

# Vamsi Deeduvanu

(765)-694-9091 | [vamsi10010@gmail.com](mailto:vamsi10010@gmail.com) | [linkedin.com/in/vamsideeduvanu/](https://www.linkedin.com/in/vamsideeduvanu/) | [github.com/vamsi10010](https://github.com/vamsi10010)

## EDUCATION

|  |                                       |
|--|---------------------------------------|
| <b>Purdue University</b><br><i>Masters of Science in Computer Science</i>  | Aug. 2024 – May 2026<br>GPA: 4.0/4.0  |
| <b>Purdue University</b><br><i>Bachelor of Science in Computer Science</i>   | Aug. 2022 – May 2025<br>GPA: 3.99/4.0 |
| <ul style="list-style-type: none"><li>• <i>Coursework:</i> OOP, DSA, Computer Architecture, Systems Programming, Compilers, AI, Machine Learning, NLP, Robotics, ML Systems, Randomized Algorithms, Statistical Theory</li><li>• <i>Honors:</i> Dean's List and Semester Honors (6x), L3Harris Scholarship, UG Research Expo Award</li></ul> |                                       |

## EXPERIENCE

|   |   |
|---|---|
| <b>Software Development Engineer Intern</b><br><i>Amazon, Project Kuiper</i>  | May 2025 – Aug. 2025<br>Redmond, WA         |
| <ul style="list-style-type: none"><li>• Developed a cost tracing service with RESTful API improving supply chain cost visibility for Project Kuiper.</li><li>• Deployed a scalable serverless data pipeline on AWS achieving sub-10ms queries on millions of cost events.</li><li>• Built an MCP agent using Neo4j and Claude enabling non-tech stakeholders to perform natural language queries.</li></ul> |   |
| <b>AI/ML Intern</b><br><i>Volvo Group</i>   | May 2024 – Aug. 2024<br>Hagerstown, MD      |
| <ul style="list-style-type: none"><li>• Designed an edge AI pipeline to identify service bottlenecks on factory floor using YOLOv8n and PaddleOCR.</li><li>• Developed a live web interface using Streamlit to monitor KPIs such as truck count and takt time on-site.</li><li>• Leveraged VAR and LSTM models to forecast service requests to reduce downtimes and improve service efficiency.</li></ul>   |   |
| <b>Undergraduate Teaching Assistant</b><br><i>Department of Computer Science, Purdue University</i>   | Aug. 2023 – May 2025<br>West Lafayette, IN  |
| <ul style="list-style-type: none"><li>• Mentored undergraduate students on foundational concepts in C programming and systems programming courses.</li><li>• Developed programming assignments and test frameworks to automate evaluation of students' understanding.</li><li>• Led weekly lab sessions and office hours for 40+ students enhancing student learning outcomes and grades.</li></ul>         |   |
| <b>Undergraduate Researcher</b><br><i>TinyML/IIoT, Purdue University</i>  | Aug. 2023 – Dec. 2023<br>West Lafayette, IN |
| <ul style="list-style-type: none"><li>• Developed a data pipeline to collect and process real-time CNC machine and sensor data using MTConnect.</li><li>• Labelled and annotated sensor data to train a CNN to predict machining failures with &gt; 90% accuracy.</li></ul>   |   |
| <b>Data Science Researcher</b><br><i>Battelle</i>   | Aug. 2022 – May 2023<br>West Lafayette, IN  |
| <ul style="list-style-type: none"><li>• Conducted research on hyperparameter tuning algorithms for LLMs and established an SOP for future projects.</li><li>• Fine-tuned BioBERT from HuggingFace to accurately identify adverse drug events in electronic health records.</li><li>• Boosted overall f1 score by more than 20% using hyperband and population-based algorithms from RayTune.</li></ul>      |   |

## PROJECTS

|   |                       |
|---|-----------------------|
| <b>LLM Uncertainty Calibration</b>   <i>Python, PyTorch, HuggingFace, scikit-learn, LMDB</i>  | Mar. 2025 - May 2025  |
| <ul style="list-style-type: none"><li>• Built an LLM calibration pipeline to quantify response uncertainty on real-world tasks during inference.</li><li>• Designed a rigorous scikit-learn evaluation suite to compute error metrics and generate calibration curves.</li><li>• Scaled experiments with LMDB tensor caching and batched GPU inference, speeding up experiments.</li></ul>          |                       |
| <b>hirehack</b>   <i>Python, JavaScript, PyTorch, HuggingFace, PRAAT, WebSpeech API</i>   | Jan. 2024 - Feb. 2024 |
| <ul style="list-style-type: none"><li>• Developed an LLM agent to automatically analyze interview performance through a Chrome extension.</li><li>• Integrated facial emotion, prosodic, and lexical features into a multi-modal model to score interview performance.</li><li>• Interfaced Mixtral-7B from HuggingFace API to interpret model output and generate feedback in real-time.</li></ul> |                       |
| <b>cgrad</b>   <i>C, C++, cmocka, Deep Learning</i>   | Aug. 2023 – Sep. 2023 |
| <ul style="list-style-type: none"><li>• Created a lightweight neural network library from scratch in C achieving 96% accuracy on MNIST dataset.</li><li>• Implemented automatic differentiation, layers, activation functions, gradient descent, and regularization options.</li><li>• Automated testing process and ensured functionality by creating unit tests using cmocka framework.</li></ul> |                       |

## SKILLS

**Languages:** Python, C, C++, Java, SQL, R, CUDA, JavaScript, TypeScript, Smithy, LaTeX, x86-64 Assembly  
**Developer Tools:** Git, Bash, UNIX, MacOS, AWS, Azure, IaC, Docker, Neo4j, REST API, uv, Agile  
**Libraries:** PyTorch, HuggingFace, RayTune, Streamlit, Tensorflow, JAX, LLVM, MLIR, TVM, CDK