Abhilash Neog

Website LinkedIn Scholar Github abhilash22@vt.edu

Research Interests

Scientific Foundation Models | LLMs | Multi-modal models | Physics-Informed ML | Time-Series Modeling

My research interests broadly lie in Scientific Foundation Models, time-series modeling (sparse data, reasoning), LLMs (for continuous data, knowledge distillation), Vision-Language Models (zero-shot evaluation, scientific applications), Physics-informed (or Knowledge-guided) Machine Learning.

Education

Virginia Tech — **Aug 2022 - May 2027**
Ph.D., Computer Science. Advisor: *Dr. Anuj Karpatne*. GPA: 4.0/4.0 — Blacksburg, USA

Virginia Tech — **Aug 2022 - Dec 2024**
M.S., Computer Science. Advisor: *Dr. Anuj Karpatne*. GPA: 4.0/4.0 — Blacksburg, USA

Birla Institute of Technology and Science (BITS), Pilani — **July 2016 – July 2020**
Bachelor of Engineering (B.E.), Computer Science, GPA: 8.08/10 — Pilani, India

Research Experience

KGML Lab, Virginia Tech | **Graduate Research Assistant** — **Jan 2023 – Present**

- Built Model-agnostic approaches for time-series modeling under **sparse data** conditions
- Exploring knowledge distillation of **LLMs** into light-weight interpretable models for time-series modeling
- Exploring **LLMs** for reasoning on non-textual modality (e.g. time series) for improved forecasting and multi-modal QA.
- Built a **Scientific Foundation Model** for spatio-temporal (probabilistic) predictions on irregular grids, leveraging multi-gpu training. Proposed an *(a) effective tokenization strategy* and *(b) Query-based forecasting*
- Developed a novel physics-guided **Diffusion Model** for Partial Differential Equation solving achieving *15-250x* faster inference
- Zero-shot evaluation of Vision-Language (**VLMs**) Models (*GPT-4, BLIP, LlaVA, Cog-VLM,* etc.) for *scene graph generation* and *reasoning ability* on VQA tasks involving scientific images

Publications

1. Atharva Pandey, **Abhilash Neog**, Gautam Jajoo. "On the Internal Semantics of Time-Series Foundation Models". *(BERT2S @ NeurIPS 2025)* [pdf]

2. **Abhilash Neog**, Arka Daw, Sepideh Fatemi, Medha Sawhney, et al. "Investigating a Model-Agnostic and Imputation-Free Approach for Irregularly-Sampled Multivariate Time-Series Modeling". Under review *(TSALM @ NeurIPS 2024)* [pdf]

3. **Abhilash Neog**, Medha Sawhney, KS Mehrab, Sepideh Fatemi, et al. "Toward Scientific Foundation Models for Aquatic Ecosystems". Under review (*FMSD @ ICML 2025)* [pdf]

4. Medha Sawhney, **Abhilash Neog**, Mridul Khurana, Anuj Karpatne. "Beyond Loss Guidance: Using PDE Residuals as Spectral Attention in Diffusion Neural Operators". Under review *(CVPRw 2025, ML4PS @ NeurIPS 2025)* [pdf]

5. Sepideh Fatemi, **Abhilash Neog**, Amartya Dutta, M. Sawhney, et al. "Scientific Equation Discovery using Modular Symbolic Regression via Vision-Language Guidance". *CVPRw 2025*

6. A. Dutta, M. Sawhney, K.S. Mehrab, **Abhilash Neog**, Mridul Khurana, et al. "Open World Scene Graph Generation using Vision Language Models". *CVPRw 2025, ICML 2025 Workshop* [pdf]

7. KS Mehrab, M. Maruf, Arka Daw, **Abhilash Neog**, HB Manogaran, et al. "Fish-Vista: A Multi-Purpose Dataset for Understanding Identification of Traits from Images". *CVPR 2025* [pdf]

8. M. Maruf, Arka Daw, KS Mehrab, HB Manogaran, **Abhilash Neog**, M. Sawhney, et al. "VLM4Bio: A Benchmark Dataset to Evaluate Pretrained Vision-Language Models for Trait Discovery from Biological Images". *NeurIPS 2024* [pdf]

9. Baviskar, A., Ramanathan, K., **Abhilash, N.**, Pawar, D. and Bangalore, K., Oracle International Corp, 2024. "Machine Learning Based Spend Classification." *U.S. Patent Application 17/903,161.* [pdf]

10. R. Ladwig, A. Daw, E.A. Albright, C. Buelo, A. Karpatne, M.F. Meyer, **A. Neog**, P. C. Hanson, and H. A. Dugan. "Modular Compositional Learning Improves 1D Hydrodynamic Lake Model Performance by Merging Process-Based Modeling With Deep Learning." *Journal of Advances in Modeling Earth Systems (JAMES) 16, no. 1 (2024)* [pdf]

11. Lavika Goel, **Abhilash Neog**, Ashish Aman, and Arshveer Kaur. "Hybrid Nature-Inspired Optimization Techniques in Face Recognition" *Transactions on Computational Science XXXVI, Springer LNCS, 2020.* [pdf]

Industry Experience

Kryptowire | **Machine Learning Intern** **May 2023 – Aug 2023**

- Developed an outlier detection model for denoising sensor-based Human Activity Recognition (HAR) **time series** data
- Built & deployed a CNN-based HAR model achieving 82% F-1 score on an android app using Keras & TensorFlow Lite

Oracle | **Data Scientist** **Sep 2020 – July 2022**

- Built & deployed Machine Learning applications into ETL pipelines, leveraging **Spark** systems, **MLOps** & **CI/CD** pipelines
- Designed and deployed a *Demand Prediction* application for time series forecasting using the DeepAR model
- Developed an unsupervised classification algorithm achieving 40% higher accuracy than then **Language Models** on a 71k-label dataset.

VMware | **Software Development Engineer Intern** **Jan 2020 – June 2020**

- Streamlined the process of fetching & filtering raw data from Workspace ONE Cloud using Spring Boot REST APIs
- Contributed to an end-user federation app on Workspace ONE Cloud, and wrote unit tests using JUnit and Mockito

Samsung Research Institute | **Summer Intern** **May 2019 – July 2019**

- Performed a feasibility study of Multi-frame Noise Reduction solutions' deployment in Live Focus for Low light conditions
- Optimized the existing HAL call flow, in C++, with considerable noise reduction in the first phase of live focus capture

Selected Projects

Can Large Vision Language Models (VLMs) Ground Fine-grained Attribute? pdf **Aug '24 – Dec '24**

- Developed a novel dual-scale attention framework for fine-grained attribute localization in Large Vision-Language Models (**LLaVa**), incorporating entropy-based head selection, maximally connected component filtering, and hierarchical constraints

Evaluating Model Reasoning & Hallucinations in Medical LLMs code pdf **Jan '24 – April '24**

- Analyzed and evaluated factual error propagation in open-source medical LLMs such as BioMistral, Asclepius, Alpacare, and PMC-LLaMA to identify variations in their efficacy and ensure reliable information dissemination in medical settings.

Convergence analysis of PINN for solving inverse PDEs code pdf **Aug '23 – Dec '23**

- Performed adaptive weighing of physics-based and data-driven loss terms in Physics-informed Neural Networks
- Achieved 50% average error reduction in PDE (Partial Differential Eq.) parameter estimation of Burgers & Allen-Cahn eq.

Mathematical Reasoning in Large Language Models (LLMs) code pdf **Aug '23 – Dec '23**

- Worked on the problem of numerical headline generation and numeral masked-fill as part of NumEval @ SemEval 2024
- Adapted **Llama**, **T5**, **BART** & **RoBERTa** models by Direct **fine-tuning** & **prompt engineering** for the respective tasks

Text Summarization of Electronic Theses and Dissertations (ETD) pdf **Sept '22 – Dec '22**

- Developed a text summarization pipeline, integrating both Transformer-based abstractive algorithms (pre-trained Pegasus & RoBERTa) and traditional extractive algorithms like TextRank, LexRank & LSA, within an ETD Info. Retrieval system

Technical Skills

Languages: Python, Java, C++, SQL, R

Frameworks: PyTorch, Git, Hydra, Tensorflow Keras, Spark

Miscellaneous

- **Reviewer**:
  - NeurIPS 2025, AAAI 2026
  - ICML 2025 Workshop on Foundation Models for Structured Data
  - ICLR 2025 Workshops: Foundation Models in the Wild, Open Science for Foundation Models, XAI4Science
  - Journal on Systems and Soft Computing
- **Talks**:
  - Invited talk at the AGU 2025 Session on AI + Water Science
  - Talk on Transfer Learning in Lake Ecosystems at the "*NSF Macrosystems Biology Meeting 2024*"
  - Lightning Talk at the "*Frontiers in Ecological Forecasting 2023,*" event at Virginia Tech
- **Awards**:
  - NSF NAIRR (National AI Research Resource) Pilot Award, 2024
  - "*Star of the Month (Dec 2021)*" within the Oracle Analytics Cloud Organization, Oracle India
- **Teaching**:
  - Graduate Teaching Assistant, CS 5805 Machine Learning, Spring 2024
  - Teaching Assistant, BITS F312 Neural Networks and Fuzzy Logic, Fall 2019