Idhant Gulati

idhant.gulati@psu.edu | idhant.xyz

Education

The Pennsylvania State University — University Park, PA
Bachelor of Science in Computer Science — *Aug. 2023 – May 2027*

Experience

Host — Janaury 2024 – Present
gradient — *Unversity Park, PA*

- Founded and lead a community-driven co-working initiative focused on helping creators and builders pursue passion projects through structured work sessions
- Design and facilitate weekly deep work sessions featuring focused work blocks, project demonstrations, and peer feedback segments
- Cultivate an engaged community of builders and creators, implementing a selective application process to maintain high-quality project standards
- Developed framework for project sharing and constructive feedback, enabling members to iterate and improve their creative endeavors

Research — August 2024 – Present
Independent

- Led systematic analysis of expert specialization in DeepSeekMoE through domain-specific routing patterns and early decoding frameworks
- Demonstrated that single top-weighted experts achieve 95% performance of full ensemble, revealing opportunities for inference optimization
- Developed novel methodology combining LogitLens and residual stream analysis to track expert contributions across English, French, and code domains
- Quantified expert specialization patterns showing concentrated expertise, with ¡5 experts handling over 50% of routing decisions

Undergraduate Research Assistant — July 2024 – Present
CAIS Lab, The Pennsylvania State University — *University Park, PA*

- Implementing a data collection system to gather various parameters from multiple 3D printers over an extended period.
- Designing machine learning model using collected data to predict to optimize the 3D printing process.
- Creating a Unity-based simulation dashboard to visualize and interact with the optimization model.
- Developed a comprehensive dashboard for real-time monitoring and analysis of 3D printing performance metrics.

Research Project: DenseTEX — June 2024 – August 2024
buildspace s5 — *(Remote) San Francisco, CA*

- Developed DenseTEX, a 95M parameter deep learning model that converts mathematical equation images to LaTeX code, integrating a DenseNet-169 CNN encoder with a GPT-2 decoder architecture.
- Pioneered the implementation of 2D Positional Encoding in Image-to-LaTeX models, significantly improving spatial awareness and mathematical notation preservation.
- Trained the model on 4xA6000 GPUs using the UniMER-1M dataset (1M samples) for approximately 20 hours, achieving a BLEU score of 0.80 and validation loss of 0.45.
- Designed architecture and custom positional encoding formulas through independent research, optimizing for complex mathematical expression transcription.

Research & Publication

Chaudhari, M., Gulati, I., Hundia, N., Karra, P., & Raval, S. (2025). *MoE Lens - An Expert Is All You Need.* Accepted at ICLR 2025 Workshop on Sparsity in LLMs (SLLM). URL: `https://openreview.net/forum?id=GS4WXncwSF`

Projects

CodeWhisper | *Python, TypeScript, Node.js, Deepgram API, Groq API, Anthropic API* October 2024

- Developed an VS Code extension enabling hands-free coding through natural speech commands
- Implemented real-time speech-to-code functionality by integrating Deepgram's speech recognition with language models using Groq's and Anthropic's API
- Designed an accessibility-focused system analyzing entire codebases for context-aware code editing and generation
- Built scalable architecture supporting multi-language coding assistance and real-time visual feedback
- Created at CalHacks, focusing on making coding more accessible for developers with visual or motor impairments

IdeaStruct | *Python, Flask, OpenAI API, Neo4j, Cytoscape.js* September 2024

- Developed an AI-powered Flask web app for generating and querying dynamic knowledge graphs
- Integrated OpenAI's GPT for NLP and implemented a flexible backend with Neo4j and in-memory databases
- Created an interactive frontend using Cytoscape.js for complex data visualization
- Engineered features like conditional data addition, URL scraping, and custom integration management
- Collaborated in a team of four to complete the project within HackMIT time constraints

Others: Glaucoma Prediction Model, Textify: CNNs from Scratch, and Vaccine Management System