Ilja **Barouski**

Machine Learning/Software Engineer

Žodzina, Belarus

+375-44-450-69-51 | iljabarouski@gmail.com | scurrra.github.io | Scurrra | iljabarouski | iljabarouski

Education

BSU (Belarusian State University), Faculty of Mechanics and Mathematics, Computer Mathematics and System Analysis — *Minsk, Belarus*

Bachelor of Mathematics and System Analysis — *Sep. 2019 - Jun. 2023*

BSU (Belarusian State University), Faculty of Mechanics and Mathematics, Computer Mathematics and System Analysis — *Minsk, Belarus*

Master of Mathematics and Computer Science — *Sep. 2023 - Jun. 2025*

Skills

Programming	Python, Julia, Rust, C, C++, Crystal, C#, Dart, Go, LaTeX, Wolfram
Databases	PostgreSQL, MongoDB, SurrealDB
Machine Learning	Data Visualization, Regression, Decision Trees, Neural Networks, Ensembles, Tokenization, NLP
Languages	English (B2-C1), Bielaruskaja (Native), Русский

Pet Projects

UBPE: Universal Byte-Pair Encoding Tokenizer

github repo, pypi, google colab

Stack: Python, C++ — *Jun. 2025 - Present*

A flexible and efficient tokenizer for general sequences, extending traditional BPE beyond text.
Key Features:

- *General-Purpose Tokenization*: Works with text, numerical sequences, or any custom alphabet.
- *Multiple Encodings*: Generates multiple encoding variants for a single sequence, allowing users to choose the best representation based on TF-IDF scores.
- *Optimized Encoding*: Supports batch substitution of token pairs and IDF-based token weighting for shorter, more efficient encodings.
- *High-Performance Backends*: Includes Python-native and Cython (C++20) backends for fast execution and low overhead.
- *PyPI Integration*: Easy to install and integrate into Python projects (‘pip install ubpe[native]‘ or ‘pip install ubpe[cython]‘).

Wavelet Neural Networks

github repo, pypi

Stack: Python (PyTorch) — *Feb. 2025 - Present*

A PyTorch library for wavelet transforms in neural networks.
Key Features:

- *Wavelet Integration*: Supports 1D and 2D wavelet transforms (e.g., Haar, Daubechies) directly within PyTorch tensors.
- *Seamless PyTorch Integration*: Compatible with PyTorch’s autograd system, enabling end-to-end training of wavelet-augmented models.
- *Applications*: Ideal for signal processing, image analysis, and time-series data augmentation.
- *Efficiency*: Optimized for GPU acceleration, leveraging PyTorch’s tensor operations.
- *Modular Design*: Easy to integrate into existing neural network architectures.

Corgi.jl

github repo

Stack: Julia — *Jun. 2022 - Present*

Machine Learning Library from scratch.

Undergraduate practice in NLP

github repo, colab notebook

Stack: Python — *Feb. 2023 - Mar. 2023*

(Mostly) CNN News analysis, including Named Entity Recognition using flair, Sentiment Analysis using TimeLMs (RoBERTa, tuned on tweets), and Extractive (textrank algorithm) and Abstractive (PEGASUS and T5) Summarization. The results could be seen in the provided colab demo.