Emilio Villa-Cueva

evillacueva@gmail.com **villacu.github.io**
+971 55 8373715 **Google Scholar**

Education

- **Mohamed bin Zayed University of Artificial Intelligence (MBZUAI)** UAE
  *PhD in **Natural Language Processing*** *2024 - Present*
  - **Supervised by**: Dr. Thamar Solorio and Dr. Alham Fikri Aji
  - **Coursework**: Advanced Natural Language Processing, Speech Processing, Advanced Computer Vision, Machine Learning
- **Mathematics Research Center (CIMAT)** Mexico
  *MSc in **Computer Science**; GPA: **9.86/10*** *2022 - 2024*
  - **Thesis**: Adaptation Techniques in Transformers for Text Classification in under-resourced settings
- **University of Guanajuato** Mexico
  *Bachelor in **Engineering Physics**; GPA: **9.79/10*** *2017 - 2022*

Research Experience

- **University of Michigan**
  Visiting Researcher *2025 - Present*
  Working on multimodal conversation understanding with **Dr. Rada Mihalcea**.
- **MBZUAI, Department of NLP**
  Research Assistant *2023 - 2024*
  Conducted research on computationally efficient few-shot cross-lingual transfer with **Dr. Thamar Solorio**.
- **CIMAT, NLP Group**
  Research Assistant *2021 - 2024*
  Conducted research on domain adaptation, few-shot classification, and question answering with **Dr. Adrian Pastor Lopez-Monroy** and **Dr. Fernando Sánchez Vega**.
- **University of Guanajuato, Division of Sciences and Engineering**
  Research Assistant *2020 - 2022*
  Designed and implemented a low-cost meteorological station to measure solar irradiance, supervised by **Dr. Modesto Sosa Aquino**.

Selected Publications

- E. Villa-Cueva, S. Ahmed, R. Chevi, J. C. B. Cruz, K. Elzeky, F. Cristobal, A. F. Aji, S. Wang, R. Mihalcea, and T. Solorio, "Moments: A comprehensive multimodal benchmark for theory of mind," *EMNLP 2025 Findings (Accepted)*, 2025
- E. Villa-Cueva, S. Bolatzhanova, D. Turmakhan, K. Elzeky, H. B. Ademtew, A. F. Aji, I. A. Azime, J. Baek, F. Belcavello, F. Cristobal, *et al.*, "Cammt: Benchmarking culturally aware multimodal machine translation," *EMNLP 2025 Findings (Accepted)*, 2025
- E. Villa-Cueva, A. P. López-Monroy, F. Sanchez-Vega, and T. Solorio, "Adaptive Cross-Lingual Text Classification through In-Context One-Shot Demonstrations," in *NAACL 2024. Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics.*, Association for Computational Linguistics, June 2024
- M. Mitchell, G. Attanasio, I. Baldini, *et al.*, "SHADES: Towards a Multilingual Assessment of Stereotypes in Large Language Models," in *NAACL 2025. Proceedings of the 2025 Annual Conference of the Nations of the Americas Chapter of the Association for Computational Linguistics.*, Association for Computational Linguistics, 2025 (Accepted.)
- D. Romero, C. Lyu, H. A. Wibowo, T. Lynn, I. Hamed, A. N. Kishore, A. Mandal, A. Dragonetti, A. Abzaliev, A. L. Tonja, *et al.*, "Cvqa: Culturally-diverse multilingual visual question answering benchmark," in *NeurIPS 2024. Proceedings of the Neural Information Processing Systems Track on Datasets and Benchmarks*, 2024

- E. Villa-Cueva, M. Valles-Silva, A. P. López-Monroy, F. Sanchez-Vega, and L.-S. J. Roberto, "Few Shot Profiling of Cryptocurrency Influencers using Natural Language Inference & Large Language Models," in *CLEF 2023 Labs and Workshops, Notebook Papers*, 2023
- E. Villa-Cueva, I. González-Franco, F. Sanchez-Vega, and A. P. López-Monroy, "NLP-CIMAT at PoliticEs 2022: PolitiBETO, a Domain-Adapted Transformer for Multi-class Political Author Profiling," in *Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2022)*, CEUR Workshop Proceedings, CEUR-WS.org, 2022

Awards

- **Best Thesis Award**: José Negrete Award, by the *Mexican Society of Artificial Intelligence* (2024)
- **First-Place Prize at PAN-CLEF 2023**: Awarded by *Symanto Research* for Few-shot Learning on Cryptocurrency Influencers (2023)
- **Full Scholarship**: Awarded by the *University of Montreal* for the Montreal Industrial Problem Solving Workshop (2023)

Projects

- **Trademark Collision Detection**: Developed a pipeline for detecting textual, semantic and phonetic collisions in trademarks on the Mexican Institute for Intellectual Property. Project carried out at the SPI Industrial Problem Solving Workshop at the Mathematics Research Center (2023).
- **Synthetic Data Privacy Attacks**: Proposed an approach for attribute inference attacks on synthetic data at the Montreal Industrial Problem Solving Workshop with *Desjardins*. (2023)

Workshops & Panels

- **Panelist at Mexican NLP Summer School**: Participated in the panel: "*Getting into NLP: Insights from Experts and Peers*" at the Mexican NLP Summer School 2024, co-located with NAACL 2024.
- **Workshop Facilitator at MexLef 2022**: Conducted a workshop on Domain Adaptation of Transformers, focusing on adapting BERT models to specific domains.

Skills

- **Languages**: Python, R, C
- **Libraries**: PyTorch, OpenCV, Transformers, Scikit-learn