YEHOR MISHCHYRIAK

ymishchyriak@wesleyan.edu|860-834-9313|LinkedIn|ymishchyriak.com
Address: 45 Wyllys Ave, Middletown, CT, US

EDUCATION

Wesleyan University, Middletown CT May 2026
BA in Computer Science and Mathematics (double-major)
Cumulative GPA: 3.64/4.00; Mathematics GPA: 3.55/4.00; Computer Science GPA: 3.53/4.00.
Dean's List: Fall 2024, Fall 2023
RELEVANT COURSEWORK:

- **Mathematics:** differential, integral, and multivariable calculus, statistics, probability, discrete mathematics, linear algebra, applied linear algebra, abstract algebra, real analysis, differential equations, combinatorics
- **Computer Science:** foundations of computer science I & II, algorithms and complexity, computer networks, artificial intelligence, audio-visual machine learning, SQL and databases, HPC, quantitative textual analysis, automata theory, program analysis

RESEARCH EXPERIENCE

Bonham Lab, Tufts University School of Medicine – Boston, MA *(supervisor email)*
Student Research Intern (remote, part-time) *November 2025 – Present*

- Designing, training, and evaluating ML models for functional annotation of proteins encoded by bacterial genes in the gut microbiome.
- Developing a Graph Attention Network on a protein co-occurrence graph derived from metagenomic profiles for semi-supervised function classification, using the dataset compiled during my prior summer internship.

Summer Research Intern (on-site, full-time) *June – July 2025*

- Built an end-to-end ML data engineering pipeline for protein functional annotation (data mining, cleaning, feature engineering, persistence).
- Encoded categorical variables as multi-hot vectors with dimensionality reduction; embedded amino acid sequences (ESM2 protein language model) and free text (OpenAI SDK with caching).
- Implemented Zarr-based persistence; packaged into a documented Python library named M2F (microbiome-to-function).
- Processed ~1k microbiome metagenomic profiles output by HUMAnN, yielding ~900k protein entries with numerically encoded UniProt annotations, compressed for ML.

Main tools: NumPy, Pandas, PyTorch, PyG, Hugging Face, OpenAI, Zarr, Tufts HPCC

Thayer Lab, Wesleyan University – Middletown, CT *(supervisor email)*
Research Assistant (hybrid, part-time) *September 2024 – Present*

- Rebuilt and optimized residue interaction network (RIN) construction and random-walk sampling pipelines developed during my summer fellowship, adding modular design, logging, error handling, shared-memory parallelization, and careful memory management; added RIN visualization and MD-trajectory animation with granular control over residues, interaction thresholds/types, and visual settings.
- Developed an embedding pipeline (PyTorch + my NumPy-based framework PureML) using DeepWalk-style random walks with interaction-weighted transition probabilities and Skip-Gram training to learn context-rich amino acid embeddings, including PCA projection and visualization; packaged everything into the SAWNERGY Python library with tests and CI/CD.
- Applying SAWNERGY to analysis of p53 tumor suppressor protein and its isoforms to discover effects of mutations and ligand binding on the protein.

Research in Science Fellow (on-site, full-time) *June – July 2024*

- Studied biophysics, molecular biology, Unix-based systems, and molecular dynamics simulations.
- Developed the initial pipeline to construct RINs from MD trajectories and sample random walks, enabling amino acid co-occurrence analysis and comparison between systems (e.g., mutant vs. wild-type) to characterize mutation- and effector-induced changes in allosteric signaling.

Main tools: NumPy, Matplotlib, Zarr, PyTorch, PyG, PureML, Wesleyan HPCC

PUBLICATIONS

- **SAWNERGY:** A Python framework for dynamic residue-interaction networks and walk-based embeddings from molecular dynamics simulations – submitted to the Journal of Open Source Software (pre-print at ymishchyriak.com/papers); deployed at PyPI.
- **PureML:** a transparent NumPy-only deep learning framework for teaching and prototyping – submitted to the Journal of Open Source Software (pre-print at ymishchyriak.com/papers); deployed at PyPI.
- Manuscript on the analysis of 12 p53 isoforms, which I am co-authoring, is currently in preparation.

AWARDS

- **Wesleyan University Summer Grant** ($5,000), 2025 — competitive grant supporting career-related summer experience
- **State Gold Medal for Academic Excellence**, Ukraine (Secondary Education), 2022

PRESENTATIONS

- **Wesleyan Summer Symposium**, 2024 – Presented initial pipeline for constructing RINs from MD trajectories and sampling random walks.
- **Annual Biophysics Retreat (Wesleyan Biophysics Program)**, 2024, 2025 – Presented work on residue interaction networks in the context of allosteric signaling (2024); presented SAWNERGY and its applications to downstream machine learning and p53 isoform analysis (2025).
- **MB&B209 Guest Lecture**, 2025 – Delivered a lecture on allosteric regulation in proteins and Thayer Lab's research in this context.
- **ACS Northeast Regional Discussion**, 2025 – Presented SAWNERGY and its applications to downstream machine learning and p53 isoform analysis.

PROJECTS

PureML: NumPy-based deep learning library | Personal Project
Creator and Maintainer *May 2025 – Present*

- Built a lightweight deep learning framework in NumPy with a graph-based reverse-mode autograd engine, Tensor API, modular layers (Affine/Linear, Embedding, BatchNorm, Dropout), activation functions, loss functions, and batched training utilities.
- Informed the framework's design and implementation by systematically working through the first 12 chapters of Deep Learning by Goodfellow, Bengio, and Courville, translating key theoretical concepts into reproducible code.
- Implemented optimizers (SGD, AdaGrad, RMSProp, Adam) with learning-rate schedulers, deterministic seeding utilities, and full checkpointing of model, optimizer, and scheduler state via compressed Zarr-based archives for reproducible experiments.
- Achieved ~98% test accuracy on MNIST with a small fully connected network implemented entirely in PureML, matching classic PyTorch baselines and validating correctness and numerical stability of the framework.
- Packaged and released PureML as an open-source Python library with logging, tests, and documentation; actively extending the framework and using it as a backend in other research codebases (like SAWNERGY).

Main tools: NumPy, Zarr

eQoScan: WesHack – Wesleyan University Hackathon | Personal Project
Developer *November 2024*

- Built a mobile/web app for Wesleyan's reusable food container system with real-time tracking (holder details, pick-up and drop-off times, locations) and automated return reminders.
- Integrated dynamic QR codes for seamless dining hall transactions.
- Developed in 24 hours; selected as a winner among 70+ U.S. university participants.

Main tools: Flask, Firebase, JS/HTML/CSS/Bootstrap, Swift (iOS)

ADDITIONAL WORK EXPERIENCE

Hazel QAC, Dep. of Math and CS: Wesleyan University – Middletown, CT *(supervisor emails: QAC, MATH/CS)*
Applied vectors and matrices teaching assistant (on-site, part-time) *January – May 2025*

- Led weekly help sessions for the class of 30 students. Assisted during in-class laboratory sessions.

Differential calculus teaching assistant (on-site, part-time) *January – May 2024*

- Led weekly help sessions for the class of 25 students. Graded HW assignments.

Instructional Media Services, Wesleyan University – Middletown, CT *(supervisor email)*
On-campus events assistant manager (on-site, part-time) *December 2023 – May 2024*

- Set up and managed AV equipment for on-campus events across multiple venues attended by dozens.
- Managed on average 10-12 events per week.
- Composed weekly schedules for 18 student workers.
- Provided support and guidance for student workers.
- Enhanced ability to respond quickly to challenges given time constraints.

Classroom support technician (on-site, part-time) *September 2022 – May 2024*

- Provided faculty and students with live technical support.
- Performed maintenance on electronic equipment, such as computers, multimedia projectors, screens, network gear, microphones, speakers, cameras, etc.

SKILLS / KNOWLEDGE

- **Programming languages:** Python, C, SML
- **Python ecosystem:** NumPy, Pandas, Matplotlib, PyTorch, PyG, Zarr, Numba, Cython, Flask
- **HPC:** Unix systems, Slurm, Bash, parallelization (threading/multiprocessing, IPC), memory management
- **Tools:** Git, SQL
- **Languages:** English (near-native), Ukrainian/Russian (native), German (elementary)