



# Goutham Vignesh

## AI RESEARCH ENGINEER

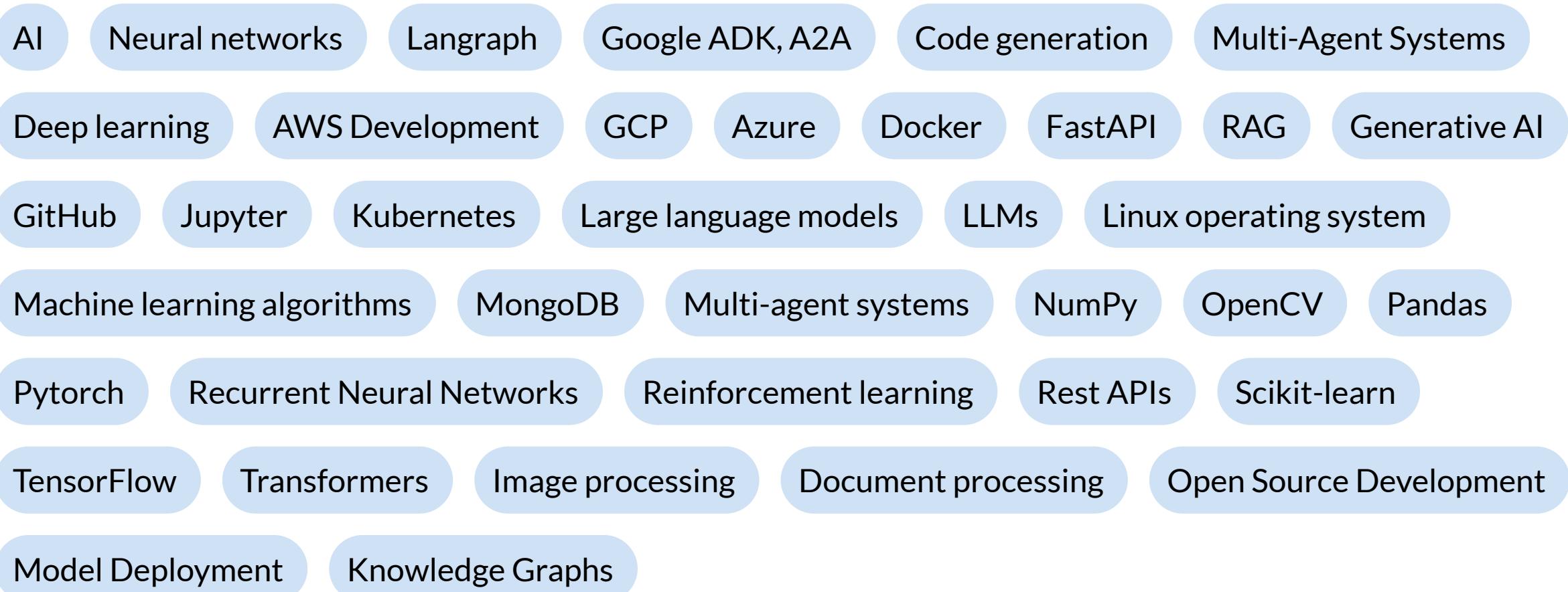
AI Research Engineer with 6 years of experience, specializing in multi-agent systems and large language models. Proficient in Python and deep learning frameworks, with a track record of developing AI-driven pipelines that accelerate app development and reduce costs. Skilled in delivering scalable AI solutions and managing end-to-end ML projects from requirements gathering and experimentation to solution design and production deployment in collaboration with cross-functional teams.

✉️ gouthamvigneshs@gmail.com ☎️ +91 9952689236 ⚙️ Bangalore, India

### WORK EXPERIENCE

- 01/2024 - Present **AI RESEARCH ENGINEER - APPLIED R&D UNIT**  
Quantiphi  
Bangalore
- Developed Qnowledge Engineer, AI pipeline for accelerating agentic development using natural language inputs.
  - Enabled collaborative agent workflows with A2A communication and MCP for efficient tool invocation using Google ADK and Langraph Agentic frameworks
  - Facilitated domain-specialised agent creation framework with custom skills for solving specific business problems.
  - Developed an agentic Coding accelerator Quantiphi's Codeaira platform with agents builder and skill planning.
  - Integrated multi-modal agentic workflows with LLMs in QDocs and Dociph, boosting document processing efficiency by 40%.
- 04/2023 - 01/2024 **RESEARCH ENGINEER (GEN AI)**  
SciSpace  
Bangalore
- Reduced training costs by 75% with LLM fine-tuning pipeline for custom datasets on open source models using PEFT and LORA techniques for multiple research tasks
  - Streamlined open source model deployment for faster inference using VLLM, NVIDIA Triton and inference optimization techniques
- 12/2019 - 03/2023 **MACHINE LEARNING ENGINEER**  
Vakilsearch  
Chennai
- Engineered RAG pipeline for legal document retrieval, enhancing accuracy in Vector DBs like Pinecone and Weaviate.
  - Achieved 90% accuracy in classifying address-proof documents with ensemble techniques, boosting document verification processes.
  - Developed Q&A chatbot with Retriever-Reader architecture, reducing backend work-load by 60% and improving response efficiency.
  - Developed robust code for scalable model consumption, boosting deployment efficiency by 30%.

## SKILLS



## PUBLICATIONS

- **IDPFlow**: A No-Code Agentic Framework for Multimodal Intelligent Document Processing | ACM Multimedia 2025
- **DocAnnot**: Accelerating Creation of Key Information Extraction Datasets | Book Chapter, 2026 .
- **DocGenie**: A Framework for High-Fidelity Synthetic Document Generation | Synthetic Data for Computer Vision Workshop @ CVPR 2025
- **SciSpace Copilot**: Empowering Researchers through Intelligent Reading Assistance | AAAI Conference 2024.
- **SciSpace Literature Review**: Harnessing AI for Effortless Scientific Discovery | ECIR 2024

## EDUCATION

2020	<b>Undergraduate</b>
Sri Krishna college of Engineering Coimbatore, India	Bachelor of Engineering - Computer Science

## LANGUAGES

- **English** - Advanced

## WEBSITES

- LinkedIn -<https://www.linkedin.com/in/goutham-vignesh-243988128/>
- Github - <https://github.com/GouthamVicky>