KIRILL MILINTSEVICH

Researcher (NLP / Medical AI)

(+33) 7 68 33 44 91 ◇ milintsevich@gmail.com

Caen, France

https://scholar.google.com/citations?user=BQNVCjYAAAAJ&hl=en ◇ https://github.com/501Good

EDUCATION AND RESEARCH EXPERIENCE

University of Caen & University of Tartu (joint program) Tartu, Estonia & Caen, France

PhD in Computer Science *Oct 2020 – Oct 2024*

Estimation of Depression Level from Text: Symptom-Based Approach, External Knowledge, Dataset Validity. A collaborative effort between the University of Caen Normandy, the University of Tartu, and the Hospital-University Federation (FHU) of Amiens, Caen, Rouen, and Lille. During this interdisciplinary project, we developed a state-of-the-art neural model for symptom-based depression prediction from text (clinical conversations), proposed an approach for integrating external knowledge into transformer-based pre-trained language models, and produced a social-media-based dataset for anhedonia detection. Five papers have been published as part of this research.

University of Tartu Tartu, Estonia

MSc in Computer Science *Sep 2018 – Jun 2020*

I was admitted as a tuition waiver student and received an achievement and specialization stipend (IT Academy). I was involved in a state-funded research project called Neural Network Based Text Analysis Models for Estonian. During this project, we developed a novel neural architecture for lemmatization in 23 languages. Additionally, we worked on evaluating multilingual pre-trained language models for Estonian. Three papers have been published as a result of this research.

Higher School of Economics Moscow, Russia

MA in Applied Linguistics (Computational Linguistics) *Sep 2016 – Jun 2018*

I was admitted as a tuition waiver student and received an achievement stipend. My thesis was on automatic code-switching detection for minor languages of Russia. Additionally, I worked on a research project on automatic word stress detection in Russian, during which we collaborated with the Russian National Corpus and trained a neural model. Two papers have been published during this research.

Far Eastern Federal University Vladivostok, Russia

BA (Hons) in Applied Linguistics *Sep 2012 – Jun 2016*

TEACHING EXPERIENCE

Teaching and Research Fellow (ATER) *Sep 2023 – Aug 2024*

Taught the following courses in French at the University of Caen Normandy, France

- Web Development, HTML & CSS (Bachelor level, 40 hours of practice sessions)
- Introduction to Programming in Python (Bachelor level, 20 hours of practice sessions)
- Introduction to OOP in Python (Bachelor level, 90 hours of practice sessions)
- Introduction to NLP (Master level, 6 hours of practice sessions)
- Advanced NLP (Master level, 4 hours of lectures on autoregressive generative models)
- Introduction to Pytorch (PhD level, 2 hours of lectures, 8 hours of practice sessions)

Teaching Assistant *Feb 2020 – Jun 2022*

Taught the following courses in English at the University of Tartu, Estonia

- Natural Language Processing, spring 2022 (Master and Ph.D. level, 64 hours of practice sessions)
- Natural Language Processing, spring 2021 (Master and Ph.D. level, 48 hours of practice sessions)
- Natural Language Processing, spring 2020 (Master and Ph.D. level, 24 hours of practice sessions)

COMMUNICATIONS

Posters

- **Automatic Text-Based Estimation of Depression Symptoms**
  15ème Congrès Français de Psychiatrie (CFP 2023), Lyon, France
- **Analyse de l'estimation du niveau de dépression basée sur les symptômes à travers le prisme de l'expertise des psychiatres**
  15ème Congrès Français de Psychiatrie (CFP 2023), Lyon, France
- **Analyzing Symptom-based Depression Level Estimation through the Prism of Psychiatrist' Expertise**
  Journée GdR TAL "Modèles de langue aux domaines de spécialité" (2023), Nantes, France
- **Automatic Text-based Estimation of Depression Symptoms**
  19th Estonian Summer School on Computer and Systems Science (ESSCaSS 2022), Tartu, Estonia

Invited Talks

- **Introducing External Knowledge for Automatic Depression Estimation from Text**
  University of Tartu Machine Learning Summer Seminar (2023), Roosta, Estonia
- **Introduction of External Knowledge into Deep Learning Models for the Diagnosis of Symptoms of Depression**
  2ème journée FHU A^2M^2P (2023), Caen, France
- **Towards Automatic Text-Based Estimation of Depression through Symptom Prediction**
  Seminar at the University of Zürich (2022), Switzerland (invitation of Prof. Martin Volk)
- **Natural Language Processing for Mental Health**
  Invited lecture for bachelor's students at the Higher School of Economics (2022), Moscow, Russia

Participation in Educational Events

- **Symposium en Santé Mentale et Intelligence Artificielle (MentalAI 2024)**
- **20th Estonian Summer School on Computer and Systems Science (ESSCaSS 2023)**
- **1ère École Saisonnière en Intelligence Artificielle sur la Santé (ESIA 2023)**
- **Advanced Language Processing Winter School (ALPS 2022)**
- **AI4Health Winter School (AI4Health 2021)**

COMPETITIONS

Clinical NLP 2023 Shared Task on Summarization of Medical Conversations
Fine-tuned LongT5 model in a multi-task setting, specifically for the generation of both short and long clinical notes. A paper has been published at the Clinical NLP 2023 workshop.
Hugging Face Whisper Fine-Tuning Event (2022)
Fine-tuned Whisper speech-to-text models for the Tatar language.

ORGANIZATIONAL ACTIVITIES

Program Committee: Clinical NLP 2023
Master Thesis Committee: Koljal, K. (2022). Predicting Depression Symptoms Based on Reddit Posts [Master's Thesis, University of Tartu].

SUPERVISION

University of Caen Normandy, Master Project
Throughout this one-year project entitled "Semi-Supervised Extraction of Depression Symptom Markers from Text," I have supervised four master-level students. The students successfully learned how to use question-answering and large language models in Python.

INDUSTRY EXPERIENCE

TransPerfect, Remote (freelance) *Summer 2021 · Summer 2022*
Developing the inverse text normalization system for the automatic speech recognition pipeline.
Medialogia, Moscow (hybrid, full-time) *Aug 2017 – Oct 2020*
NLP Engineer (Sentiment Analysis & Information Extraction).
Tolstoy Digital, Remote (freelance) *Oct 2018 – Oct 2019*
Developing the pipeline for automatic conversion of Tolstoy's texts into TEI format.

LANGUAGES

French ◇ *Advanced (C1)* **English** ◇ *Advanced (C1)* **Russian** ◇ *Native*

RESEARCH PUBLICATIONS

1. **Milintsevich, K.**, Sirts, K., & Dias, G. (2024). Evaluating Lexicon Incorporation for Depression Symptom Estimation. In *Proceedings of the 6th Clinical Natural Language Processing Workshop (**Clinical NLP**)* at *NAACL 2024*.
   </> Code: `https://github.com/501Good/dialogue-classifier`
2. Agarwal, N.*, **Milintsevich, K.***, Métivier, L., Rothärmel, M., Dias, G., & Dollfus, S. (2024). Analyzing Symptom-based Depression Level Estimation through the Prism of Psychiatric Expertise. In *Proceedings of the Joint International Conference on Computational Linguistics, Language Resources and Evaluation (**LREC-COLING 2024**)*.
   **equal contributions*
3. **Milintsevich, K.**, Sirts, K., & Dias, G. (2024). Your Model Is Not Predicting Depression Well And That Is Why: A Case Study of PRIMATE Dataset. In *Proceedings of the 9th Workshop on Computational Linguistics and Clinical Psychology (**CLPsych**)* at *EACL 2024*.
4. **Milintsevich, K.** & Agarwal, N. (2023). Calvados at MEDIQA-Chat 2023: Improving Clinical Note Generation with Multi-Task Instruction Fine-tuning. In *Proceedings of the 5th Clinical Natural Language Processing Workshop (**Clinical NLP**)* at *ACL 2023*.
   </> Code: `https://github.com/501Good/MEDIQA-Chat-2023-Calvados`
5. **Milintsevich, K.**, Sirts, K., & Dias, G. (2023). Towards automatic text-based estimation of depression through symptom prediction. ***Brain Informatics**, 10(1), 1-14.*
   </> Code: `https://tinyurl.com/3ssw4rcf`
6. **Milintsevich, K.** and Sirts K. (2021). Enhancing Sequence-to-Sequence Neural Lemmatization with External Resources. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume* at ***EACL 2021***. Association for Computational Linguistics.
   </> Code: `https://github.com/501Good/lexicon-enhanced-lemmatization`
7. **Milintsevich, K.** and Sirts K. (2020). Lexicon-Enhanced Neural Lemmatization for Estonian. In *Proceedings of the Ninth International Conference of **Baltic HLT 2020***. Frontiers in Artificial Intelligence and Applications.
8. Kittask, C., **Milintsevich, K.** & Sirts K. (2020). Evaluating Multilingual BERT for Estonian. In *Proceedings of the Ninth International Conference of **Baltic HLT 2020***. Frontiers in Artificial Intelligence and Applications.
9. Chernyak, E., Ponomareva, M., & **Milintsevich, K.** (2019). Char-RNN for Word Stress Detection in East Slavic Languages. In *Proceedings of the Sixth Workshop on NLP for Similar Languages, Varieties and Dialects (**VarDial**)* at *NAACL-HLT 2019*. Association for Computational Linguistics.
10. Ponomareva, M., **Milintsevich, K.**, Chernyak, E., & Starostin, A. (2017). Automated word stress detection in Russian. *In Proceedings of the First Workshop on Subword and Character Level Models in NLP (**SCLeM**)* at *EMNLP 2017*. Association for Computational Linguistics.

LIST OF REFEREES

- **Prof. Gaël Dias** (PhD Supervisor) — University of Caen Normandy (France), CNRS GREYC UMR 6072 — gael.dias@unicaen.fr
- **Prof. Kairit Sirts** (PhD Supervisor) — University of Tartu (Estonia), Institute of Computer Science — kairit.sirts@ut.ee
- **Prof. Sonia Dollfus** (PhD Co-supervisor) — University of Caen Normandy, CHU of Caen (France) — dollfus@cyceron.fr
- **Dr. Ekaterina Artemova** (Master's Supervisor, Co-author) — Toloka.AI — ekaterina.l.artemova@gmail.com