

Giorgos Vernikos

Chemin de Mornex 5B, Lausanne, Switzerland

+41 0786786698 • georgios.vernikos@epfl.ch • georgevern.github.io
google scholar • github • linkedin • twitter

Research Interests

My research focus lies in the intersection of Machine Learning and Natural Language Processing. I am interested in large language models (LLMs), machine translation (MT) systems and evaluation.

Professional Experience

Research Intern **Zurich, Switzerland**
Google, Google Assistant *10/2022–02/2023*

- Conducted research on improving the performance of large language models for specific tasks.
 - Proposed the use of smaller models to correct the predictions of LLMs, *published* at EACL 2024.
- Supervisors: [Eric Malmi](#) and [Aliaksei Severyn](#)

Applied Scientist Intern **Santa Clara, CA, USA**
Amazon Web Services, AI Labs *04/2022–08/2022*

- Conducted research on document-level machine translation evaluation which was *published* at WMT 2022.
 - Proposed document-level metrics that have been used as baselines in subsequent [WMT metrics shared tasks](#).
- Supervisors: [Brian Thompson](#) and [Marcello Federico](#)

Machine Learning Engineer **Athens, Greece**
DeepSea Technologies (startup, acquired by Nabtesco) *09/2018–02/2020*

- Contributed to the foundational team as one of the first 10 employees.
- Developed neural networks for vessel monitoring and cost optimization, deployed in production.

Education

Ph.D candidate **Lausanne, Switzerland**
École Polytechnique Fédérale de Lausanne (EPFL), Electrical Engineering (EDEE) *08/2020–10/2024*

- Thesis: *The Use of External Knowledge in Machine Translation*
- Contributed to applied projects with external partners:
 - FamilyMT: Multilingual MT for families of languages and dialects, supported by Armasuisse
 - UniSub: Unsupervised MT with innovative multilingual subword models, supported by Armasuisse
- Supervisor: Prof. [Andrei Popescu-Belis](#)

Master of Science (MSc) in Data Science and Machine Learning **Athens, Greece**
National Technical University of Athens (NTUA) *11/2018–07/2020*

- GPA: 9.22/10 (excellent)
- Thesis: *Adversarial Fine-tuning of Pretrained Language Models*

Erasmus exchange student **Madrid, Spain**
Universidad Carlos III de Madrid (UC3M), School of Electrical Engineering *01/2016–06/2016*

BEng & MEng in Electrical and Computer Engineering **Athens, Greece**
National Technical University of Athens (NTUA) *10/2011–11/2017*

- GPA: 8.02/10 (top 15% of class)
- Thesis: *Machine Learning Algorithms and Data Stream Storage Techniques for Sentiment Analysis*

Invited talks

- Machine Translation reading group, Microsoft / virtual, on-line 02/2024
- [ELLIS & Unbabel Seminars](#), Lisbon ELLIS Unit (LUM LIS) and Unbabel / virtual, on-line 02/2024
- [Workshop on Natural Language Processing](#), Archimedes Research Center / Athens, Greece 07/2023
- [ELLIS & Unbabel Seminars](#), Lisbon ELLIS Unit (LUM LIS) and Unbabel / virtual, on-line 10/2022

Publications

- [1] [Vernikos, G.](#), and Popescu-Belis, A. (2024). *Don't Rank, Combine! Combining Machine Translation Hypotheses Using Quality Estimation*. ACL 2024. [\[paper\]](#) [\[code\]](#)
- [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 (oral). [\[paper\]](#) [\[code\]](#)
- [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). [\[paper\]](#) [\[code\]](#)
- [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). [\[paper\]](#) [\[code\]](#)

Honors and Awards

- **Facebook Scholarship for Attendance at EurNLP 2019**
Recipient of travel grant to attend the EurNLP summit in London, UK.
- **Nationwide University Entrance Examination**
Scored 19.398/20 on the national qualification exams (top 1% nationwide).

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:** L^AT_EX, Git, Flask

Languages

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