

Sonal Sannigrahi

+33 0662574908 | sannigrahisonal@gmail.com | sonalsannigrahi.github.io | **GitHub:** sonalsannigrahi

EDUCATION

University of Saarland, Saarbrücken, Germany

2021-expected August 2023

- MSc. in Language Science and Technology | **GPA: -**
- **Research Fellow** at German Center of Artificial Intelligence within the Multilingual Technologies Lab
- **Relevant Coursework:** Ethics in NLP, Foundations in Linguistics, Sociophonetics, Computational Linguistics, Advances in Question Answering, Foundations in Empirical Methods, Slavic Linguistics

École Polytechnique, Palaiseau, France

2018-August 2021

- BSc. in Mathematics and Computer Science | **GPA: 4.21/4.0+** (most recent, 3.89 overall)
- Honours Grant of 10500 Euros upon admission | **Google Generation Scholarship 2020**
- **Thesis:** Low Resource Machine Translation with Cross-Lingual Mapping (A+) under Dr. Jesse Read
- **Relevant Coursework:** Design and Analysis of Algorithms, Machine Learning, Computer Architecture, Intro to Networks, OOP in C++, Functional Programming, Probability and Statistics, Algebra, Analysis, Formal Languages, Deep Learning in Computer Vision (Graduate Course), Concurrent Programming, Image Analysis, Numerical ODEs
- *Other honours:* **vGHC Scholarship EMEA 2021**, Next Genius full tuition scholarship finalist

TECHNICAL EXPERIENCE and RESEARCH

INRIA, Paris, France- **NLP Research Intern**

June-Sept 2021

- Joined the ALMAAnaCH (Automatic Language Modelling and Analysis & Computational Humanities) team under the advisory of Dr. Rachel Bawden to research interpretability of large multilingual models and to improve current techniques in low resource translation via tokenisation.
- Designing experiments to test how different granularities in tokenisation (subword, character, byte) affects language sharing in multilingual machine translation for low resource scenarios with lexical overlap
- Conducted large scale literature reviews in multilingual representation and cross lingual learning.

LIX, Palaiseau, France- **Bachelor Thesis Research Intern**

Jan- Apr 2021

- Developed a novel method to compute cross-lingual word embeddings which solves the isometry problem
- Applied my findings to the case of supervised machine translation for Nepali-English and improved benchmark performance by over 75%
- The word embeddings themselves beat both Mapping and Joint methods in Eigen similarity as well as Bilingual Lexicon Induction accuracy (*paper in submission*)

Linagora, Paris, France- **NLP Research and Development Intern**

June-Sept 2020

- Built an on-device next word predictor in French and English similar to mobile keyboards, for emails
- Evaluated several architectures for the task: Bi-LSTM, Transformers, CNN-LSTM, GRU
- Comparing different word2vec representations to attain better accuracy on word level language models
- Used knowledge distillation to compress models, experimented with Flask for web deployment
- Presented my work in an oral presentation as well as in a medium article on the company blog

École Polytechnique, Palaiseau, France- **Research Assistant**

June-Sept 2019

- Worked under Prof. Pierre Boyer to write code in Python to calculate Twitter penetration rates for a given region and time period to study the impacts of social media involvement in local legislation
- Used R to create several different visualisation techniques at city, county, state levels of accuracy and trained Naïve Bayes Classifier model for sentiment analysis on bilingual tweets
- Used Twitter REST API and Get OAuth methods to geolocate twitter JSON data via the Carmen library.

SKILLS AND ADDITIONAL INFORMATION

Technical:

- **Programming:** Python, C++, familiar: HTML/CSS, JavaScript, R, Shell, OCaml, C, Prolog, Coq
- **Tools and Packages:** pyTorch, Keras, Tensorflow, NumPy, SciPy, Pandas, scikit-learn, and more

Languages: English, Hindi, Bengali (oral native tongue), French (professional fluency)

Google Get Ahead'20, 21: Invite-only development programme in EMEA, weekly assignments in Python

Professional Service: Reviewer at NLP Beyond Text Workshop at EMNLP 2020 and at TWC 2021

Invited Talks: "APIs and Language Processing with Python for Twitter" at PyCON FR 2019

"Moving in Sync: Self-supervised learning of n-human interactions" at PyCON SA 2021