Amirtha Varshini A S

+1 (470) 812-5349 | amirtha255@gmail.com | linkedin.com/in/amirtha-varshini-as | amirtha255.github.io

EDUCATION

- **Georgia Institute of Technology**, Atlanta, GA — Aug. 2021 - May 2023
  Master of Science in Computer Science (ML Concentration) — **GPA: 4.0/4.0**
- **National Institute of Technology Tiruchirappalli**, India — Jul. 2014 - May 2018
  Bachelor of Technology in Electronics and Communication Engineering — **GPA: 8.90/10**

TECHNICAL SKILLS

- **Languages**: Python, C++, C, SQL, Bash
- **Machine Learning & Frameworks**: PyTorch, TensorFlow, Scikit-learn, HuggingFace,CUDA, W&B, MLflow
- **Cheminformatics & Modeling**: RDKit, CReM, Docking - MOE and Boltz2 , rdfilters
- **Data & Systems**: AWS, GCP, Redun, DuckDB, Linux
- **Graduate Coursework**: Machine Learning with Limited Supervision, Deep Reinforcement Learning, Computer Vision, Graduate Algorithms, Advanced Machine Learning, Deep Learning Specialization (Coursera)

EXPERIENCE

- **Montai Therapeutics**, Cambridge, MA - *Machine Learning Scientist II* — Jul 2023 – Present
  - Worked on extending **SynFlowNet**, a reaction-based GFlowNet model that generates synthesizable molecules from chemical reactions and building blocks. Trained models with the **Boltz-2 reward function** and bioactivity model scores, enabling probabilistic sampling of high-quality compounds with favorable physicochemical properties; multiple candidates were **validated by synthesis**.
  - **Designed fragment-based de novo workflows** with the CReM framework, generating chemically valid and synthetically accessible structures by design. Implemented **multi-objective optimization** with ADME property filters, bioactivity predictions, and reward conditioning to balance potency, novelty, and drug-likeness; contributed to the **open-source CReM library**.
  - **Built predictive ML models** for molecular property, ADMET and SAR potency prediction (Spearman >0.8). Improved in-house enrichment factor by **28%** through large-scale pretraining (MolData) and delivered an additional **30% gain** in collaboration with Pfizer using PubMed-derived biochemical datasets.
  - **Advanced deep learning approaches**: integrated graph neural networks, transformers, Chemprop, and MegaMolBART embeddings with cheminformatics descriptors. Applied hybrid modeling with uncertainty quantification for robust compound nomination (**AUC ~0.8 across diverse targets**).
  - **Built interpretability frameworks** for GNNs and Transformers, applying Monte Carlo Tree Search and counterfactual generation to extract mechanistic substructure rationales. Built **interpretability dashboards** in Streamlit using RDKit and SMARTS filters to visualize substructure rationales, highlight predictions, and flag structural alerts, translating model insights into actionable medicinal chemistry decisions.
  - Engineered **reproducible ML pipelines** integrating cheminformatics modules for scalable generative and interpretability workflows; contributed to a **patent-pending method** for identifying core structures in large chemical datasets.
  - Designed and deployed **interactive interpretability dashboards** using Streamlit (frontend) and FastAPI services with Redis caching and Amazon S3 storage. Integrated embedding visualizations (t-SNE, UMAP, TMAP) for clustering and molecule exploration, and served predictions via Ray Serve for scalable inference and retrieval, directly supporting medicinal chemistry design workflows

- **Amazon Robotics**, Westborough, MA - *Software Development Engineer Intern* — May 2022 - Aug. 2022
  - Built an Augmented Reality-based real-time tracking application on HoloLens 2 combining computer vision and 3D spatial mapping to automate package identification, demonstrating scalable integration of CV and systems software.

- **Qualcomm**, Bengaluru, India - *Software Engineer* — Jul. 2018 - Aug. 2021
  - *ADAS team* -Designed Minidump feature on a QNX Real-time operating system to capture a snapshot of a system post-crash. Brought down the download time by **70%** and the size from **12GB to 300 MB**, enabling faster analysis
  - Developed a GDB-based Python and C parser to extract debug information from the collected kernel dump.
  - Built FastRPC framework to offload high-compute tasks from CPU to Digital Signal Processors, improving performance
  - Implemented tools to monitor system metrics such as watchdog timer, heartbeat, power consumption, and memory usage
  - Worked on bring-up of an SoC with ARM Architecture and developed tools for power, temperature & memory metrics

RESEARCH

- **Interpreting GFlowNets for Drug Discovery**-*Montai Therapeutics*   Aug. 2025 – Present
  - ○ **Contributed to scientific dissemination**: accepted poster at NeurIPS WiML and MoML (Molecular Machine Learning Conference) at MIT, open-source code contributions; regularly presented research to cross-functional teams.
  - ○ Designed the first **interpretability framework for hierarchical GFlowNets** in molecular design, advancing transparency of deep generative models for drug discovery.
  - ○ Engineered **gradient-based saliency maps** and **SMARTS-driven counterfactuals** to produce atom-level attributions with causal evidence of substructure importance, accelerating molecular optimization.
  - ○ Applied **sparse autoencoders and linear probes** to SynFlowNet embeddings, disentangling drug-likeness (QED) into interpretable latent factors such as size, polarity, and lipophilicity.
  - ○ Recovered chemically meaningful motifs (functional groups, rings, halogens) from embeddings, bridging ML representations with medicinal chemistry reasoning. **Paper | OpenReview | Code**

- **Explainability for Graph Neural Networks in Proactive Robot Assistance**   Aug. 2022 – Apr. 2023
  - ○ *Research advised by Prof Sonia Chernova at Georgia Tech.* Developed explainability methods for a **dynamic spatio-temporal GNN** performing real-time object tracking and future movement prediction in home environments.
  - ○ Applied **GNNExplainer** with novel counterfactual approaches to derive edge- and time-based explanations, yielding **human-aligned, intuitive outputs**.
  - ○ Proposed new evaluation metrics (**Edge-Time F1, Time Recall**) and validated on the real-world HOMER activity dataset. Performed ablation studies with saliency, attention weights, and temporal perturbations to assess explanation faithfulness.
  - ○ Contributed to **PyTorch Geometric** explainer modules through metric design and code reviews. **GitHub**

- **Text-to-video generation using Latent Diffusion**   Aug. 2022 - Dec. 2022
  - ○ Trained a transformer to generate future video frame embeddings on top of the Stable Diffusion encoder.
  - ○ Outperformed the TGANv2 baseline by **26%** improvement in Frechet Video Distance score by using a novel combination of loss functions and video interpolation components.   **Code**

- **Deep Reinforcement Learning (RL) based autonomous driving**   Jan. 2022 - May 2022
  - ○ Built TQC (Truncated Quantile Critics) algorithm with experience replay and increased rewards by **17%** for navigation in a self-driving simulator Donkeycar. Improved rewards by **42%** by training a Variational Autoencoder to compress inputs.
  - ○ Generated a semantic segmentation mask using a pretrained autoencoder to visualize the model decisions. **Demo |Report**

- **Semantic Similarity and Toxicity Detection of Questions in Quora**   Sep. 2021 - Dec. 2021
  - ○ Using PyTorch, compared the results of BERT, Bi-LSTM, Bi-RNN, and Bi-GRU models with NLP word-embedding techniques TF-IDF Vectorization and Word2Vec to predict intent similarity and toxicity of questions on Quora. **Link**
  - ○ Achieved F1-score of **0.7** by fine-tuning BERT to predict question sincerity and accuracy **0.89** for questions' similarities.

- **Computer Vision Tools for Non-verbal Communication in Interviews**   Aug. 2021 - Dec. 2021
  - ○ Devised a K-Nearest Neighbours(KNN) model to estimate head pose in videos with accuracy **83%**. Obtained features as the difference in minima and maxima of first-order pitch differences, from OpenFace Keypoints output on AMI corpus. **Link**
  - ○ **Runner-up** at Innovation Competition 2022, an Entrepreneurial challenge of VentureLabs, Georgia Tech.

- **Low-cost intelligent vision in automotive (LIVA)**   Jun. 2019 - Oct. 2019
  - ○ Collected dataset of depth images using Kinect V2 mounted on a moving car. Achieved object detection accuracy **85%** in real-time to recognize pedestrians and vehicles by fine-tuning YOLO V3 model with depth images and COCO dataset.
  - ○ Converted model to DLC and ran inference on Linux Vehicle platform using SNPE (Snapdragon Neural Processing Engine)
  - ○ Top 6 finalists out of the 230+ applicants in Maker's Challenge of QBuzz 2019, Qualcomm's annual tech conference.

- **Real-Time Hand Gesture Recognition system**   Jan. 2018 - May 2018
  - ○ Fine-tuned Inception V3 Architecture on ASL dataset to detect gestures with **98%** accuracy and controlled a custom-built robotic arm. Published a paper as the **first author**: Amirtha Varshini, A.S. and et.al, "Real-time Hand Gesture Recognition for Robotic Arm and Home Automation", **(ISEEIE 2021) Link**
  - ○ Best Final Year Project Award by Sonata and Centre for Entrepreneurship Development and Incubation, NIT Trichy.

ACHIEVEMENTS

- Granted scholarship to represent College of Computing, Georgia Tech at **Grace Hoppers Conference, 2022**
- Runner-up at **Innovation Competition 2022**, an Entrepreneurial challenge of VentureLabs, Georgia Tech.
- Recipient of **K. C. Mahindra Scholarship** for Post Graduate Studies Abroad, 2021
- Top 6 finalists out of the 230+ applicants in Maker's Challenge of QBuzz 2019, Qualcomm's annual tech conference.
- Recipient of **AIEEE Merit Scholarship for Rank 1448 (Top 0.1% amongst 1,350,000 candidates)** in JEE Main'14
- Received **two Qualstar recognitions** for innovation and excellent delivery of results.
- Poster on 'Automated Bug Triage with ML' selected for Qualcomm ML Summit'19