

Katerina Margatina

Ph.D. Student in Natural Language Processing

✉ k.margatina@sheffield.ac.uk  <https://katerinamargatina.github.io>
 Twitter  Github  LinkedIn  Google Scholar  Semantic Scholar

Research Interests

active learning, large language models, dynamic adversarial data collection, evaluation & benchmarking, language modeling, robustness, fairness, temporal drift, disagreement, cross-cultural natural language processing

Education

Oct. 2019 – Ongoing	University of Sheffield, UK <i>Ph.D. in Computer Science</i> <ul style="list-style-type: none">Natural Language Processing Group (SheffieldNLP)Supervisors: Prof. Nikolaos AletrasFunding: Amazon Alexa FellowshipThesis: “<i>Active Learning Algorithms for Data Efficient Language Models</i>”Expected Graduation: November 2023
Jan. 2016 – Jul. 2016	Universidad Carlos III de Madrid, Spain <i>MEng in Computer Engineering (Erasmus Mobility in Masters)</i>
Sep. 2012 – Jul. 2019	National Technical University of Athens, Greece <i>BEng & MEng in Electrical and Computer Engineering</i> <ul style="list-style-type: none">GPA: 8.06/10 (top 15%)Major: Electronics and SystemsThesis: “<i>Transfer Learning and Attention-based Conditioning Methods for Natural Language Processing</i>” - supervised by Prof. Alexandros Potamianos

Professional Experience

Aug. 2022 – Mar. 2023	Meta, UK <i>Research Scientist Intern at Meta AI Research (FAIR Labs) in London</i> <ul style="list-style-type: none">Active Learning Principles for In-Context Learning with Large Language Models (paper published at Findings of EMNLP 2023)
Apr. 2022 – Aug. 2022	Amazon Web Services, USA <i>Applied Scientist Intern at Amazon Comprehend in NYC</i> <ul style="list-style-type: none">Dynamic Benchmarking of Masked Language Models on Temporal Concept Drift with Multiple Views (paper published in EACL 2023)
Jun. 2021 – Feb. 2022	University of Copenhagen, Denmark <i>Research Intern / Visiting Ph.D. in Computer Science at CoAStAL</i>
Feb. 2020 - Jun. 2021	University of Sheffield, UK <i>Teaching Assistant at the Department of Computer Science</i> <ul style="list-style-type: none">COM4513/6513 Natural Language Processing (2020 & 2021)COM3110/4115 Text Processing (2020)
Nov. 2018 - Jul. 2019	DeepSea Technologies, Piraeus, Greece <i>Machine Learning Engineer at the AI team</i>

Teaching & Other Academic Experience

- 2021 | **MSc. Dissertation Co-Supervisor**
- Co-advised Ard Snijders, a MSc student from the University of Amsterdam, along with Douwe Kiela on the thesis project *Investigating Multi-source Active Learning for Natural Language Inference* (published in EACL 2023).
 - Co-advised [Atsuki Yamaguchi](#), a MSc student from the University of Sheffield, along with George Chrysostomou and Nikolaos Aletras on the project *Frustratingly Simple Alternatives to Masked Language Modeling* (published in EMNLP 2021).
- 2019 - Ongoing | **Teaching Assistant**
- Lead lab sessions of the Text Processing and Natural Language Processing modules.
 - Marked lab reports following a pre-determined scheme set by the Course Leader.
 - Experience in face-to-face teaching with students that needed extra guidance.
- 2018 - Ongoing | **Program Committee**
- Reviewer for ACL, EMNLP, NAACL, EACL, AACL, WASSA, NLLP, DADC, ARR
 - Co-organizer of the First Workshop on Dynamic Adversarial Data Collection at NAACL 2022 ([website](#))

Invited Talks & Presentations

- Jul. 2023 | Archimedes Summer School ([slides](#))/ in-person in Athens
- Jun. 2023 | Meta AI ([slides](#))/ in-person in London
- Jun. 2022 | Bloomberg, AI Group ([slides](#)) / in-person in NYC
- Apr. 2022 | [NLPhD Speaker Series](#), University of Saarland ([slides](#)) / remote

Honors & Awards

- 2023 | **Microsoft:** Awarded a grant of \$20,000 in credits from Accelerate Foundation Models Research for our proposal: *“The PERDI Project: PERsonalised and Diverse Feedback for Humans-and-LLMs-in-the-loop”* (with Hannah Rose Kirk, Paul Röttger, Bertie Vidgen and Scott Hale).
Archimedes: Awarded 500€ from the [Archimedes](#) Center for Research in Artificial Intelligence, Data Science and Algorithms to do an in-person presentation at the Archimedes Summer School in July 2023.
- 2022 | **Dynabench:** Awarded a research grant of \$46,825 from Meta for our [proposal](#): *“Adversarial NERDs: Optimising feedback between humans-and-model-in-the-loop”* (with Hannah Rose Kirk and Scott Hale).
- 2020 | **EMNLP:** Awarded a scholarship to participate to the virtual Conference on Empirical Methods in Natural Language Processing ([EMNLP 2020](#)).
MLSS: Selected to participate to the highly competitive Machine Learning Summer School ([MLSS](#)) organized by the Max Planck Institute for Intelligent Systems (180 out of 1300+ applications were accepted).
- 2019 | **Amazon Alexa Fellowship:** Awarded an [Amazon Alexa Fellowship](#) for a 3-year Ph.D. program at the University of Sheffield under the supervision of Dr. Nikolaos Aletras.
EETN Scholarship Award: Awarded a scholarship of from the Hellenic Artificial Intelligence Society (EETN) to attend the first Athens Natural Language Processing Summer School ([AthNLP](#)).
ACM-W Scholarship Award: Awarded an [ACM-W scholarship](#) of \$600 to attend the Annual Meeting of the Association for Computational Linguistics in Florence ([ACL 2019](#)).
- 2018 | **WASSA IEST:** Ranked 3rd out of 30 participants at the WASSA 2018 Implicit Emotion Shared Task ([IEST](#)) conducted by the [EMNLP 2018](#) Conference.

Publications

- 2023
1. Alajrami, A., **Margatina, K.** & Aletras, N. Understanding the Role of Input Token Characters in Language Models: How Does Information Loss Affect Performance? *EMNLP* (2023).
 2. **Margatina, K.** & Aletras, N. *On the Limitations of Simulating Active Learning in Findings of ACL* (2023).
 3. **Margatina, K.**, Schick, T., Aletras, N. & Dwivedi-Yu, J. Active Learning Principles for In-Context Learning with Large Language Models. *Findings of EMNLP* (2023).
 4. **Margatina, K.** et al. *Dynamic Benchmarking of Masked Language Models on Temporal Concept Drift with Multiple Views in EACL* (2023).
 5. Snijders, A., Kiela, D. & **Margatina, K.** *Investigating Multi-source Active Learning for Natural Language Inference in EACL* (2023).
- 2022
6. Hershcovich, D. et al. *Challenges and Strategies in Cross-Cultural NLP in ACL* (2022).
 7. **Margatina, K.**, Barrault, L. & Aletras, N. *On the Importance of Effectively Adapting Pretrained Language Models for Active Learning in ACL* (2022).
- 2021
8. **Margatina, K.**, Vernikos, G., Barrault, L. & Aletras, N. *Active Learning by Acquiring Contrastive Examples in EMNLP* (2021).
 9. Yamaguchi, A., Chrysostomou, G., **Margatina, K.** & Aletras, N. *Frustratingly Simple Alternatives to Masked Language Modeling in EMNLP* (2021).
- 2020
10. Vernikos, G., **Margatina, K.**, Chronopoulou, A. & Androutsopoulos, I. *Domain Adversarial Fine-Tuning as an Effective Regularizer in Findings of EMNLP* (2020).
- 2019
11. **Margatina, K.** *Transfer Learning and Attention-based Conditioning Methods for Natural Language Processing. Thesis, NTUA* (2019).
 12. **Margatina, K.**, Baziotis, C. & Potamianos, A. *Attention-based Conditioning Methods for External Knowledge Integration in ACL* (2019).
- 2018
13. Chronopoulou*, A., Margatina*, K., Baziotis, C. & Potamianos, A. *NTUA-SLP at IEST 2018: Ensemble of Neural Transfer Methods for Implicit Emotion Classification in Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis* (2018).

Coding

Expert: Python, Bash / PyTorch, HuggingFace transformers

Familiar with: C, C++, Matlab, SQL, Java / Tensorflow, Keras

Other: Git, ~~TeX~~, Unix, Flask, AWS

Languages

Greek (*Native*), English (*C2*), Spanish (*B2*), French (*B1*)