NLP Researcher

Pere-Lluís Huguet Cabot

Natural Language Processing researcher, about to graduate as a National Artificial Intelligence Ph.D. at Sapienza University of Rome. Vast experience with Information Extraction models, with millions of downloads in Huggingface. Recently part of the core team behind the Italian LLM Minerva.

✉ perelluis1993@gmail.com | ⌂ littlepea13.github.io | ▣ littlepea13 | ▣ perelluis | 🎓 Pere-Lluís Huguet Cabot

Experience

Marie Sklodowska Curie PhD *Sapienza University and Babelscape* — Nov/2020 - Oct/2024

Knowgraphs Part of an European Network of Researchers on Knowledge Graphs.

Research

- Creator of the Relation Extraction model REBEL (+400 stars in Github, +2M downloads at Hugginface).
- Core team behind Minerva, the first family of Italian LLM trained from scratch (1B, 3B and 7B).
- Developed the closed Information Extraction model ReLiK (+250 stars in Github).
- Served as a Teaching Assistant for Natural Language Processing M.Sc. course.

Tech Stack Python, Pytorch, Huggingface, Transformers, LLM-foundry

InStep Intern *Infosys Bangalore* — Summer 2019

Responsibilities

- Internship at Infosys. Ranked #1 Best internships 2020.
- Work on a QA system for insurance documents using Transformers.

Tech Stack Jupyter, Colab, Pytorch, Tensorflow

NLP Developer *Amsterdam Data Collective* — 12/2018 - 06/2020

Responsibilities

- Worked on a recommending system for scientific articles.
- Implementing Transformer models in different client projects.

Tech Stack Dash, sklearn, Docker, Pytorch

SOFTWARE DEVELOPER *BaseTIS*   `02/2017 - 08/2018`

Responsibilities

- Web developer using Angular.

- Machine Learning programmer with Python and R.

Tech Stack   Python, R, Angular, JavaScript, PHP, Hyperledger, Unity

Publications

Orlando, R., **Huguet Cabot, Pere-Lluís**, Barba, E., & Navigli, R. (2024, August). ReLiK: Retrieve and LinK, fast and accurate entity linking and relation extraction on an academic budget. *Findings of the Association for Computational Linguistics ACL 2024*, 14114–14132.

Stranisci, M., **Huguet Cabot, Pere-Lluís**, Bassignana, E., & Navigli, R. (2024, August). Dissecting biases in relation extraction: A cross-dataset analysis on people's gender and origin. *Proceedings of the 5th Workshop on Gender Bias in Natural Language Processing (GeBNLP)*, 190–202.

Martinez Lorenzo, A. C., **Huguet Cabot, Pere-Lluís**, Ghonim, K., Xu, L., Choi, H.-S., Fernández-Castro, A., & Navigli, R. (2024, August). Mitigating data scarcity in semantic parsing across languages with the multilingual semantic layer and its dataset. *Findings of the Association for Computational Linguistics: ACL 2024*, 14056–14080.

Conia, S., Barba, E., Martinez Lorenzo, A. C., **Huguet Cabot, Pere-Lluís**, Orlando, R., Procopio, L., & Navigli, R. (2024, June). MOSAICo: a multilingual open-text semantically annotated interlinked corpus. *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)*, 7990–8004.

Huguet Cabot, Pere-Lluís, Tedeschi, S., Ngonga Ngomo, A.-C., & Navigli, R. (2023, July). RED^{FM}: a filtered and multilingual relation extraction dataset. *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 4326–4343.

Martínez Lorenzo, A. C., **Huguet Cabot, Pere-Lluís**, & Navigli, R. (2023a, July). AMRs assemble! learning to ensemble with autoregressive models for AMR parsing. *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, 1595–1605.

Martínez Lorenzo, A. C., **Huguet Cabot, Pere-Lluís**, & Navigli, R. (2023b, July). Cross-lingual AMR aligner: Paying attention to cross-attention. *Findings of the Association for Computational Linguistics: ACL 2023*, 1726–1742.

Vasylenko, P., **Huguet Cabot, Pere-Lluís**, Martínez Lorenzo, A. C., & Navigli, R. (2023, July). Incorporating graph information in transformer-based AMR parsing. *Findings of the Association for Computational Linguistics: ACL 2023*, 1995–2011.

Huguet Cabot, Pere-Lluís, & Navigli, R. (2021, November). REBEL: Relation extraction by end-to-end language generation. *Findings of the Association for Computational Linguistics: EMNLP 2021*, 2370–2381.

Huguet Cabot, Pere-Lluís, Abadi, D., Fischer, A., & Shutova, E. (2021, April). Us vs. them: A dataset of populist attitudes, news bias and emotions. *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*, 1921–1945.

Huguet Cabot, Pere-Lluís, Dankers, V., Abadi, D., Fischer, A., & Shutova, E. (2020, November). The Pragmatics behind Politics: Modelling Metaphor, Framing and Emotion in Political Discourse. *Findings of the Association for Computational Linguistics: EMNLP 2020*, 4479–4488.

Education

PH.D. IN ARTIFICIAL INTELLIGENCE *Sapienza University of Rome*   `EXPECTED 11/2024`

M.SC. IN ARTIFICIAL INTELLIGENCE *University of Amsterdam*   `06/2020`

B.SC. IN PHYSICS *University of Barcelona*   `06/2017`

B.SC. IN MATHEMATICS *University of Barcelona*   `06/2017`

Additional activity

Reviewer   `2021-2024`

At *ACL conferences, EMNLP and ARR.

Languages

Catalan: Native language
Spanish: Native language
English: Fluent (C1)
Italian: Working fluency
German: Beginner