

# Javier García Gilabert

☎ (+34) 618 29 22 09 | ✉ javier.garcia1.bsc@gmail.com | 🏠 Scholar | 🐙 Github | 🔗 Linkedin | 🏠 Homepage

## Education



### Master's Exchange Student (Erasmus)

Helsinki, FI

AALTO UNIVERSITY

Sep. 2023 - Jan. 2024

Took part in advanced Master's courses covering topics such as statistics, computer vision, complex networks, and computational genomics during my Master's exchange program.



### Master in Artificial Intelligence

Barcelona, ES



POLYTECHNIC UNIVERSITY OF CATALONIA (UPC) & UNIVERSITY OF BARCELONA (UB) & UNIVERSITY OF ROVIRA I VIRGILI (URV)

2022 - 2024

- Interuniversity master (taught in English).
- Achieved a mean degree mark of 8.85/10.
- Master Thesis: *Improving Multilingual Neural Machine Translation by projecting language representations.*



### Bachelor's degree in Data Science

Valencia, ES

POLYTECHNIC UNIVERSITY OF VALENCIA (UPV)

2018 - 2022

- Got the best student award in Data Science. Honor granted for achieving the highest marks in my university degree.
- Achieved a mean degree mark of 9.3/10.
- Subjects passed with honors: 13.
- Received a Special Mention for my Bachelor Thesis: *Image Captioning using pre-trained GPT-2 models.*

## Experience



### Research Engineer in NLP

Barcelona, ES

BARCELONA SUPERCOMPUTING CENTER

Jan. 2024 - Present

Working as a research engineer in the Language Technologies Group at the Barcelona Supercomputing Center.

- NLP research engineer in Eloquence (European Project), ILENIA (Spanish National Project), AINA (Catalan Project).
- Developed Multilingual Neural Machine Translation models LLM-based (SalamandraTA, SalamandraTA-7b-instruct, Plume).
- Worked in low-resource machine translation.
- Experience in training LLMs in multi-node distributed environments using DeepSpeed and Nemo frameworks.



### Research Engineer in NLP

Barcelona, ES

MT@UPC (MACHINE TRANSLATION GROUP AT UPC)

Oct. 2022 - Sep. 2023

National Spanish Project ROB-IN (PLEC2021-007859). The primary goal of this project was to create a robot designed for continuous personalized assistance with self-explanatory capabilities. My role in this project involved contributing to the development of the Natural Language Processing (NLP) component. Duties included:

- Developing the dialogue system of the robot using the RASA framework.
- Investigating the phenomenon of hallucinations in neural machine translation models.



### Research Collaboration Fellowship

Valencia, ES

PRHLT (PATTERN RECOGNITION AND HUMAN LANGUAGE TECHNOLOGY RESEARCH CENTER)

Oct. 2021 - Aug. 2022

I received a collaboration fellowship with the PRHLT group from the Spanish Ministry of Education, where I focused on researching state-of-the-art encoder-decoder architectures for image captioning. This research led to the development of my bachelor's thesis under the supervision of Francisco Casacuberta Nolla.



### Undergraduate Research Assistant

Valencia, ES

DEPARTMENT OF APPLIED STATISTICS, OPERATIONS RESEARCH AND QUALITY AT UPV

Oct. 2019 - Aug. 2022

My role was mainly as a data scientist, dealing with data collection, processing, analysis, and the application of machine learning techniques to derive meaningful insights and solutions.

## Publications



**Javier Garcia Gilabert**, Aleix Sant, Carlos Escolano, Francesca De Luca Fornaciari, Audrey Mash, and Maite Melero. 2024. BSC Submission to the AmericasNLP 2024 Shared Task. *In Proceedings of the 4th Workshop on Natural Language Processing for Indigenous Languages of the Americas (AmericasNLP 2024)*, pages 143–149, Mexico City, Mexico. Association for Computational Linguistics.



**Javier Garcia Gilabert**, Carlos Escolano, and Marta R. Costa-Jussà. 2024. ReSeTOX: Re-learning attention weights for toxicity mitigation in machine translation. *In Proceedings of the 25th Annual Conference of the European Association for Machine Translation*, pages 37–58, Sheffield, England. European Association for Machine Translation.

Preprint

**Javier Garcia Gilabert**, Carlos Escolano, Aleix Sant, Francesca De Luca Fornaciari, Audrey Mash, Xixian Liao, and Maite Melero. 2024. Investigating the translation capabilities of Large Language Models trained on parallel data only. *arXiv preprint arXiv:2406.09140*.



Aleix Sant, Daniel Bardanca, José Ramon Pichel Campos, Francesca De Luca Fornaciari, Carlos Escolano, **Javier Garcia Gilabert**, Pablo Gamallo, Audrey Mash, Xixian Liao and Maite Melero. 2024. Training and Fine-Tuning NMT Models for Low-Resource Languages using Apertium-Based Synthetic Corpora. *The Ninth Conference on Machine Translation (WMT24)*.



**Javier Garcia Gilabert**, Carlos Escolano, Audrey Mash, Xixian Liao, and Maite Melero. 2024. MT-LENS: An all-in-one Toolkit for Better Machine Translation Evaluation. *arXiv preprint arXiv:2412.11615*. *Accepted for publication in the Conference of the North American Chapter of the Association for Computational Linguistics (NAACL) 2025*.

## Awards and Grants

- |      |  |                 |
|------|--|-----------------|
| 2022 | <b>Best Student award in Data Science</b> Honor granted for achieving the highest marks in my university degree.   | UPV             |
| 2022 | <b>Beca Santander Estudios Progreso 2021-2022</b> The "Becas Santander Progreso" program is a grant given by Banco Santander S.A., whose objective is to facilitate access to education for students with the best academic records at a national level who are pursuing official Bachelor's or Master's degree studies. | Banco Santander |
| 2021 | <b>Awards to Graduates of the UPV with Best Academic Records</b> Honor granted for achieving the highest marks in the second academic year of my University degree.  | UPV             |
| 2021 | <b>National Datathon Cajamar UniversityHack 2022</b> Water prediction model presented in the Cajamar Water Footprint 2022 challenge of the national datathon University Hack 2022. We were able to arrive to the national final round.   | Cajamar         |

## Service in the field

- |      |                                  |
|------|----------------------------------|
| 2025 | <b>Reviewer</b> MT Summit 2025   |
| 2025 | <b>Reviewer</b> NAACL Demo track |

## Certificates

- |      |   |           |
|------|---|-----------|
| 2023 | <b>DeepLearn 2023 summer: 10th International Gran Canaria School on Deep Learning</b> | DeepLearn |
| 2017 | <b>First Certificate of English (FCE)</b>   | Cambridge |

## Skills

<b>Programming</b>	Python, R, LaTeX, Bash.
<b>Frameworks</b>	PyTorch, numpy, pandas, nemo, deepspeed, slurm, gradio, vllm, fastapi, lm-evaluation-harness, accelerate, transformers.
<b>Languages</b>	Spanish (Native), Catalan (Native), English (Fluent).

## Volunteering Activities

<b>Mentor for international exchange students</b>	UPV
MENTOR	2022
Assisting international exchange students in their adaptation to the university and the city of Valencia. As part of the university's MENTOR support program: <a href="https://www.upv.es/entidades/OPII/infoweb/pi/info/820230normali.html">https://www.upv.es/entidades/OPII/infoweb/pi/info/820230normali.html</a>	

## **Mentor for new university students**

PATU

Advising new Data Science students in their transition to the university, as part of PATU (Programa de Acción Tutorial):  
<https://www.upv.es/perfiles/futuro-alumno/integra-patu-es.html>

UPV  
2019-2021