Constanza Fierro

(+45) 9388 8347
constanza.fierro94@gmail.com
https://constanzafierro.github.io

Education

University of Copenhagen — Copenhagen, Denmark
Ph.D. in Natural Language Processing and Machine Learning — Jul 2021 - May 2025
- Advisor: Prof. Anders Søgaard
- Research area: Interpretability, reliability, and knowledge representations

University of Chile — Santiago, Chile
Computer Science Engineer — March 2013 - July 2019
- Cum. Grade: 6.5/7.0 (GPA 4)

Experience

Google DeepMind — London, United Kingdom
Student Researcher — Oct 2022 – March 2023
- Led experiments on generating text attributed (citing) to source documents.
- Trained T5-XXL models to generate text with planning and attribution.
- Led the preparation of the data between 3 people constructing the pipeline.
- Defined and developed the evaluation pipeline: text and attribution quality, and the relevance of the plan.

Apple (Siri Understanding) — Cambridge, United Kingdom
Research Intern — June 2022 – Sept 2022
- Analyzed the performance of the latest multilingual parser model.
- Proposed and experimented on improvements to the model by training modular language components.
- Improved the performance of the model on low resource languages.

Google (Youtube Video Classification) — Paris, France
Software Engineer — Oct 2019 – June 2021
- Youtube classification platform for training, evaluating and governing classifiers.
- Designed and implemented pipelines for automatic sampling, human labeling, training, and evaluation of classifiers.
- In charge of building and maintaining 3 different binary classification models.
- Developed evaluation analysis for the classifiers and collecting data pipelines.

Research Projects

How Do Multilingual Language Models Remember? — Pre-print
- Mechanistic interpretability analyses of knowledge recall in multilingual LMs.
- Addressing: Do the same mechanisms drive recall in different languages? and, when in the recalling process does language play a role?

Defining Knowledge: Bridging Epistemology and LLMs — EMNLP 2024
- Addressed what it means to say that an LLM knows a propositional fact.
- Surveyed epistemological definitions and formalized interpretations for LLMs.

Learning to Plan and Generate Text with Citations — ACL 2024
- Internship at Google DeepMind (See above).

MuLan: A Study of Fact Mutability in Language Models — NAACL 2024
- Created a dataset and investigated if LLMs recognize that facts change over time.
- Implemented minimum description length to probe the latent representations of the models for fact mutability.
- Tested the accuracy of predictions and updates for different mutability levels.

Do Vision and Language Models Share Concepts?
A Vector Space Alignment Study — TACL 2024
- Measured the alignment of vector spaces in vision and language models with Procrustes.
- Contributed with direction of data preparation, baselines, and paper writing.

RESEARCH PROJECTS

Factual Consistency of Multilingual Language Models   Findings of ACL 2022
- Studied how consistent factual predictions are in different languages.
- Used a mixture of neural translation models to translate a resource of English paraphrases to 45 languages.

Challenges and strategies in cross-cultural NLP   ACL 2022
- Surveyed and discussed strategies for cross-cultural and multicultural NLP.
- Contributed with the "Common ground" dimension that differs among cultures.

AWARDS AND HONORS

- **Distinction**, Graduated with the highest academic distinction   2019
- **Dean's list**, All Semesters   2013-2018
- **Scholar**, IEEE Latin American Summer School in Computational Intelligence   2017
- **11th Place / 450**, Ascenci Dev Cup (France) Finalist   2015

SKILLS

Languages: English, Spanish, French.

Programming: Python, C++, SQL.

Libraries: PyTorch, NumPy, HuggingFace, Pandas, Scikit-learn.

ACADEMIC SERVICES

Reviewers for: *ACL Rolling Review (ARR)*.

PUBLICATIONS

1. **Constanza Fierro**, Negar Foroutan, Desmond Elliott, Anders Søgaard. How Do Multilingual Language Models Remember? Investigating Multilingual Factual Recall Mechanisms. Pre-print.

2. **Constanza Fierro**, Ruchira Dhar, Filippos Stamatiou, Nicolas Garneau, and Anders Søgaard. 2024. Defining Knowledge: Bridging Epistemology and Large Language Models. *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Miami, Florida, USA.

3. **Constanza Fierro**, Reinald Kim Amplayo, Fantine Huot, Nicola De Cao, Joshua Maynez, Shashi Narayan, and Mirella Lapata. 2024. Learning to Plan and Generate Text with Citations. *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers) (ACL)*. Bangkok, Thailand.

4. **Constanza Fierro**, Nicolas Garneau, Emanuele Bugliarello, Yova Kementchedjhieva, and Anders Søgaard. 2024. MuLan: A Study of Fact Mutability in Language Models. *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 2: Short Papers) (NAACL)*, Mexico City, Mexico.

5. Jiaang Li, Yova Kementchedjhieva, **Constanza Fierro**, and Anders Søgaard. 2024. Do Vision and Language Models Share Concepts? A Vector Space Alignment Study. *Transactions of the Association for Computational Linguistics (TACL)*.

6. **Constanza Fierro** and Anders Søgaard. 2022. Factual Consistency of Multilingual Pretrained Language Models. *Findings of the 60th Association for Computational Linguistics (ACL)*. Dublin, Ireland.

7. Daniel Hershcovich, Stella Frank, Heather Lent, Miryam de Lhoneux, Mostafa Abdou, Stephanie Brandl, Emanuele Bugliarello, Laura Cabello Piqueras, Ilias Chalkidis, Ruixiang Cui, **Constanza Fierro**, Katerina Margatina, Phillip Rust, and Anders Søgaard. 2022. Challenges and Strategies in Cross-Cultural NLP. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers) (ACL)*. Dublin, Ireland.