Hai Le

hai.le@skol.tech
https://github.com/highly0

INTEREST   Natural Language Processing, Large Language Models, Explainable AI, Computer-Aided Education

EDUCATION   **Skolkovo Institute of Science and Technology**   Moscow, Russia
M.S. Mathematics and Computer Science   Sep. 2021– Jun. 2023
Supervised by Professor Alexander Panchenko and Mikhail Salnikov.
- Distinction/Cum Laude (GPA: 4.85/5.0)
- Thesis: Knowledge Graphs Question Answering: Sequence to Sequence (Seq2Seq) Model with Redirects, Subgraphs Extraction & Re-ranking.
- Relevant Courses: Machine Learning, Deep Learning, Deep Learning for Natural Language Processing, Numerical Linear Algebra, Introduction to Recommender System, Neuroimaging and Machine Learning for Biomedicine, Introduction to Blockchain.

University of Maryland, College Park   Maryland, USA
B.S. Computer Science   Aug. 2017– May 2021
- Relevant Courses: Computer Network & Security Computer Algorithms, Object-Oriented Programming, Cryptography, Programming Handheld Systems, Advanced Data Structures, Organization of Programming Languages, Game Programming.

HONORS AND
AWARD
Skolkovo Institute of Science and Technology Fellowship, 2021-2023.
Academic Excellence Scholarship, Skolkovo Institute of Science and Technology, 2022-2023.
Industrial Immersion Excellence award, 2022
Outstanding project for MIT's Global Startup Lab, 2023.
Dean's List, University of Maryland, 2019-2021.

PUBLICATION   Mikhail Salnikov[*], Hai Le[*], Prateek Rajput, Irina Nikishina, Pavel Braslavski, Valentin Malykh and Alexander Panchenko. **Large Language Models Meet Knowledge Graphs to Answer Factoid Questions.** *In proceedings of Pacific Asia Conference on Language, Information and Computation* (PACLIC 2023).

RESEARCH
EXPERIENCE
Center for Artificial Intelligence Technologies, Huawei & Skoltech   Moscow, Russia
Research Engineer (Advisor: Professor Alexander Panchenko)   Jul. 2022 – Current
- Read and gather ideas from prior state-of-the-art research papers on top level NLP conferences and journals.
- Research and develop a Proof of Concept pipeline for Question Answering (QA).
- Conduct experiments on different QA datasets, including English and multilingual.
- Write and publish experiment results on reputable conferences.

Samsung AI Center   Moscow, Russia
Research Intern (Advisor: Dr. Alexander Limonov)   Jun. 2022 – Aug. 2022
- Worked on indoor positioning tasks for robot navigation and 3D models of indoor environment.
- With stationary ultrasound beacons responsible for mapping coordinates via triangulation, implemented an optimization algorithm that can minimize the error discrepancies from 0.179m to 0.81m.
- Won Industrial Immersion Excellence award.

TEACHING   **Introduction to Data Science**, University of Maryland   College Park, USA
Teaching Assistant   Aug. 2020 – May. 2021
- CMSC320 in the Department of Computer Science.

- Covered basic data science concepts such as scraping, cleaning, regression & classification techniques in R.
- Held office hours every week, answered questions in person and online, marked assignments, and proctored exams.

Introduction to Object-Oriented Programming, University of Maryland   College Park, USA
Teaching Assistant   Jan. 2021 – May. 2021
- CMSC131 and CMSC132 in the Department of Computer Science.
- Covered basic object-oriented concepts (polymorphism, inheritance) and basic data structures (linked list, graphs, queues, stacks, hashed-maps, and others).
- Ran discussion for over 40 students twice a week, had office hours every week, answered questions in person and online, marked assignments, and proctored exams.

Geeklama   California, USA (Remote)
Computer Science Instructor   Feb. 2022 – Feb. 2023
- Taught students Introduction to Python Programming and rudimentary computer science concepts/ideas such as expressions, statements, methods, conditionals, and others.
- Collaborated with fellow teachers to develop the lesson plans and coursework such as assignments and projects.

PROJECTS

Image Caption Generator Case Study
Deep Learning & Deep Learning for Natural Language Processing   Skoltech, 2022
- Implemented and benchmarked the image captioning performance of several models in the encoder-decoder framework with backbones such as convolution neural network (`VGG16 Densenet161`, `InceptionV`), recurrent neural network (`LSTM`, `GRU`) and transformers (`DiET`, `ViT`) [Code]

Image Restoration with SwinIR
Machine Learning   Skoltech, 2022
- Re-implemented `SwinIR` for image restoration using Transformers by adding Gaussian noise compared performance using `PSNR` and `SSIM` metrics. [Code]

Tensor Decomposition Case Study
Numerical Linear Algebra   Skoltech, 2023
- Implemented Canonical Polyadic decomposition (CPD) and Tucker Decomposition. Examined performance of classical CNN architectures such as `Resnet` and `Densenet` with the addition of the tensor decomposition. [Code]

Case Study of Hybrid Recommender Architecture
Introduction to Recommender System   Skoltech, 2023
- Implemented and compared the performance of a hybrid model - `DeepFM` (utilizing both low and high order user-item interaction) against several baselines - `SVD`, `LightFM`, `DSSM`. [Code]

COMPETENCES

Programming Languages Python, C/C++, Java, R, GraphDB, MATLAB, LaTeX.
Programming Libraries & Tools Pytorch, Hugging Face, Transformers, Tensorflow, Keras, Scikit-learn, Scipy, NumPy, Git, Docker.
Languages English (*native*), Vietnamese (*native*)

REFERENCES

Alexander Panchenko
Associate Professor, Head of Skoltech NLP Group   A.Panchenko@skol.tech
Center for Artificial Intelligence Technologies,   +7-916-370-0376
Skolkovo Institute of Science and Technology   Google Scholar

Mikhail Salnikov
Research Engineer, Skoltech NLP Group   M.Salnikov@skol.tech
Center for Artificial Intelligence Technologies,   +7-927-172-3059
Skolkovo Institute of Science and Technology   Google Scholar