Surendra Kumar Yadav

Greater Noida | Mail | Phone | LinkedIn | Github

Skills

- **Languages:** C++, Java, Python, SQL, JavaScript, Typescript
- **Frameworks:** React.js, Next.js, Express.js, Node.js, Tailwind CSS, FastAPI, Pytorch, Tensorflow, RestAPI
- **Tools:** Zustand, Clerk, Prisma, MongoDB, PostgreSQL, MLFlow, Langchain, Git, Docker
- **Other:** AWS EC2, S3, Redshift, Object-Oriented Programming, End-to-End Deployment

Education

Bennett University, B.tech in Computer Science — Sept 2022 – May 2026

- GPA: 8.86/10
- **Coursework:** Data Structures & Algorithms, Computer Networks, Operating System

Experience

CODE HELP – Intern — March 2024 – Sept. 2024

- This boot camp covered important Data Structures and Algorithms concepts, including Basic Programming, Looping, Functions, arrays, Strings, stacks, and topics including OOPS.
- During the boot camp, I exclusively used Leetcode and GFG as my main platform to tackle complex data structures and algorithms problems.

Projects

Flow Genius

- Developed a robust workflow automation application integrating Google Drive, Discord, and Notion, utilizing React, Next.js, and Prisma for seamless cross-platform functionality.
- Implemented real-time updates and dynamic connections, enhancing system responsiveness by 40% while optimizing API calls for efficient data synchronization.
- Integrated Clerk for user authentication, streamlining sign-up and OAuth flows for Google and Discord, improving authentication process efficiency by 30%.

Instruct tuning Llama2 with RLHF loop

- Fine-tuned LLaMA2 models (up to 13B parameters) with pre-training, SFT, and RLHF, achieving a 20-30% improvement in instruction-following accuracy over baseline models
- Utilizing techniques like LoRA and PPO, the pipeline reduces training time by 50% while maintaining 95% model performance compared to full-scale training, all within a 24GB GPU memory budget
- Training a 7B model with 50K+ samples, the model achieves an 85% success rate on human evaluation tasks, with a 10x reduction in fine-tuning cost using quantization and FSDP.

VisionVoice-AI

- The image-to-text generation step uses Salesforce's BLIP model, processing images at 95% accuracy for context understanding, with captions generated in under 3 seconds per image.
- ESPnet TTS model converts text to audio with human-like prosody, achieving a 99% intelligibility rate, while processing and generating speech in under 10 seconds per story.
- The full pipeline (image upload to audio output) completes in less than 2 minutes, delivering a seamless user experience with high efficiency and minimal latency.

Achievements

LeetCode, GFG

Solved 350+ DSA questions

5* Rating in C++ at Geeks for Geeks.