

JUAN VÁSQUEZ

[✉](mailto:juanmvs@pm.me) juanmvs@pm.me | [☎](tel:+17206687342) (+1) 720-668-7342 | [in](#) LinkedIn | [GH](#) GitHub | [GS](#) Google Scholar

EDUCATION

University of Colorado Boulder

August 2023 – May 2026

Ph.D. in Computer Science – Focus: Neurosymbolic and Conversational AI

National Autonomous University of Mexico

August 2021 – May 2023

M.S. in Computer Science – Focus: Natural Language Processing

Ibero-American University

January 2015 – December 2019

B.S. in Engineering Physics

EXPERIENCE

Conversational Agent Consultant

December 2023 – May 2026

Neurona

Mexico City, Mexico

- Built a Llama3.3-powered conversational agent providing 24/7 confidential support to workplace violence survivors, connecting users with legal resources and crisis intervention services.
- Designed and deployed a RAG pipeline achieving 100% source retrieval accuracy and 93% keyword coverage on domain-specific legal queries.
- Partnered with legal experts to curate a knowledge base of trauma-informed victim support protocols aligned with restorative justice frameworks.

Graduate Researcher

August 2023 – May 2026

University of Colorado Boulder

Boulder, CO

- Designed a GPU-accelerated podcast analysis pipeline combining WhisperX transcription, pyannote.audio diarization, and 8 transformer models for multi-label hate speech detection, achieving 4.5x speedup through torch.compile optimization and BF16 quantization on NVIDIA H100/A100 GPUs.
- Built a 5-stage speaker identification system using spaCy NER, cross-episode analysis, and LLM-based role classification to automatically replace speaker IDs with actual names across 21+ podcast shows.
- Implemented semantic embeddings-based dehumanization pattern detection using sentence transformers, achieving 50-100x GPU acceleration while maintaining configurable taxonomy-based similarity matching.
- Developed a neurosymbolic architecture combining Abstract Meaning Representations (AMR) with Graph Neural Networks, improving on interpretable text classification analysis.
- Processed 11,000 documents through AMR parsing pipelines to construct semantic graph datasets for GNN training.
- Evaluated GAT, GCN, and GraphSAGE architectures on AMR-derived graphs, establishing baselines for neurosymbolic document understanding.

NLP Consultant

March 2023 – August 2023

University of Arizona

Mexico City, Mexico

- Built text preprocessing pipelines for a 61K-article multilingual corpus of crime-related news from Central America.
- Fine-tuned mBERT for multi-label classification, achieving 87% macro-F1 on crime type detection across Spanish-language news.
- Published findings in a peer-reviewed paper analyzing cross-country classification transfer for low-resource NLP.

Abiogenesis Project

- Led development of *Abiogenesis*, an AI art installation using real-time computer vision to transform physical paintings into interactive digital experiences.
- Curated and annotated a custom dataset to train an object detection model for artistic element recognition.
- Engineered real-time image processing pipelines for live visual transformations during gallery exhibitions.

SELECTED PUBLICATIONS

- **Dehumanization of LGBTQ+ Groups in Sexual Interactions with ChatGPT.** A. Leto, J. Vásquez, A. Palmer, M. L. Pacheco. *Queer in AI Workshop, NAACL, 2025.*
- **HOMO-MEX: A Mexican Spanish Annotated Corpus for LGBT+phobia Detection on Twitter.** J. Vásquez, S. Andersen, G. Bel-Enguix, H. Gómez-Adorno, S. Ojeda-Trueba. *Workshop on Online Abuse and Harms (WOAH), ACL, 2023 (28 citations).*
- **Bert-based Approach for Sentiment Analysis of Spanish Reviews from TripAdvisor.** J. Vásquez, H. Gómez-Adorno, G. Bel-Enguix. *IberLEF @ SEPLN, 2021 (21 citations).*
- **HeteroCorpus: A Corpus for Heteronormative Language Detection.** J. Vásquez, G. Bel-Enguix, S. Andersen, S. Ojeda-Trueba. *Workshop on Gender Bias in NLP (GeBNLP), NAACL, 2022 (14 citations).*
- **Classifying Organized Criminal Violence in Mexico using ML and LLMs.** J. Osorio, J. Vásquez. *Workshop on Socio-political Events (CASE), RANLP, 2023.*

SKILLS

- **Programming:** Python, Rust, Julia, C++, Java, Bash
- **ML/NLP:** PyTorch, HuggingFace Transformers, PyTorch Geometric, scikit-learn, spaCy, NLTK, FAISS, pyannote, SentenceTransformers, bitsandbytes, Networkx, BERTopic
- **Languages:** Spanish (Native), English (Advanced), French (Intermediate)
- **Data:** Pandas, Polars, NumPy, Matplotlib, Seaborn, W&B, Pydantic, Pydantic AI, Pyvis, SQL
- **CLI:** Rich, Click
- **Infrastructure:** Git, HPC clusters, Jupyter, Gradio, Modal, Docker
- **AI CLI Tools:** Claude Code, Copilot CLI, Gemini CLI, Codex, Ollama, llm-mlx
- **Others:** LaTeX, Typst, Markdown, jq, F2, uv, ruff

HONORS & SERVICE

- **1st Place**, Rest-Mex Shared Task on Sentiment Analysis, IberLEF 2021, Málaga, Spain.
- **Keynote Speaker**, “The State of NLP in Latin America”, ACL 2023, Toronto, Canada.
- **Reviewer:** NAACL 2024, LREC-COLING 2024, MICAI 2022/2024, LREC-COLING 2025.
- **Organizer:** Queer in AI Workshop (NAACL 2024, 2025), oSTEM Conference 2024, POPL 2025, HOMO-MEX Shared Tasks (2022–2024).

CERTIFICATIONS & TRAINING

- **Neurosymbolic Programming Summer School**, 2022, California Institute of Technology.
- **Lisbon Machine Learning Summer School (LxMLS)** 2023, Lisbon, Portugal.
- **Program Committee Member**, COMIA 2024 (XVI Mexican Congress on AI), Celaya, Mexico.
- **Participant**, NAACL 2022 (Seattle, WA), 2024 (Mexico City, Mexico) & 2025 (Albuquerque, NM).
- **Participant**, ACL 2022 (Dublin, Ireland), 2023 (Toronto, Canada).
- **Participant**, POPL 2025 (Denver, CO).