

# Anubhav Vashishtha

(+91) 8192939336 | anubhavvashishtha.iitd@gmail.com | [linkedin](#) | Github ([Second](#) , [First](#)) |

---

## EDUCATION

### INDIAN INSTITUTE OF TECHNOLOGY, DELHI (IITD)

2022 - 2026

- **Coordinator , Student Incubation Cell , IIT Delhi** : Led startup ecosystem initiatives at IIT Delhi, organizing competitions, fostering innovation, and supporting entrepreneurial growth.

---

## WORK EXPERIENCE

### Machine Learning Intern , INNOGICAL PVT. LTD.

Jun 2023 - Jul 2023

- Developed AI-Powered Student Analytics using machine learning to assess and improve student performance in government schools.
- Enabled Data-Driven Decision-Making by providing educators and policymakers with actionable insights for personalized student development.
- Achieved Socio-Educational Impact by leveraging AI to enhance learning outcomes in underprivileged sectors.
- Recognized at a Government Level with the project inaugurated by the Governor of Uttar Pradesh, highlighting its significance.

### AI Intern , Data Design Oy

Jun 2025 - Aug 2025

- Built a chat agent for a **Finland-based real estate company**, enabling direct database querying through natural language.
- Optimized Google's **A2A Protocol** for multi-agent parallel processing, enhancing performance and scalability of the chat agent.
- Delivered a robust, database-integrated AI solution that streamlined client interactions with real-time data access.
- Secured **Pre-Placement Offer to continue work in Finland starting September 2026** based on exceptional technical leadership and project delivery.

### AI Engineer , Data Design Oy

Sep 2025 - Dec 2025

- Lead Development of **Text-to-SQL Chat Agent**: After the success of the previous project, I was appointed **Tech Lead** for the next month's initiative.
- Architected a robust text-to-SQL system that interfaced with any client production databases using connection string , requiring **minimal to no fine-tuning** for seamless querying within minutes.
- Achieved **69% score on BIRD Benchmark**, demonstrating strong performance on complex real-world database querying tasks.

---

## PROJECTS

### SmartCouncil : My own Multi Agent architecture

Dec 2025

- Designed a multi-agent orchestration framework using **Llama-3.1-8B-Instruct** with **automated task decomposition, dependency resolution, and parallel execution optimization**.
- Built a **SQLite-backed** dynamic task scheduler with **real-time dependency-graph traversal** for maximizing concurrent execution across four specialized domain agents.
- Implemented an intelligent task-optimization pipeline that removes redundant subtasks and prevents over-decomposition to reduce overall workflow latency.
- Developed a stateful execution **lifecycle manager** with selective **context propagation using JSON-based dependency tracking** for coherent multi-agent workflows.

- Engineered a high-performance orchestration stack leveraging Llama 3.1, HuggingFace Inference API, SQLite, and **ThreadPoolExecutor for scalable multi-agent execution.**

#### MOE for edge devices

Nov 2025

- Developed an intelligent multi-agent mobile mental-health chatbot with five **Gemma-3-1B LoRA-fine-tuned models** running fully offline on **Android via TensorFlow Lite**.
- Engineered an **orchestrator using TF-IDF vectorization** and Logistic Regression to classify and route user queries with **87% accuracy**.
- Built an **automated data pipeline** that extracts medical-literature knowledge from PDFs and generates conversational **Q&A pairs using Llama-3.2-1B**.
- Optimized on-device models with **4-bit quantization and MediaPipe task packaging**, reducing **each model to ~150MB** while preserving response quality.
- Leveraged Gemma, Llama, LoRA/PEFT, TFLite, scikit-learn, AI Edge Torch, and Transformers to deliver a fully offline, high-performance mental-health assistant.

#### AI search engine for hotel search

Jan 2025

- Developed an LLM-Based Search Engine:** Designed a hotel search system that **eliminates traditional filters, allowing users to find hotels through natural language queries**.
- Implemented Vector-Based Similarity Matching:** Converted hotel descriptions into vector embeddings and used cosine similarity to retrieve the most relevant results.
- Optimized Hotel Data Processing:** Precomputed hotel feature vectors from the dataset and stored them in a MongoDB database for efficient real-time retrieval.
- Built a Scalable Backend Architecture:** Developed a robust backend using Node.js, Express, and Flask to handle API requests and vector processing.
- Enhanced User Experience with Dynamic UI: **Utilized Pug templating** to create an interactive and seamless search interface.

#### RL stock market bot

Feb 2025

- Developed an **RL-Powered Trading Agent**: Built a reinforcement learning model that learns optimal stock trading strategies from historical market data.
- Designed a Custom Trading Environment: **Simulated real-world market conditions** to train and test the RL agent effectively.
- Implemented Deep Reinforcement Learning: **Utilized neural networks** to determine the best trading policy, optimizing buy/sell decisions.
- Automated Training and Model Optimization: Saved and reloaded the best-performing model weights to enhance trading performance.
- Preprocessed and Analyzed Market Data: Designed data pipelines using Pandas and NumPy for importing and processing stock data.

#### T20 prediction model

Jun 2024

- Developed a T20 Cricket Match Prediction Model: **Built a machine learning model to forecast match outcomes** with high accuracy.
- Comprehensive Data Analysis: **Conducted thorough EDA** and used advanced imputation techniques to handle missing data effectively.
- Modeled Key Match Influencers: Factored in ground details, historical match stats, and batting performance to enhance predictions.
- National-Level Recognition: **Secured a top-10 rank among premier IIT and IIM teams across India**.
- Expertise in Feature Engineering: Leveraged innovative techniques and data-driven insights to deliver impactful solutions.

#### Visual Taxonomy Attribute Prediction Model

Nov 2024

- Developed a Multi-Attribute Prediction Model:** Designed a robust deep learning model to predict multiple visual attributes using a custom dataset.

- **Leveraged TensorFlow & ResNet50:** Built the model using TensorFlow with ResNet50 architecture for feature extraction.
- **Advanced Data Processing:** Employed data augmentation for better generalization and innovative preprocessing to handle missing values.
- **Efficient Multi-Output Design:** Integrated separate output layers for each attribute with categorical cross-entropy loss for precise predictions.
- **Optimized Training Pipeline:** Created a scalable data pipeline with tf.data for real-time image loading and implemented Batch Normalization and Dropout for improved model performance.

#### **Hand Gesture Cursor Controller**

**Aug 2024**

- Inspired by Meta's Orion Project: Developed an innovative **hand gesture-based cursor control** system using YOLO keypoint detection.
  - Real-Time Fingertip Detection: Accurately identifies fingertip positions to facilitate seamless mouse event handling via hand gestures.
  - Integrated OpenCV & PyAutoGUI: Used **OpenCV for video input and PyAutoGUI** for system navigation, enabling intuitive gesture-based cursor control and "pinch" gestures for mouse clicks.
  - **Custom YOLO Model for Precision:** Trained a custom YOLO model to ensure accurate fingertip detection, enhancing system responsiveness and robustness.
  - Practical Application: Designed for real-world usability, making it ideal for hands-free control and accessibility.
- 

## **SKILLS**

- **Programming Languages** - Python, JavaScript (Vanilla), HTML, CSS (Tailwind, SCSS, CSS Modules, Component-based)
- **Frameworks & Tools** - FastAPI, Express.js, Node.js, React, Next.js, RESTful APIs, Postman, Git, PUG, MongoDB, PostgreSQL, Jupyter Notebooks, HuggingFace Transformers, MediaPipe
- **AI & Machine Learning** - LLMs (Large Language Models - Llama 3.2, Gemma 3), Prompt Engineering, Multi-Agent Systems (A2A Protocol, MCP), OpenAI API, Claude API, Grok API, PydanticAI, TensorFlow, TensorFlow Lite, PyTorch, Fine-tuning (LoRA, PEFT), Model Quantization (4-bit, Dynamic int4), NLP (Text Classification, Tokenization, TF-IDF), Supervised & Unsupervised Learning, Reinforcement Learning, Deep Learning, Data Analysis (Tableau, PowerBi, NumPy, Pandas), scikit-learn, On-device ML Deployment, AI Edge Torch, Model Optimization