [Tower-v2](#), a state-of-the-art LLM for translation, and the winning submission of the WMT24 General MT Shared task.
- Core developer of [Tower](#), a suite of open state-of-the-art LLMs for translation and related tasks (spotlight at COLM 2024). I lead the effort of continued pre-training, and co-lead the stream of evaluation.
- Developer on the [EuroLLM](#) project, having been involved in performing automatic evaluation focused on multilingual capabilities, and translation-related tasks.
- Assisted in the evaluation and subsequent productization of internal versions of Tower models, which lead to substantial cost savings.
- Researched methods to improve the company's automatic translation quality estimation technology, contributing to the development of [COMETKiwi23](#), the winning system at the WMT 23 QE Shared Task.

### Feedzai

07/21 - 02/23

#### Machine Learning Research Intern

- Part of the Fairness, Accountability, Transparency, and Ethics (FATE) research group, dedicated to implementing fair and trustworthy AI in the context of the company's fraud detection product.
- Researched methods to understand bias and unfairness along the various stages of a high-stakes, dynamic ML pipelines..
- Co-developed BAF: the first realistic, [publicly available](#), large-scale bank account opening fraud dataset suite, which resulted in a NeurIPS 2022 paper, and a patent granted by the US Patent Office.
- Published work on data bias and algorithmic fairness in static and performative prediction settings to workshops in the ICML 2022, KDD 2022, and ICLR 2023 conferences.
- Co-supervised a university research project on rejection inference and the selective labels problem, and their impact on the performance and fairness of fraud detection models.

### McKinsey & Company

06/20 – 07/20

#### Business Analyst intern

- Designed two corporate reorganization and post-merger integration workshops, and developed analytical models to assess the sales capabilities by region and product of a company the client wished to acquire.

## EDUCATION

### Instituto Superior Técnico

Lisbon, Portugal

#### Doctoral Studies (PhD) in Computer Science and Engineering (AI/ML)

09/24 – Present

- Industrial ELLIS PhD co-located at Unbabel.
- Advised by André F.T. Martins and Ricardo Rei.

### Instituto Superior Técnico

Lisbon, Portugal

#### Masters in Data Science and Engineering

09/20 – 12/22

- **Final GPA:** 19/20
- **Relevant Coursework:** Machine Learning (19/20); Deep Learning (19/20); Natural Language Processing (18/20); Information Retrieval (19/20); Information and Database Systems (19/20)
- **MSc thesis:** *Towards a Data-Centric Understanding of Observational Unfairness in Supervised Learning*
  - Developed a holistic theoretical and empirical analysis of the effects of data bias on algorithmic fairness, and proposed a suite of methods to diagnose data bias. Awarded a perfect grade.

**Nova School of Business and Economics***Bachelor of Science in Economics*

Lisbon, Portugal

**09/17 – 06/20**

- **Final GPA:** 18/20
- **Relevant Coursework:** Econometrics (19/20); Finance (18/20); Advanced Microeconomics (18/20)

**BI Norwegian School of Business***Exchange Program*

Oslo, Norway

**08/19 – 12/19****Technical Skills**

---

**Professional proficiency:**

- **Python** (pytorch, sklearn, optuna, pandas). Worked extensively with LLMs and other applications in *NLP*, *fraud detection*, *fair ML*, *MT*, *QE*) (**5+ years**)
- **SQL** (PostgreSQL, experience with dealing with large databases) (**2+ years**)
- **Git** (academic and professional collaboration) (**5+ years**)

**Academic proficiency:** **R** (applied projects in Finance), **Shell**, **JavaScript** (web development experience in d3.js); **Microsoft Office toolkit**;

---

**Additional Information**

---

**Languages:** Portuguese (native); English (fluent, A in CPE Exam); Mandarin Chinese (HSK1); French (beginner);

**Interests:** *Music:* jazz-trained guitarist and self-taught pianist; *Sports:* played futsal for 3 years and basketball for 4 years in the school's teams; *Reading:* interest in history, fiction, and dystopian literature; *Extracurricular activities:* consultant at NJC, a student-run consultancy firm at Nova SBE; economics tutor for high school students.