

Sushant Gundla

Bengaluru, India

+91 7738328985

gundla.sushant@gmail.com

<https://www.linkedin.com/in/sushantgundla>

<https://sushantgundla.com>

Technical Lead – AI/ML & Innovation with 7+ years of experience building **Agentic AI, RAG, LLM, NLP, Computer Vision, and ML/DL** solutions at scale. Founding member of MyPDI, PDI Technologies’ AI-powered platform serving 200,000+ retail locations. Proven track record of **current advancements in AI** and passionate about leveraging **cutting-edge technologies** to build innovative AI solutions that drive business growth. Experienced in **transforming complex data** into **actionable intelligence**, developing **scalable AI models**, and integrating both **ML/DL and GenAI** approaches to solve **real-world challenges**.

Skills

Development	<ul style="list-style-type: none">• Python, FastAPI, Flask, Pydantic, Docker, Kubernetes, Terraform, Git• AWS Sagemaker, EMR, ECS, EC2, Opensearch, Azure AKS, OpenAI
Data Science & Machine Learning	<ul style="list-style-type: none">• Pandas, Numpy, Scikit-Learn• PyTorch, Tensorflow
NLP and CV	<ul style="list-style-type: none">• HuggingFace Transformers, Spacy, vLLM, Ollama• OpenCV, Google OCR, Azure AI Services, YOLO
Generative AI	<ul style="list-style-type: none">• CrewAI, Agno, LangChain, LlamaIndex, LangGraph, MCP, A2A• ChromaDB, SingleStore, PGVector, Opensearch, Pinecone, Milvus• Agentic AI, Agents, RAG, Prompt Engineering, Pretraining, Finetuning, PEFT• OpenAI GPT, Claude, Google Gemini, LLaMa, DeepSeek, Qwen• Azure AI Foundry, AWS Bedrock, Vertex AI, Groq, Anthropic
Professional	<ul style="list-style-type: none">• Problem Solving, Team Leadership, Cross-functional Collaboration, Mentoring

Experience

OCT 2025 – PRESENT

Technical Lead – AI/ML & Innovation | PDI Technologies

1. **Founding Member – MyPDI AI Platform**

Leading all AI initiatives for MyPDI, PDI Technologies’ AI-powered platform that centralizes analytics, AI-driven insights, and operations for 200,000+ convenience retail and energy locations across 50+ countries, managing \$150B+ in annual transactions.

2. **Enterprise Agentic AI Framework**

Architecting and leading development of a config-driven agentic framework for building, deploying, and orchestrating AI agents across PDI’s retail, fuel, and enterprise product suite. Supports multi-agent orchestration, MCP integrations, tool use, and memory.

3. **RAG Pipelines & Intelligent Automation**

Building scalable RAG pipelines and chatbot systems to power AI-driven insights, personalized

recommendations, and automated workflows across PDI's convenience retail and petroleum wholesale ecosystem.

4. **ML Solutions for Enterprise Products**

Developing and deploying ML models for demand forecasting, price optimization, and operational efficiency across retail and fuel verticals, enabling data-driven decision-making at scale.

FEB 2024 – SEP 2025

Senior AI Developer II | Connectwise

5. **Built an autonomous SOC Analyst (Agno framework) with multi-agent orchestration for end-to-end threat analysis and mitigation. Reduced manual effort by 40% with 98% Recall.**
6. **Built an agentic ticket resolution system leveraging millions of historical tickets, device/user databases, IT docs, and RPA bots. Reduced average resolution time by 60% and support costs by 40%. Stack: Python, FastAPI, OpenAI GPT, AWS Kubernetes.**
7. **Designed a config-driven agentic AI framework with conversational capabilities, user memory, MCP integrations, and multi-tool orchestration via REST API.**
8. **Architected a low-code RAG framework supporting multiple LLM providers, vector DBs, and scalable data ingestion. Adopted by all internal teams.**
9. **Built and scaled RAG and prompt engineering apps (script generation, RPA bots, workflow planning, IT docs) serving millions of users across 3 regions.**
10. **Built an internal chatbot integrating HR policies, university resources, cybersecurity docs, and product knowledge bases. Reduced client support resolution time by 50%.**

JAN 2023 – FEB 2024

Data Scientist III | Draup

1. **Curie – RAG-based Talent Chatbot (Blog / Product)**
 - Spearheaded Curie using GPT-3.5/4, LangChain, and Chain-of-Thought prompting. Queried 400M+ job postings with 5–10s response times at under 7 cents/query.
 - Engineered the system to query 400M+ job postings with 5–10 second response times, driving high user satisfaction.
2. **JDLM – Talent Domain Specific Language Model**
 - Pre-trained an in-house RoBERTa model on 60M job descriptions (5%+ downstream gain). Reduced GPU training costs by 75% through efficient training techniques.
3. **Led 10+ knowledge-sharing sessions on LLM pretraining, instruction tuning, and parameter-efficient fine-tuning (PEFT).**
4. Established standardized code structure and data/model storage guidelines for ML projects across the team.

SEP 2021 – DEC 2022

Data Scientist II | Draup

- a. Built a hybrid collaborative/content-based recommendation system for talent reports (0.3 Precision@5) with online retraining pipeline.
- b. Revamped skill clustering algorithm to classify skills into 110+ classes.
- c. Diagnosed and resolved legacy data, model, and workflow issues to restore classification pipeline reliability.

OCT 2020 – SEP 2021

Senior ML/AI Developer | Accolite Digital

- i. Built real-time ID card localizer (SSD) and keypoint detector (0.85 R²) with CNN classifiers (90–95% F1) across 40+ card types in 13 countries.
- b. Built a GPU-based image processing pipeline (blur, sharpen, super-resolution) to enhance ID images for accurate OCR extraction.
- c. Implemented real-time OCR-based data extraction from 50+ ID card types, streamlining sign-up verification.
- Deployed Django REST microservices on GPU nodes using Docker and Azure AKS. Designed end-to-end ML pipeline architecture.

JUL 2019 – OCT 2020

ML/AI Developer | Accolite Digital

1. Developed a CNN-LSTM model for radiology report criticality classification, achieving 92% F1 and improving clinical prioritization.
2. Engineered an LSTM-based context-aware profanity detector achieving 97% F1 for real-time content moderation.
3. Led Mixed-GAN implementation using NVIDIA StyleGAN for 15th-century portrait generation (avatar creation).
 - a. Developed a Siamese network for real-time face matching with liveness detection (blink, smile, pose, iris-tracking). Sub-100ms inference via TensorFlow.js on mobile.

Education

AUG 2017 – JUN 2019

M.Tech. | COEP, Pune (Computer Science)

CGPA: 9.54

AUG 2012 – MAY 2016

B.Tech. | KJ Somaiya COE, Vidyavihar (Computer Science)

CGPA: 6.79

Projects

Weapon Detection and Classification using Deep Learning

- Led the creation of a comprehensive **dataset comprising 11,000 images**, with more than 20,000 instances of weapons enabling precise detection and classification of weapons in images and videos.
- Implemented state-of-the-art Object Detection models including **Faster-RCNN, YOLO, SSD**, etc., resulting in a top-performing model achieving **66.5% mean Average Precision (mAP) score**.

Achievements

- Star Performer Award (Q3 2024) at ConnectWise for outstanding contributions.
- **Fast Tracker Award (2022) at Draup for exceptional performance and rapid career progression.**
- Best Performer Award for three consecutive quarters (Q2'20, Q1'21, Q2'21) at Accolite Digital.
- **Winner, Open Pune Data Hackathon'19 (Pune Municipal Corporation & IEEE).**