

ABHINAY KOTLA

+1 469 674 1021 ◊ abhinaykotla@gmail.com

[GitHub](#) [Portfolio](#)

SUMMARY

Software Engineer with over 3 years of experience building and deploying AI systems. Skilled in **Machine Learning, Data Structures, and Algorithms**. Developed AI-powered chatbots, optimized computer vision pipelines, and fine-tuned NLP models for production use. Experienced in full-stack software development, integrating ML with scalable systems, and delivering real-world business impact through data-driven automation.

WORK EXPERIENCE

IT Operations <i>UT Arlington Office of Information Technology, Arlington, TX</i>	Sep 2024 – Present
<ul style="list-style-type: none">Designed and developed an AI-powered chatbot using Python and NLP for campus IT, automating support queries through model-based intent detection.Improved ticket resolution time by 18% and deployed retraining pipelines with ServiceNow integration.Supported IT operations by developing automation scripts and optimizing infrastructure workflows.	
OIT Tech Support <i>UT Arlington Office of Information Technology, Arlington, TX</i>	Aug 2024 – Sep 2024
<ul style="list-style-type: none">Supported ~5,000+ users, resolving hardware, software, and network issues in Linux, macOS, and Windows systems.Documented IT workflows and created knowledge-base SOPs, reducing ticket escalations by 20%.Assisted in device imaging, switch configurations, and network maintenance across campus labs.	
Full Stack Developer – Saintechinc <i>Python, React, Node.js, REST APIs, Authentication, Deployment</i>	Feb 2023 – Nov 2023
<ul style="list-style-type: none">Built and deployed a responsive full-stack application with secure user roles and real-time analytics.Achieved 99.9% uptime and improved SEO metrics by 40% through structured data and optimized routing.Applied component-based architecture for modular scalability and maintainability.	
Machine Learning Engineer – 1StopAI <i>Python, scikit-learn, Audio Processing, ML Deployment</i>	Nov 2021 – Jan 2022
<ul style="list-style-type: none">Developed real-time mood and gender classification models for customer support calls.Improved routing accuracy by 15% through emotion detection and ML inference optimization.Deployed trained models in production using lightweight inference APIs.	
ML Content Developer – SmartKnowers	Jun 2022 – Jul 2022
<ul style="list-style-type: none">Designed ML learning modules on model evaluation, data pipelines, and cloud deployment.Delivered interactive Jupyter notebooks and exercises that increased course engagement by 30%.	

SKILLS

Programming	Python (Advanced), C++, Java, Scala, R, SQL, Bash
Core CS Areas	Machine Learning, Data Structures, Algorithms, Systems Design
ML & AI	Deep Learning, Generative AI, LLMs, Transfer Learning, NLP, GANs, Feature Engineering
Computer Vision	CNNs, Vision Transformers (ViTs), Image Inpainting, Object Detection
Frameworks	PyTorch, TensorFlow, Keras, scikit-learn, OpenCV, Pandas, Spark
Cloud & Big Data	AWS, Google Cloud, Azure, Spark, Hadoop, Kubernetes
DevOps & MLOps	Docker, GitHub Actions, CI/CD, Model Serving, Terraform, Monitoring
Data Engineering	ETL Pipelines, Data Wrangling, Kafka Streams, SQL Optimization
Web Development	React, Next.js, Node.js, Firebase, REST APIs, CSS3, HTML5

PROJECTS

Edge- and Color-Aware Adversarial Image Inpainting

Python, GANs, Computer Vision, Deep Learning, Image Processing, PyTorch, Feature Matching, Loss Engineering

- Designed a dual-stage GAN (G1 edge generator, G2 color generator) for realistic completion on the CelebA dataset.
- Reduced perceptual loss by 7% via color-hint regularization, feature-matching loss, and adversarial training.
- Trained G1 on masked Canny edges and G2 for structure-texture fusion, improving fine-grained details.

TuneParams.ai Community Platform

React, Firebase, Firestore, Node.js, CSS, Authentication, Cloud Deployment, Web Security

- Built and deployed a scalable community platform for AI/ML discussions with public-read access and moderation.

- Implemented role-based authentication, Firestore security rules, and modular component architecture.
- Integrated user/payments dashboards and managed serverless functions for analytics and forum data.

Efficient CV Models with Knowledge Distillation

Python, PyTorch, CNNs, Quantization, Model Compression, Optimization

- Compressed a 669MB CNN into a 6.5MB model (~99% smaller) while maintaining 97% accuracy.
- Applied pruning, FP16/FP8 quantization, and dynamic residual learning for low-latency inference.
- Achieved real-time inference with ~10× less GPU memory and faster throughput.

News Summarization using T5 Transformer

Python, T5, Transformers, NLP, Hugging Face, API Integration

- Fine-tuned T5 on a news dataset for abstractive summarization with semantic coherence.
- Achieved ROUGE-1: 0.532 and cosine similarity of 0.80 against human references.
- Exposed the summarizer as an API with adjustable temperature and output length.

Road Surface Classification Tool

Python, scikit-learn, Qt Designer, Data Visualization, Predictive Analytics, GUI Design

- Built a GUI-based application to classify road surface defects and predict lifespan using ML.
- Trained and evaluated on 3.5k samples; achieved 85% accuracy across defect classes.
- Added visualization and reporting tools for field engineers using PyQt and matplotlib.

[More Projects on GitHub: github.com/Abhinaykotla]

EDUCATION

Masters in Computer Science

The University of Texas at Arlington

GPA: 4.0/4.0

BE in Computer Science

Gandhi Institute of Technology and Management, Hyderabad

CGPA: 8.47

Certifications

- Neural Networks and Deep Learning (DeepLearning.AI)
- Robotic Process Automation (RPA)
- IBM Big Data with Spark and Hadoop
- Wordcloud Using NLP and TF-IDF
- Google Technical Support Fundamentals

LEADERSHIP & COMMUNITY

Mentor – SCAI Club, University of Texas at Arlington

Jun 2024 – Present

- Led workshops on Machine Learning and Data Science fundamentals.
- Supervised a student project on cryptocurrency trend prediction using real-time inference.

Head – Coding and Gaming Division, GITAM COGAAN

Jul 2020 – Jan 2023

- Organized 10+ technical events, doubling participation through competitive coding and ML workshops.
- Conducted hands-on training on Algorithms, Data Structures, and gaming AI.