

Matteo Guida

PHD RESEARCHER · UNIVERSITY OF MELBOURNE

✉ matteo.guidaa@gmail.com |  LinkedIn |  Google Scholar |  GitHub

Personal Profile

Ph.D. candidate in AI and NLP at the University of Melbourne, with hands-on software engineering experience building and shipping production systems. Designed RAG pipelines, fine-tuned LLMs, and deployed full-stack applications on cloud infrastructure (GCP, AWS, Azure). Five peer-reviewed publications at ACL, TACL, and EMNLP.

Education

University of Melbourne

Melbourne, Australia

Ph.D. in Computer Science (NLP/AI)

April 2024 - April 2027

- Thesis: Towards Models of Framing Detection and Effect Analysis in News Coverage of Contemporary Polarizing Issues

University of Konstanz

Konstanz, Germany

Masters in Speech and Language Processing

Oct 2021 - Feb 2024

- Specialization: Natural Language Processing and Deep Learning

University of Trieste

Trieste, Italy

B.A. in Applied Linguistics

Oct 2018 - July 2021

- Thesis: German in a Non-European Multilingual Context: The Case of Namibia

Industry Experience

University of Melbourne

Melbourne, Australia

Casual Software Developer

May 2026 - Present

- Designed and built an end-to-end NLP pipeline to analyse lecture transcripts, measuring clarity, pacing, and alignment with slide content
- Developed a Canvas LTI dashboard delivering per-lecture feedback to lecturers, integrating with Canvas and Echo360

Alfa Cruise Srl

Remote/Italy

AI Consultant

May 2025 - Present

- Supporting and consulting for the design of AI-native digital platforms with a focus on NLP, user profiling, recommendation systems, and intelligent matching
- Contributing to the development of scalable AI products, particularly systems based on connections, interests, and user behavior

DHL Group

Bonn, Germany

Working Student — Natural Language Processing

September 2023 - February 2024

- Architected and deployed end-to-end NLP solutions on cloud infrastructure (Azure ML, Google Cloud, AWS) for enterprise-scale text processing and content generation
- Led training and fine-tuning of large language models and deployed on Kubernetes (Kubeflow) for production deployment serving business-critical applications
- Engineered retrieval-augmented generation (RAG) systems using LangChain, integrating vector databases and LLM orchestration for intelligent question-answering across multilingual corporate documentation
- Developed and productionized custom neural machine translation models for 15+ language pairs, implementing automated evaluation pipelines and A/B testing frameworks
- Designed domain-specific NER and text summarization models
- **Technologies:** PyTorch, Transformers, LangChain, Kubeflow, Azure ML, Google Cloud (Vertex AI), AWS (SageMaker), Vector DBs, REST APIs, Docker, Git

Deutsche Post DHL Group

Bonn, Germany

Intern — Linguistic Data Management & NLP Engineering

March 2023 - August 2023

- Designed and trained custom NER models for domain-specific terminology extraction from corporate documentation
- Fine-tuned large language models to optimize large-scale document processing and editing workflows for enterprise applications.
- Built data annotation workflows using Prodigy
- Evaluated neural machine translation systems and implemented quality metrics
- **Technologies:** Python, spaCy, Prodigy, Google Cloud AutoML, AWS, Tableau, REST APIs

Research Experience

University of Melbourne

Melbourne, Australia

AI-based App Developer – Interactive Performance Tools

July 2025 - Present

- Developed and deployed interactive web application (depth-perception.onrender.com) for real-time audience data collection across live performance venues
- Built full-stack solution using HTML/JavaScript frontend with AWS S3 backend for data storage and analysis
- Implemented pipeline for statistical analysis of audience perception data
- **Technologies:** HTML/JavaScript, Python, AWS S3, LLMs, Statistical Analysis (R)

University of Melbourne

Melbourne, Australia

AI-based App Developer – Climate & Content Analysis

June 2025 - Present

- Built and deployed LLM-based web applications for analyzing framing in climate news and moderating online comments (climate-narrative-analyzer.onrender.com, comment moderation agent)
- Developed full-stack applications using HTML/JavaScript with MongoDB backend, deployed on Render and Modal with AWS S3 integration
- Implemented custom annotation interface (metaphor-annotation-study.onrender.com) for large-scale collaborative text analysis
- **Technologies:** Python, HTML/JavaScript, LLMs, MongoDB, Render, Modal, AWS S3, Streamlit

University of Melbourne

Melbourne, Australia

Research Assistant – Information Acceleration & Crisis Dynamics

August 2024 - Present

- Contributing to "Measuring Social Media Speed and the Acceleration of Informational Crisis" project, supervised by Dr. Philipp Pond
- Built scalable data pipelines for large-scale social media data collection and temporal analysis
- Analyzing information dissemination patterns during crisis events using network analysis and NLP techniques

University of Konstanz

Konstanz, Germany

Research Assistant – Multimodal Sarcasm Detection

Nov 2022 - Mar 2023

- Investigated contextual triggers for sarcasm detection using multimodal features (text, audio, visual)
- Implemented and evaluated transformer-based models (BERT variants) for contextual understanding
- Conducted data preprocessing, feature engineering, and model evaluation with Python and R

Publications

- **Matteo Guida**, Yulia Otmakhova, Ed Hovy, Lea Frermann (*to appear*). *Article and Comment Frames Shape the Quality of Online Comments*. Findings of the Association for Computational Linguistics (ACL) 2026.
- **Matteo Guida**, Yulia Otmakhova, Ed Hovy, Lea Frermann. *LLMs for Argument Mining: Detection, Extraction, and Relationship Classification of Pre-defined Arguments in Online Comments*. Australasian Language Technology Association Workshop (ALTA) 2025.
- **Matteo Guida**, Yulia Otmakhova, Ed Hovy, Lea Frermann (*to appear*). *Retain or Reframe? A Computational Framework for the Analysis of Framing in News Articles and Reader Comments*. Transactions of the Association for Computational Linguistics (TACL).
- Yulia Otmakhova, **Matteo Guida**, Lea Frermann. Not All ANIMALS Are Equal: Metaphorical Framing through Source Domains and Semantic Frames. Findings of the Association for Computational Linguistics (ACL) 2026.
- Jackson Trager, Francielle Vargas, Diego Alves, **Matteo Guida**, Mikel K Ngueajio, Ameeta Agrawal, Flor Plaza-del-Arco, Yalda Daryanai, Farzan Karimi-Malekabadi. *MFTCXplain: A Multilingual Benchmark Dataset for Evaluating the Moral Reasoning of LLMs through Hate Speech Multi-hop Explanation*. Findings of Empirical Methods in Natural Language Processing (EMNLP) 2025.
- Francielle Vargas, Jackson Trager, Diego Alves, Surendrabikram Thapa, **Matteo Guida**, Berk Atil, Daryna Dementieva, Andrew Smart, Ameeta Agrawal (*to appear*). *Self-Explaining Hate Speech Detection with Moral Rationales*. Findings of the Association for Computational Linguistics (ACL) 2026.

References available upon request