

# Sachin Bhat

Mobile No.: +65 8544 8214 | Email: [sachubhat17@gmail.com](mailto:sachubhat17@gmail.com)

LinkedIn: <https://www.linkedin.com/in/sbhat17> | Github: <https://github.com/Sachin-Bhat> | Gitlab: <https://gitlab.com/Sachin-Bhat>

## SUMMARY

Driven and innovative Undergraduate with hands-on experience in AI and system architecture at Dell Technologies and Schaeffler. Demonstrated expertise in developing robust AI solutions, agentic frameworks, enhancing system integrations, and leading cost-saving initiatives that significantly improved digital assistant functionalities. Proven problem-solving skills and a strong commitment to leveraging advanced technology to drive business transformation and operational efficiency.

## SKILLS & INTERESTS

- Programming/Scripting Languages: C, C++, Java, Python, R, Rust, Bash
- Databases: MariaDB (SQL), MongoDB, Redis, PostgreSQL, Qdrant, Apache AGE, Neo4j, FalkorDB, Kafka, RabbitMQ
- Productivity: git, SVN, Mercurial, Microsoft Office, LibreOffice, Microsoft Azure, Docker
- Operating Systems: Linux, RTOS, BSD, Solaris
- Frameworks: Langchain, PyTorch, CUDA, NumPy, SciPy, Pandas, Polars, FastAPI, Flask, Rasa, ROS, Gradio, Airflow, MLFlow
- Research Interests: Large Language Models (LLMs), NLP, Reasoning, Knowledge Extraction, Context Compression

## WORK EXPERIENCE

### Tamira Technologies

#### AI & Data Engineer (Python & Rust)

Jul 2025 - Present

- Led migration from Python microservice stack to **Rust-based agentic router**, which improved system throughput and reliability, resulting in 3× **latency reduction** and near-zero agent crashes.
- Created a patent-pending **declarative recipe** orchestration framework which improved context reuse and reduced unnecessary token usage, resulting in **40–60%** token cost savings across core workflows.
- Developed a secure, role-aware TUI with **global guardrails and permission enforcement** which improved operational safety, resulting in zero unsafe or privilege-escalated executions in testing.
- Implemented automated **recipe vulnerability scanner** which improved detection of unsafe instructions and execution paths, resulting in **70% reduction** in malicious or unsafe recipes during QA.

#### Part-time Intern, Innovation Engineer

Feb 2025 - Jun 2025

- Created patent-pending **Form Navigator** using NER + LLM pipelines, **improving** form comprehension and entity-extraction accuracy, resulting in **35–40%** reduction of operator input time.
- Built AI-powered **Work-Order Report Generator** which **improved** consistency and completeness of technician reports, **resulting in 25% decrease** in revision cycles.

### Schaeffler AG

Aug 2024 - Jan 2025

#### Part-time Intern, Innovation Engineer

- Developed advanced bin-picking solutions for industrial applications using state-of-the-art pose estimation models like FoundationPose and BuildSDF.
- Integrated open multimodal models, including Mistral and Llama-3.2, into **patent-pending** agentic task planner for complex scene understanding and behavior automation.
- Contributed to over **€1 million** project by designing a novel "Iteration of Thought" architecture supporting behavior-tree generation, code synthesis, and video-based imitation learning through natural language input.

### Dell Technologies

Jan 2024 - May 2024

#### Intern, Undergraduate Gen-AI Scientist

- Redesigned orchestration layer of Dell Support Digital Assistant, transitioning to FastAPI with asynchronous support, enhancing system performance and scalability.
- Enhanced NLU capabilities by integrating advanced rules and stories, improving precision and responsiveness of AI-driven customer interactions.
- Led development and integration of a custom Retrieval-Augmented Generation (RAG) fusion architecture, reducing hallucination rates to **0.3%**, boosting reliability of content generation.
- Architected and implemented an in-house Information Retrieval (IR) solution, phasing out a costly third-party service and saving company over **\$1.5 million** annually.
- Developed 3 innovative, **patent-pending** frameworks and products, positioning Dell to capitalise on advanced AI technologies and potentially saving millions in licensing fees.

## CO-CURRICULAR ACTIVITIES

### Open Source Contributor

Aug 2021 - Present

#### Volunteer

- Engineered critical features to Wayland Compositors including LabWC such as supporting configurations for workspaces, an automated pipeline to set-up Systemd environment variables, helping in system extensibility and playing non-native games on Linux.
- Created open-source [Agent Orchestration Framework](#) for Rust to power agentic router for my company, achieving over 200+ downloads within the first week of launch.
- Contributed over 1,000 lines of code to Langchain, implementing key features like SmartLLMChain, structured output parsing with LLMs, and advanced indexing strategies such as IVFFlat for pgvector, resulting in **15%** improvement in query performance and adoption by 50+ developers.

### WorldQuant LLC & Nanyang Technological University

Aug 2023 - Sep 2023

#### Worldquant Brain NTU X CAO Alphathon

- Participated in NTU's first Alphathon by regularly researching and submitting Alphas on WorldQuant's BRAIN platform.
- Achieved Gold League by ranking within TOP 50 of participants due to minimising turnover and increasing Sharpe and achieving an average fitness of **1.5** within in-sample tests.

## PROJECTS

### [Final Year Project: Exploration on Effect of Chunking for Downstream RAG Tasks](#)

Aug 2024 - May 2025

#### Skills utilised: git LFS, Python, Rust, Docker, Qdrant, RAG

- Developed Rust-based TUI, **Unarxiver**, to efficiently download and manage research papers from sources like ArXiv.
- Created **Tattva**, a PDF-to-Markdown extraction pipeline achieving **98%** multimodal extraction accuracy by integrating models like DocLayoutYOLO, LayoutLMv3, and heuristic-driven rules.
- Implemented advanced OCR techniques using PaddleOCR and extracted tabular data with StructEqTable and TableMaster.
- Designed Python library for customisable chunking strategies and evaluated methods like Recursive Text Splitting and Semantic Chunking against proposed Schematic Chunker on HuggingFace datasets
- Proposed chunking strategy led to **5%** improvement on simple tasks and **10% improvement with minimum deviation** on complex multi-turn and multi-hop retrieval efficiency calculated as combination of **5 different metrics**.

### [Neural Memory Retrieval: Custom Retrieval through Neural Memory](#)

Mar 2025 - May 2025

#### Skills utilised: Python, LLMs, Information Retrieval, Neural Memory, VectorDBs, CLIP, LangChain, RAGAS, Multi-Hop Reasoning

- Designed and implemented novel multi-component retrieval architecture (**NeuralMemoryRetriever**) integrating abstract neural memory, concrete contextual memory, and compressed working memory to enhance LLM reasoning over long contexts **>2M tokens**.
- Developed key components including **stateful neural memory module** based on titans-pytorch for dynamic, learned guidance and LLM-based compression mechanism (**LightThinkerCompressor**) for effective context management.
- Evaluated architecture's performance on **3** diverse datasets (CoQA, ai-arxiv2, TextVQA) using RAGAS framework, achieving strong results for factual retrieval on ai-arxiv2 with mean Answer Relevancy of **0.997** and mean Context Precision of **0.843**, and authored a detailed report, resulting in publication at **ICDM - SENTIRE**.

### [Model Context Protocol \(MCP\): Simple Terminal Language Assistant for Linux Users](#)

Mar 2025 - May 2025

#### Skills utilised: Python, Bash, MCP, Ollama, Claude Desktop, TUI (Textual), Agentic Automation, Linux

- Built a personal terminal assistant using MCP to enable local LLMs (e.g., Ollama, Claude) to access shell history, filesystem, and live system metrics resulting in **over 30 downloads** within **1 week** of launch on Smithery.
- Integrated with Claude Desktop, Goose and custom MCP tools to allow natural language queries to translate to Linux commands
- Implemented system call wrappers and command analyzers in Python to safely interpret and execute shell operations based on AI recommendations.
- Designed an intuitive TUI interface for interacting with the assistant using Textual for smooth UX in the terminal.
- Emphasized **privacy and local-first design** with Ollama, ensuring secure access to sensitive command-line environments.

### [Natural Language Processing: Twitter Tweet Sentiment Analysis](#)

Jan 2022 - May 2022

#### Skills utilised: Python, Jupyter, Analytics

- Retrieved and engineered data from the Twitter API by writing a web scrapper with BeautifulSoup and requests, driving project efficiency by at least **10%**.
- Performed data cleaning, exploratory data analysis and fitted Logistic Regression, Random Forest, Support Vector Machine and XGBoost models on tweets with Pandas, Numpy, TensorFlow and Scikit-Learn.
- Attained distinction grade by applying multiple evaluation metrics and further back-testing on additional datasets.

## EDUCATION

### Nanyang Technological University, Singapore

Singapore, SG

#### Bachelor of Science in Data Science and Artificial Intelligence

### Legacy School Bangalore

Bengaluru, IN

#### International Baccalaureate Diploma