William Eduardo Soto Martinez

williamsotomartinez@gmail.com | +33 07 54 13 00 59 | Nancy, France

<u>Github: Sotwi | Gitlab: Williamsotomartinez | Inria Gitlab: Wsotomar | LinkedIN: Wilsoto | sotwi.github.io</u>

WORK EXPERIENCE

Centre National de la Recherche Scientifique, France

PhD Researcher | October 2021 - Present

- Conduct research for one of the largest fundamental science agencies in Europe.
- Apply generative AI, in particular LLMs, to multilingual generation and evaluation of text from Knowledge Graphs and Semantic Graphs.
 - **High Level Publications:** Multiple publications in conferences indexed by the ACL Anthology including ACL, IJCNLP / AACL, and INLG.
 - **Open Source Code and Models:** Publicly available open source code and models for multilingual text generation and evaluation.

M2 Researcher | March 2021 - August 2021

• Development of a Multilingual Paraphrase Evaluation metric.

M1 Researcher | June 2020 - July 2020

• Development of a Language Identification model.

Universidad de Costa Rica, Costa Rica

Research and Teaching Assistant

• Research in Data Anonymization, Multi-Agent Simulation, and Cloud Architecture.

EDUCATION

Université de Lorraine, France

PhD in Informatics

• Thesis subject: Multilingual Graph-to-Text Generation and Evaluation.

MSc in Natural Language Processing

- Thesis subject: X-ParEval: A Multilingual Metric for Paraphrase Evaluation.
- Grade Average: 16.40/20.00 Rank: 4/20 Mention: Très Bien

Universidad de Costa Rica, Costa Rica

 $BSc\ in\ Computer\ Sciences$

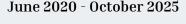
• Grade Average: 8.7 / 10

FIRST AUTHOR PUBLICATIONS

- Semantic Evaluation of Multilingual Data-to-Text Generation via NLI Fine-Tuning: Precision, Recall and F1 scores (Soto Martinez et al, ACL 2025) *To be published*
- <u>Generating from AMRs into High and Low-Resource Languages using Phylogenetic Knowledge and</u> <u>Hierarchical QLoRA Training (HQL) (Soto Martinez et al., INLG 2024)</u>
- <u>Phylogeny-Inspired Soft Prompts For Data-to-Text Generation in Low-Resource Languages (Soto</u> <u>Martinez et al., IJCNLP-AACL 2023)</u>

SKILLS AND INTERESTS

- Languages: Spanish (Native), English (C1), French (B1).
- Technologies: Python, PyTorch, Transformers, SBERT, SciPy, Scikit-Learn, Pandas, SPARQL, SQL.
- Interests: Languages, Creative Writing, Literature, Game Development.



October 2025 (Pending defense)

March 2017 - August 2019

August 2021

December 2017