

# Dr. Abelardo Carlos Martínez Lorenzo

AI RESEARCH SCIENTIST | NLP & LANGUAGE MODELS | RESEARCH ENGINEERING

☎ (+34) 653903060 | ✉ abelardomartinezlorenzo@gmail.com | 🌐 abelardocarlos.com | 📄 CarlosML26 | 🌐 carlosml26 | 🎓 Abelardo Carlos

NLP researcher with a PhD and publications in ACL, NAACL, LREC-COLING and AACL. Build neural models, multilingual datasets and evaluation workflows at the World Bank, combining research with retrieval and generation systems and leadership of six direct reports.

## Experience

### Lead AI Scientist

Washington, DC / Rome

WORLD BANK (EXTENDED-TERM CONSULTANT)

Sep 2024 - Present

- Lead an AI science team of **six direct reports**; own **architecture, core backend implementation** and code review for a conversational **evidence-synthesis platform** supporting meta-analysis and development policy.
- Guide and co-author research on **causal effect prediction from natural-language queries**, using a **72,000+ synthetic-query benchmark** and structured LLM-to-encoder pipeline; evaluate cross-domain generalization.
- Design **human-in-the-loop evaluation**, LLM judges and **retrieval benchmarks**; improve prompts and models through annotator feedback, **user sessions** and error analysis.
- Optimize **inference cost and latency** through domain-specific **fine-tuning**, encoder models, batching and caching; build Vertex AI training pipelines and synthetic-data workflows.
- Build **Python/FastAPI REST APIs** for PDF parsing, information extraction, **RAG**, entity linking, SQL analysis and summarization; integrate streaming, failure recovery and automated tests.
- Develop LLM-based **taxonomy generation and refinement** with human validation to structure and retrieve research evidence.
- Own **GCP infrastructure**, resource planning and cloud budgets; establish deployment standards and automate GitHub-to-GCP **CI/CD** with Docker, Cloud Build and Cloud Run.
- Build **terabyte-scale news pipelines** for **food-crisis research** with Pandas, PyArrow and Parquet; guide model evaluation and code improvements for **societal AI research**.
- Translate **user needs** into **LLM integration plans**; coordinate **research roadmaps**, grant proposals, dependencies and resources.

### Senior AI Scientist

Washington, DC / Rome

WORLD BANK (SHORT-TERM CONSULTANT)

Jan 2024 - Sep 2024

- Designed and implemented a **two-stage entity-linking system** using an **e5 retriever** followed by a **fine-tuned DeBERTa encoder** to map research text to an impact-evaluation taxonomy.
- Built **FastAPI/Docker NLP services**; managed **GCP resources**, container orchestration, API deployment and latency optimization.
- Built and evaluated **LLM prompting pipelines** for information extraction and semantic parsing using GPT-4o, LLaMA and Gemini.

### NLP Researcher

Rome, Italy

SAPIENZA NLP, SAPIENZA UNIVERSITY OF ROME

2023 - 2024

- Designed **compact graph representations and encoder-decoder parsers** in PyTorch Lightning, cutting **training time by 82% and inference time by 78%** versus SPRING (BART-large, AMR 3.0) while preserving parsing quality.
- Developed **multilingual semantic-parsing datasets** through structured annotation and **GPT-3.5 fine-tuning**; evaluated on 1,100 human-annotated sentences across 11 languages.
- Led research with PhD students** on multilingual language understanding and LLM-based **parsing, alignment and tagging** workflows.
- Served as a **teaching assistant** for MSc NLP courses and **mentored students** in research methods and implementation.

### NLP Researcher

Rome, Italy

BABELSCAPE

2020 - 2023

- Designed **cross-lingual meaning representations** and **BART/T5** parsing and generation models; delivered Dockerized inference demos.
- Built alignment-aware training with **cross-attention supervision**; co-developed **graph-informed adapters and self-distillation** for encoder-decoder models.
- Developed **autoregressive ensembling and graph validation**, reducing invalid outputs from **13.7% to 0.3%** versus Graphene-SMATCH at equal Smatch on AMR 3.0.
- Co-developed a **multilingual annotated corpus** with hundreds of millions of silver annotations across **four language-understanding tasks and five languages**.

### Research Assistant

Berlin, Germany

TECHNISCHE UNIVERSITÄT BERLIN

Sep 2019 - Jun 2020

- Designed and led implementation** of a **neuro-symbolic framework for knowledge-graph completion** with faculty collaborators, combining embedding models and ontology-based reasoning.
- Built **interchangeable model/OWL2 reasoner interfaces**; evaluated TransE, ComplEx and DistMult on DBpedia and LUBM, improving DistMult MRR from **0.28 to 0.35** on LUBM.

### Software Engineer

Málaga, Spain

FREELANCE

2017 - 2018

- Built a **real-time, multiplatform emergency coordination** app with chat, peer-to-peer **WebRTC** voice, maps, incidents and team assignment.
- Developed the backend with **Node.js and LoopBack**, using **MongoDB** for persistence, and implemented the frontend with **AngularJS**.

Education

PhD in Deep Learning and Natural Language Processing

Rome, Italy

SAPIENZA UNIVERSITY OF ROME

- Marie Skłodowska-Curie fellowship, Horizon 2020 KnowGraphs. Advisor: Roberto Navigli.
- Thesis: Enhancing Semantic Parsing in the Age of Pre-trained Language Models.

MSc in Big Data Management and Analytics

Berlin, Barcelona, Brussels

ERASMUS MUNDUS JOINT PROGRAMME: TU BERLIN, UPC BARCELONA AND ULB BRUSSELS

- Final grade: 9.0/10. Focus: large-scale analytics, big data and business intelligence.

BSc in Software Engineering

Málaga, Spain

UNIVERSITY OF MÁLAGA

- Final grade: 9.0/10. Erasmus exchange: Charles University, Prague.

Publications & Preprint

|      |                                                                                                                                                                                                                                                                                                                      |                       |
|------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------|
| 2026 | Giuliano Martinelli, Piriya Korn Piriya Tamwong, <a href="#">Abelardo Carlos Martínez Lorenzo</a> , Jasmin Baier, Riccardo Orlando, Satvik Garg, Sharif Kazemi, Linxi Wang, Arianna Legovini, Samuel P. Fraiberger. <b>Predicting Causal Effects from Natural Language Queries using Structured Representations.</b> | arXiv preprint        |
| 2025 | <a href="#">Abelardo Carlos Martínez Lorenzo</a> , Alexander Perfilyev, Volker Markl, Martha Clokie, Thomas Sicheritz-Ponten, Zoi Kaoudi. <b>Modular Neuro-Symbolic Knowledge Graph Completion.</b>                                                                                                                  | VLDB Workshop: LS-NSL |
| 2024 | <a href="#">Abelardo Carlos Martínez Lorenzo</a> , Pere-Lluís Huguet Cabot, Karim Ghonim, Lu Xu, Hee-Soo Choi, Alberte Fernández-Castro, Roberto Navigli. <b>Mitigating Data Scarcity in Semantic Parsing across Languages with the Multilingual Semantic Layer and its Dataset.</b>                                 | Findings of ACL       |
| 2024 | <a href="#">Abelardo Carlos Martínez Lorenzo</a> , Roberto Navigli. <b>Efficient AMR Parsing with CLAP: Compact Linearization with an Adaptable Parser.</b>                                                                                                                                                          | LREC-COLING           |
| 2024 | Simone Conia, Edoardo Barba, <a href="#">Abelardo Carlos Martínez Lorenzo</a> , Pere-Lluís Huguet Cabot, Riccardo Orlando, Luigi Procopio, Roberto Navigli. <b>MOSAICo: a Multilingual Open-text Semantically Annotated Interlinked Corpus.</b>                                                                      | NAACL                 |
| 2023 | <a href="#">Abelardo Carlos Martínez Lorenzo</a> , Pere Lluís Huguet Cabot, Roberto Navigli. <b>Cross-lingual AMR Aligner: Paying Attention to Cross-Attention.</b>                                                                                                                                                  | Findings of ACL       |
| 2023 | <a href="#">Abelardo Carlos Martínez Lorenzo</a> , Pere Lluís Huguet Cabot, Roberto Navigli. <b>AMRs Assemble! Learning to Ensemble with Autoregressive Models for AMR Parsing.</b>                                                                                                                                  | ACL                   |
| 2023 | Pavlo Vasylenko, Pere Lluís Huguet Cabot, <a href="#">Abelardo Carlos Martínez Lorenzo</a> , Roberto Navigli. <b>Incorporating Graph Information in Transformer-based AMR Parsing.</b>                                                                                                                               | Findings of ACL       |
| 2022 | <a href="#">Abelardo Carlos Martínez Lorenzo</a> , Marco Maru, Roberto Navigli. <b>Fully-Semantic Parsing and Generation: the BabelNet Meaning Representation.</b>                                                                                                                                                   | ACL                   |
| 2022 | Roberto Navigli, Rexhina Blloshmi, <a href="#">Abelardo Carlos Martínez Lorenzo</a> . <b>BabelNet Meaning Representation: A Fully Semantic Formalism to Overcome Language Barriers.</b>                                                                                                                              | AAAI Blue Sky         |

Skills

|                      |                                                                                                                     |
|----------------------|---------------------------------------------------------------------------------------------------------------------|
| Programming & Data   | Python, SQL, Java, Bash, Pandas, NumPy, PyArrow, Parquet, MySQL, PostgreSQL / pgvector                              |
| ML & NLP             | PyTorch / Lightning, Hugging Face Transformers, scikit-learn, fine-tuning, self-distillation, entity linking        |
| LLMs & Evaluation    | RAG, embeddings, Qdrant, prompting, OpenAI / Gemini APIs, vLLM, human evaluation, retrieval metrics                 |
| Backend & Cloud      | FastAPI, Pydantic, REST APIs, Docker, GCP, Vertex AI, Cloud Run / Build / SQL / Storage, GitHub Actions, AWS Lambda |
| Research & Tools     | Git, Linux, Jupyter, Weights & Biases, Hydra, Poetry, automated testing                                             |
| Technical Leadership | Research planning, code review, team coordination, cloud budgets, technical proposals, mentoring                    |
| Spoken Languages     | English (fluent), Spanish (native), Italian (fluent), Portuguese (intermediate)                                     |

Honors & Academic Service

|      |                                                                                                                                          |
|------|------------------------------------------------------------------------------------------------------------------------------------------|
| 2023 | 2× ISCRA Type C compute allocations. CINECA HPC Center                                                                                   |
| 2023 | Best Reviewer Award. ACL                                                                                                                 |
| 2020 | Marie Skłodowska-Curie PhD Fellowship. Horizon 2020 KnowGraphs                                                                           |
|      | Reviewer. ACL, EMNLP and EACL (2023); LREC (2022); EMNLP (2021); ACL Rolling Review (October and December 2023; February and June 2024). |