Neerav Kaushal

kaushalneerav@gmail.com | 906-370-6376 | Github | Google Scholar | LinkedIn

EDUCATION

PhD, Engineering Physics, *Michigan Technological University* 2018 – 2022 | Houghton, MI, USA
Thesis - *Toward Deep Learning Emulators for Modeling the Large-Scale Structure of the Universe* (LINK)

MS, Physics, *Panjab University* 2015 – 2017 | Chandigarh, India

BS, Physical Sciences, *Himachal Pradesh University* 2012 – 2015 | Shimla, India

INDUSTRY EXPERIENCE

Deep Learning Scientist II, Drug Discovery, Jul 2024 – present | Cambridge, MA
Flagship Pioneering (3 months)

- Built and deployed an NLP generative AI model for small molecule drug design, capable of generating **1 million** novel drug candidates in **under 4 minutes**, outperforming SOTA industry model in all cheminformatics metrics
- Executed **3** successful deep learning projects as on drug design & delivery, nanoparticle AI programming, and RNA biology. Implemented project proposal and AI budget logistics as company **ML representative** at 2 industry collaborations

Deep Learning Scientist I, Drug Discovery, Jan 2023 – Jun 2024 | Cambridge, MA
Flagship Pioneering (17 months)

- Solely developed an in-house generative NLP model for **93%** functionally **optimized** circular RNA components expanding the therapeutic space of RNA
- Designed a pipeline of NLP models to identify and isolate novel RNA-targeting small molecule drugs achieving more than **1900% hit rate** compared to baseline models
- Executed project planning, deliverables identification, resource allocation, and timeline estimation of AIML projects for Q3 2023 and Q1-Q3 2024

ACADEMIC EXPERIENCE

Graduate Researcher, *Michigan Tech University (4 years, 3 months)*

- Built a deep Convolutional Neural Network (**DCNN**) on **5TB** 3D data to model the large-scale structure of Universe (supported in part by **NASA** award no. NNX15AJ20H, Michigan Space Grant Consortium) generating dark matter simulations with **99.3% accuracy** achieving **>100x** runtime **speedup** than baseline simulations
- Designed python programs to streamline **410 GB** of NASA satellite sensor data, **decreasing** program run time from **60 days to 2 days** by regridding low resolution data, using IDW Interpolation at 500m resolution, and temporal upsampling to 3 Hz range
- Streamlined ML **anomaly detection** algorithms in **150-featured** galaxy catalogs using Isolation Forests (IFs), Elliptic Envelopes, UMAPs, DBSCAN, Spectral Clustering, CNNs and VAEs to identify **856 anomalies**
- Developed a novel algorithm in C++, Fortran and Python to reconstruct muon trajectories in HAWC-like Water Cherenkov Detectors with **less than 1%** reconstruction error

SKILLS

ML/DL (5 years) — Generative ML, Discriminative ML, Computer Vision, Natural Language Processing (NLP), Large Language Models (LLM), Geometric Deep Learning, Transfer learning, Ensemble Learning, Representation Learning, Dimensionality Reduction (VAE, cVAE), Outlier/Anomaly Detection, Bayesian Neural Nets, Graph Neural Nets (GNN), Generative Adversarial Nets (GAN), Convolutional Neural Nets (CNN), Diffusion models (DDPMs), Transformers

Data Science (5 Years) — Exploratory Data Analysis (EDA), Data Fusion, High Performance Computing (HPC), Cloud Computing, Version Control Systems (VCS), Algorithm Development, Statistics

Programming & Tools (5 Years) — LANGUAGES: Python, C++, Bash | ML: Pytorch, Scikit-learn, Keras, H2O, W&B, HuggingFace | SOFTWARE: Git, Amazon Web Services (AWS S2, EC2, Sagemaker) Docker | DRUG DISCOVERY: RDKit, DeepChem, Chemprop, MolFormer, MolFlux

ML Field Knowledge — Drug Discovery, Cheminformatics, Small Molecules, Genomics, RNA/DNA Therapeutics, Animal Studies Data, Proteomics

Soft — Excellent Written and Oral Communication, Audience-oriented Presentations, Interpersonal & Analytical Skills, Highly Collaborative

ACADEMIC ACHIEVEMENTS

Publications

- 9 peer-reviewed publications (*Google Scholar Link*)

Conferences

- NeurIPS 2021: A Quasi-Universal Neural Network to Model Structure Formation in the Universe (Kaushal et al. ML4PS, 2021)
- ML X Cosmology (Flatiron Institute): Learn the Universe: Using CNNs to map from approximate to full N-body simulations (N. Kaushal et al., 2021)
- 240th AAS (American Astronomical Society) Meeting: Towards a Universal Field-level Deep Learning Cosmological Emulator (Kaushal et al. AAS, June 2022)
- AGU (American Geophysical Union) Fall Meeting 2019: Global observations of the diffuse radiation fertilization effect through the OCO-2 solar-induced chlorophyll florescence product (Xi et al., AGU 2019)

ML Reviewing

- **Reviewed 12 ML manuscripts**: Invited peer review panelist for NASA ROSES/APRA for the Software and Data Analysis proposals, reviewer for Nature Machine Intelligence, The Journal of Supercomputing, The Astrophysical Journal, Machine Learning: Science and Technology, and NeurIPS 2023 and 2022 conference

VOLUNTEER EXPERIENCE

- **Life Science Cares** (*Boston*): As part of a 5-person team, assembled 400 hygiene care kits for the Boston Health Care for the Homeless Program to ensure dignified access to high-quality healthcare for individuals experiencing homelessness in the downtown Boston community.
- **The Charles River Conservancy (CRC)**: Performed litter removal, raking leaves/sticks, and invasive plant management on spring cleanup projects in Cannalonga Park, Watertown, MA.