

JOSÉ MARIA POMBAL

+351 912 662 992 | pombal.josemaria@gmail.com | <https://zeppombal.github.io>

WORK EXPERIENCE

Unbabel

03/23 - Present

AI Research Scientist

- Part of Unbabel's AI research team, focused on leveraging the current advances in LLM technology to improve the company's end-to-end translation service.
- Working on leveraging the capabilities of existing LLM's for machine translation (MT) via finetuning: released [Tower](#), a suite of state-of-the-art LLMs for translation and related tasks.
- Researched methods to improve the company's automatic translation quality estimation services, contributing to the winning system at the WMT 23 QE Shared Task.
- Conducting research on building LLMs from scratch that excel at various tasks in different languages.

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

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 in a series of real-world settings, and proposed a suite of methods to diagnose data bias.

Nova School of Business and Economics

Lisbon, Portugal

Bachelor of Science in Economics

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

Oslo, Norway

Exchange Program

08/19 – 12/19

Technical Skills

Professional proficiency:

- **Python** (deep knowledge of 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) **C** (self-taught), **Shell**, **JavaScript** (web development experience in d3.js); **Microsoft Office toolkit**;

Publications

Rei, R., Guerreiro, N. M., Pombal, J., van Stigt, D., Treviso, M., Coheur, L., ... & Martins, A. F. (2023). Scaling up COMETKIWI: Unbabel-IST 2023 Submission for the Quality Estimation Shared Task. *arXiv preprint arXiv:2309.11925*.

Alves, D. M., Guerreiro, N. M., Alves, J., Pombal, J., Rei, R., de Souza, J. G., ... & Martins, A. F. (2023). Steering Large Language Models for Machine Translation with Finetuning and In-Context Learning. *arXiv preprint arXiv:2310.13448*.

Pombal, J., Saleiro, P., Figueiredo, M. A., & Bizarro, P. (2023). Fairness-Aware Data Valuation for Supervised Learning. *arXiv preprint arXiv:2303.16963*. (*ICLR 2023 Workshop on Pitfalls of limited data and computation for Trustworthy ML*)

Jesus, S., Pombal, J., Alves, D., Cruz, A., Saleiro, P., Ribeiro, R., ... & Bizarro, P. (2022). Turning the tables: Biased, imbalanced, dynamic tabular datasets for ml evaluation. *Advances in Neural Information Processing Systems*, 35, 33563-33575.

Pombal, J., Cruz, A. F., Bravo, J., Saleiro, P., Figueiredo, M. A., & Bizarro, P. (2022). Understanding Unfairness in Fraud Detection through Model and Data Bias Interactions. *arXiv preprint arXiv:2207.06273*. (*KDD 2022 Workshop on Machine Learning in Finance*)

Pombal, J., Saleiro, P., Figueiredo, M. A., & Bizarro, P. (2022). Prisoners of Their Own Devices: How Models Induce Data Bias in Performative Prediction. *arXiv preprint arXiv:2206.13183*. (*ICML 2022 Workshop on Responsible Decision Making in Dynamic Environments*)

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.