

Nathan Godey

LinkedIn: <https://www.linkedin.com/in/nathan-g-114298175/>

Github: <https://github.com/NathanGodey>

Personal blog: <https://nathangodey.github.io/>

Email : godeynathan@gmail.com

Mobile : +33 06 98 21 42 25

PUBLICATIONS

- **On the Scaling Laws of Geographical Representation in Language Models:** COLING'24 (review)
- **Headless Language Models: Learning without Predicting with Contrastive Weight Tying:** ICLR24(review)
- **Anisotropy is Inherent to Self-Attention in Transformers:** ACL-SRW'23, EACL'24 (review)
- **MANTa: Efficient Gradient-Based Tokenization for Robust End-to-End Language Modeling:** EMNLP'22 Findings

EDUCATION

- **Inria/ALMAAnaCH, Université Paris Sorbonne** Paris, France
PhD. in Natural Language Processing (pending) Dec 2021 - now
Subject: Cheap and Expressive Contextual Neural Representations for Textual Data
- **Université Paris-Saclay** Saclay, France
Master MVA courses (no diploma; GPA: 4.0 (17.3/20)) Sep 2020 - Jun 2021
Courses: Speech and NLP, Computational Optimal Transport, Image Denoising, Graphs in ML, Modelling in Neuroscience, Deep Learning
- **École Nationale des Ponts et Chaussées** Paris, France
Engineering Degree ; GPA: 3.65 Sep 2017 - Jun 2021
Courses: Advanced Analysis, Optimization, Machine Learning, Advanced Algorithmic, Operations Research, Probability and Statistics

EXPERIENCE

- **EPITA (Engineering School)** Paris, France
Responsible of the Advanced NLP course (Masters) Dec 2023 - Mar 2024
- **Université Pierre et Marie Curie** Paris, France
Teaching Assistant in Python (Bachelors) Jan 2023 - May 2023
- **Université Paris 1** Paris, France
Teaching Assistant in Linear Algebra (Bachelors) Jan 2022 - May 2022
- **Ministry of Defense** Paris, France
R&D internship in Natural Language Processing May 2021 - Nov 2021
 - **NLP in low-resource setting:** Wrote a codebase from scratch to benchmark different state-of-the-art approaches (BERT, RNNs, CRF) on in-domain noisy datasets
 - **Improved performance for the specific use-case:** Leveraged semi-supervised learning, multilingual alignment and synthetic noise to improve on state-of-the-art methods

SKILLS SUMMARY

- **Languages:** Python, C++, Javascript, SQL, Unix scripting
- **Tools:** PyTorch (+ Lightning), Hydra, Pytest, ReactJS, Git, Flask, ExpressJS, AirFlow, Postgres

ACADEMIC PROJECTS (AVAILABLE ON GITHUB)

- **Cycle-consistent Voice Conversion:** Developed a new method based on CNN to convert the speech of some speaker into the same speech as said by another speaker. Trained for several days on Colab to obtain interesting results.
- **Stochastic simulation of Argon gas:** Implemented a simulation of an Argon particles set in C++ following stochastic differential equations to verify thermodynamical properties as a Statistical Physics project.
- **Metaheuristics for the Minimum Connected k-Cover problem in a Network of Sensors:** Used a Genetic Algorithm to tackle a problem about covering an area with connected sensors. Added advanced features to improve diversity at convergence and improve results.
- **Simulation of Neurofeedback-Induced Striatal Learning:** Studied Neuro-feedback via simulation using partial differential equations and exposed interesting properties that were studied in practice.