

Giorgos Vernikos

Chemin de Mornex 5B, Lausanne, Switzerland

☎ +41 0786786698 • ✉ georgios.vernikos@epfl.ch • 🌐 georgevern.github.io
🔍 google scholar • 🐙 github • in linkedin • 🐦 twitter

Research Interests

multilingual & monolingual natural language processing, large language models, (un)supervised machine translation, domain adaptation, subword segmentation algorithms, machine translation evaluation, active learning

Education

Ph.D candidate

École Polytechnique Fédérale de Lausanne (EPFL), Electrical Engineering (EDEE)

Lausanne, Switzerland

08/2020–Present

○ Supervisor: Prof. [Andrei Popescu-Belis](#)

○ Graduation (expected): April 2025

Master of Science (MSc) in Data Science and Machine Learning

National Technical University of Athens (NTUA)

Athens, Greece

11/2018–07/2020

○ GPA: 9.22/10 (excellent)

○ Thesis: *Adversarial Fine-tuning of Pretrained Language Models*

○ Supervisors: Prof. [Andreas Stafylopatis](#) and Dr. [Georgios Siolas](#)

Erasmus exchange student

Universidad Carlos III de Madrid (UC3M), School of Electrical Engineering

Madrid, Spain

01/2016–06/2016

BEng & MEng in Electrical and Computer Engineering

National Technical University of Athens (NTUA)

Athens, Greece

10/2011–11/2017

○ GPA: 8.02/10 (top 15% of class) Digital Signal Processing, Neurofuzzy Industrial and Robot Control

○ Thesis: *Text-content based analysis of emotions: A Comparative Study of Machine Learning Algorithms and Data Stream Storage Techniques*

○ Supervisor: Prof. [Iakovos Venieris](#)

Professional and Research Experience

Research Intern

Google, Google Assistant

Zurich, Switzerland

10/2022–02/2023

Research on efficient adaptation of large language models.

Supervisors: [Eric Malmi](#), [Jakub Adamek](#) and [Aliaksei Severyn](#)

Applied Scientist Intern

Amazon Web Services, AI Labs

Santa Clara, CA, USA

04/2022–08/2022

Research on machine translation evaluation.

Supervisors: [Brian Thompson](#), [Prashant Mathur](#) and [Marcello Federico](#)

Research Assistant

HEIG-VD, Institute for Information and Communication Technologies (IICT)

Yverdon-Les-Bains, Switzerland

08/2020–Present

○ Projects:

- [DOMAT](#): On-demand Knowledge for Document-level Machine Translation, supported by SNSF

- FamilyMT: Multilingual NMT for families of languages and dialects, supported by Armasuisse

- UniSub: Unsupervised NMT with innovative multilingual subword models, supported by Armasuisse

- Teaching:
 - Analysis of Text Data (AnTeDe) (2020 & 2021)

Machine Learning Engineer

DeepSea Technologies, AI Research Department

Vessel monitoring and cost optimization using noisy and missing data.

Pireaus, Greece

09/2018–02/2020

Program Committee

- Reviewer for ACL, EMNLP, NAACL, EACL, ARR, DADC, HCRL
- AAAI 2024 Workshop on Human-Centric Representation Learning ([HCRL 2023](#))

2019–Present

Compulsory Military Service

Communications and Network Center

12/2017–8/2018

Private Tutor

Freelancer

Athens, Greece

09/2011–06/2017

Publications

- [1] **Vernikos, G.**, and Popescu-Belis, A. (2024). *Don't Rank, Combine! Combining Machine Translation Hypotheses Using Quality Estimation*. [\[paper\]](#)
- [2] **Vernikos, G.**, Bražinskas, A., Adamek, J., Mallinson, J., Severyn, A., and Malmi, E. (2023). *Small Language Models Improve Giants by Rewriting Their Outputs*. EACL 2024. [\[paper\]](#)
- [3] Wolleb, B., Silvestri, R., **Vernikos, G.**, Dolamic, L. and Popescu-Belis, A. (2023). *Assessing the Importance of Frequency versus Compositionality for Subword-based Tokenization in NMT*. EAMT 2023. [\[paper\]](#) [\[code\]](#)
- [4] Popescu-belis, A., Atrio, À.R., Bernath, B., Boisson, E., Ferrari, T., Theimer-lienhard, X., and **Vernikos, G.** (2023). *GPoeT: a Language Model Trained for Rhyme Generation on Synthetic Data*. LaTeCH-CLfL 2023 (oral). [\[paper\]](#) [\[code\]](#)
- [5] **Vernikos, G.**, Thompson, B., Mathur, P. and Federico, M. (2022). *Embarrassingly Easy Document-Level MT Metrics: How to Convert Any Pretrained Metric Into a Document-Level Metric*. WMT22 (oral). [\[paper\]](#) [\[code\]](#)
- [6] **Vernikos, G.**, and Popescu-Belis, A. (2021). *Subword Mapping and Anchoring across Languages*. EMNLP 2021 Findings. Presented at WMT21 (oral). [\[paper\]](#) [\[code\]](#)
- [7] Margatina, K., **Vernikos, G.**, Barrault, L., and Aletras, N. (2021). *Active Learning by Acquiring Contrastive Examples*. EMNLP 2021 (oral). [\[paper\]](#) [\[code\]](#)
- [8] Atrio, À.R., Luthier, G., Fahy, A., **Vernikos, G.**, Popescu-Belis, A. and Dolamic, L. (2021). *The IICT-Yverdon System for the WMT 2021 Unsupervised MT and Very Low Resource Supervised MT Task*. WMT21. [\[paper\]](#)
- [9] **Vernikos, G.**, Margatina, K., Chronopoulou, A., and Androutsopoulos, I. (2020). *Domain Adversarial Fine-Tuning as an Effective Regularizer*. EMNLP 2020 Findings. Presented at SustaiNLP (oral). [\[paper\]](#) [\[code\]](#)

Invited talks

- [Workshop on Natural Language Processing](#), Archimedes Research Center / Athens, Greece
- [IST & Unbabel Seminars](#), Instituto Superior Técnico (IST) / virtual, on-line

Honors and awards

- **Facebook Scholarship for Attendance at EurNLP 2019**

Recipient of travel grant in order to attend the EurNLP summit in London, UK.

- **Nationwide University Entrance Examination**

Scored 19.398/20 on the national qualification exams (top 1% nationwide).

Attended seminars

- Lisbon Machine Learning School ([LxMLS](#)), virtual, on-line
- Athens Natural Language Processing Summer School ([AthNLP](#)), Athens, Greece

Programming skills

- **Languages:** Python, R, C, Matlab, Bash, Assembly, SQL
- **ML Frameworks:** Pytorch, Tensorflow, Keras, HuggingFace, scikit-learn
- **Big Data Tools:** Apache Storm, Apache Spark, Apache Flume
- **Other:** \LaTeX , Git, Flask

Languages

Greek (native), English (C2), Spanish (B2), German (B1)