

# Francisco Rodriguez, Ph.D.

✉ [franct14@gmail.com](mailto:franct14@gmail.com) | 📞 +41 78 342 21 20 | [in francisorodriguezsanchez](https://www.linkedin.com/in/francisorodriguezsanchez) |  
🌐 [francisorodriguez92](https://github.com/francisorodriguez92) | [francisorodriguez92.github.io](https://francisorodriguez92.github.io) |  
📍 Saillon, Valais (Wallis), Switzerland



## Summary

---

Senior Data Scientist with a Ph.D. in Natural Language Processing and 9+ years of experience at the intersection of research and industry. Proven track record in designing and deploying end-to-end ML solutions, with expertise in NLP, LLM fine-tuning, predictive analytics, and scalable data pipelines on AWS and GCP. Delivered impactful projects at Hilti and UST, including customer satisfaction modeling and multilingual NLP applications, by integrating modern ML tools into production. Multilingual communicator (English, French and Spanish) with peer-reviewed publications, combining technical depth with business-oriented innovation. Currently based in Valais, Switzerland, working in a hybrid setup.

## Work Experience

---

### Hilti AG

*Senior Data & Research Scientist*

2021 – Ongoing

Paris, France (Hybrid)

- Designed, developed, and deployed LLM use cases to enhance customer experience through analysis of text feedback
- Built and operationalized a predictive ML pipeline leveraging operational KPIs, improving customer satisfaction by 5%
- Architected and deployed production-ready ML pipelines for marketing campaigns, increasing targeting efficiency and scalability.
- Regularly presented LLM/ML research insights to senior management, influencing strategic planning and product direction

### Telefónica (via UST Spain)

*Data Scientist*

2016 – 2021

Madrid, Spain

- Built and deployed ML pipelines for customer satisfaction prediction, enabling proactive customer retention strategies
- Developed ML solutions for early detection of network failures, reducing downtime and improving service reliability
- Designed and maintained ETL workflows for large-scale data extraction and processing on GCP and AWS

### Telefónica

*Talentum Intern*

2015 – 2016

Madrid, Spain

- Developed a big data processing platform using Spark and R (Shiny), deployed on AWS
- Processed and analyzed 1TB+ of network traffic data, providing insights for infrastructure optimization

## Research Experience

---

### UNED, Research group in NLP

*Research Collaborator*

2019 – 2025

Madrid, Spain (remote)

- Conducted research on deep learning for NLP with a focus on bias mitigation and multi-task text classification for sexism detection; experimented with transformer/LLM fine-tuning (LoRA, RLHF) and tool-using agents.
- Principal researcher and co-organizer of the EXIST 2021–2022 shared task, coordinating 50+ teams across 10+ countries and establishing a multilingual benchmark for sexism detection.
- Created and released the *MeTwo* dataset — the first corpus of sexist expressions in Spanish Twitter and an early multilingual resource for sexism research — enabling reproducible benchmarking and follow-on studies.
- Peer reviewer for *IEEE Access*, *IberLEF*, and *CLEF*.

UPCT, Mathematics & Statistics Group
2014 – 2016

Research Intern
Cartagena, Spain

- Investigated time-series modelling and supervised/unsupervised learning for municipal water-consumption forecasting in Murcia (Spain), informing capacity planning and anomaly detection.

## Education

UNED
2019 – 2025

Ph.D. in Natural Language Processing (*summa cum laude*)
Madrid, Spain

UNED
2016 – 2019

M.Sc. in Machine Learning and Natural Language Processing
Madrid, Spain

UPCT
2010 – 2016

B.Eng. & M.Eng. in Telecommunications Engineering
Cartagena, Spain

## Selected Publications

For a comprehensive list, please refer to my [Google Scholar profile](#).

Expert Systems: The Journal of Knowledge Engineering
2024

- “Leveraging Unsupervised Task Adaptation and Semi-Supervised Learning With Semantic-Enriched Representations for Online Sexism Detection.” F. Rodríguez-Sánchez, J. Carrillo-de-Albornoz, and L. Plaza. [doi:10.1111/exsy.13763](#)

Applied Intelligence
2024

- “Detecting Sexism in Social Media: An Empirical Analysis of Linguistic Patterns and Strategies.” F. Rodríguez-Sánchez, J. Carrillo-de-Albornoz, and L. Plaza. [doi:10.1007/s10489-024-05795-2](#)

IEEE Access
2021

- “Automatic Classification of Sexism in Social Networks: An Empirical Study on Twitter Data.” F. Rodríguez-Sánchez, J. Carrillo-de-Albornoz, and L. Plaza. [doi:10.1109/ACCESS.2020.3042749](#)

## Skills

Programming	Python (NumPy, pandas, scikit-learn), R (dplyr, ggplot, Shiny), SQL
ML / NLP	PyTorch, Hugging Face Transformers, LLM fine-tuning (LoRA, RLHF), multi-task learning
Data / Cloud	AWS, GCP, Spark, Hadoop (Hive, Pig, Oozie)
Deployment	Docker, Git, CI/CD, ML model deployment and monitoring
Visualization	Tableau, Power BI

# Projects

---

PhD Sexism Detection System	Achieved state-of-the-art results in the EXIST 2021 and 2022 shared tasks on sexism detection, ranking #1 among 80+ international teams; methodology published in <a href="#">Expert Systems (2024)</a> .
EXIST Shared Task	Principal organizer of EXIST 2021, the first international benchmark on sexism detection; the initiative has grown into an annual CLEF evaluation campaign, continuing through 2025 ( <a href="#">Link</a> ).
MeTwo Corpus	Creator of <i>MeTwo</i> , the first multilingual dataset of sexist expressions in Spanish Twitter, publicly released and widely adopted as a benchmark resource ( <a href="#">GitHub</a> ).

# Awards

---

Ph.D. Distinction	<i>Summa cum laude</i> distinction; nominated for Best Ph.D. Thesis Award at UNED (2025).
Odilo Award	Winner, Best Final Project in Web Applications (M.Eng.), UPCT, sponsored by Odilo (2017).
Talentum Award	Winner, Best Startup Project at UPCT Talentum program, Telefónica (2016).
B.Eng. & M.Eng. Thesis	Both undergraduate and master’s theses nominated for Best Final Project at UPCT (2014 & 2016).

# Languages

---

Spanish	Native
English	Professional working proficiency
French	Professional working proficiency
German	Beginner