

# Sagar Chhabriya

+92-332-336-2473 • Karachi, Pakistan • [sagarchhabriya6@gmail.com](mailto:sagarchhabriya6@gmail.com) • [sagarchhabriya.github.io](https://sagarchhabriya.github.io) • [linkedin.com/in/sagar-chhabriya](https://linkedin.com/in/sagar-chhabriya) • [github.com/SagarChhabriya](https://github.com/SagarChhabriya)

## Research Interests

My research lies at the intersection of NLP, Information Retrieval, and Machine Learning. I study agentic RAG systems that enable autonomous question-answering and knowledge discovery. I am particularly interested in text mining for insight extraction from unstructured data and retrieval systems that enhance efficiency and relevance in large-scale digital environments.

## Experience

### Machine Learning Teaching Assistant @ [atomcamp](#)

Islamabad, Pakistan

Aug 2025 - Present

- Conduct weekly live sessions on ML, DL, NLP, and MLOps, combining theory with hands-on coding via Google Meet
- Lead end-to-end ML projects, teaching Python, Git/GitHub, Streamlit, and FastAPI; mentored students in Dockerization and CI/CD deployment on Streamlit Cloud
- Lead MLOps implementation modules, covering data acquisition, model training, serialization, experiment tracking (Weights & Biases), and automated CI/CD workflows leveraging GitHub Actions + Google Cloud (Artifact Registry, Cloud Run)
- Mentor ~30 learners; 25+ deployed ML projects with improved performance via hyperparameter tuning and feature engineering
- Provide 1:1 mentorship and code reviews via official communication channels (WhatsApp and email), ensuring consistent student progress
- Design learning modules on ML, DL, CV, Time Series, Generative AI, and Agentic AI for Atomcamp Arabia (KSA international bootcamps)

### NLP Research Assistant (Full-time) @ [Fast Nuces](#)

Karachi, Pakistan

June - Aug 2025

Project: Automatic multi-modal classification of religious hate content from social media (HASOC 2025) – NRPU 16153

- Contributed to HASOC 2025 on multilingual hate speech detection from OCR-extracted meme text in Hindi, Romanized Hindi, and English
- Engineered custom multilingual lexical resources (vulgar word and stop word dictionaries) that improved model performance by 2–5% across classification tasks
- Fine-tuned Multilingual BERT (mBERT) and benchmarked against ML baselines (SVM, Naive Bayes) to enhance multilingual text classification accuracy
- Conducted a comprehensive literature review on hate speech detection and multilingual NLP to inform model design and evaluation strategies
- Leveraged GPU-accelerated deep learning frameworks (TensorFlow & PyTorch) for large-scale experimentation and evaluation

## Education

### Bachelor of Science in Computer Science

Sukkur, Pakistan

Sukkur IBA University, Sukkur, Pakistan | CGPA: **3.67** / 4.0

Sept 2022 – Sept 2026

## Publications

### Agentic Retrieval-Augmented Generation: A Compact Review

Karachi, Pakistan

- Presented at [1<sup>st</sup> ICCA @ Salim Habib University](#)
- Sagar Chhabriya\*, [Sumra Khan](#), Dr. Rizwan Ahmed Khan, [Mansoor Ahmed](#)

May 2025

## Skills

**Programming Languages & Tools:** Python • SQL (MySQL) • MongoDB • Git • GitHub • GitHub Codespaces/Actions • Google Cloud Run • Vertex AI • Docker • DVC

**GenAI & LLMs:** Large Language Models • RAG • ReAct • LangChain • LangGraph • LangSmith • CrewAI • Chroma • FAISS • Pinecone • Unislosh

**Libraries and Frameworks:** Scikit-learn • TensorFlow • Keras • PyTorch • NLTK • spaCy • Streamlit • FastAPI • Flask • MLflow • Wandb • Hugging Face Transformers

## Projects

### CallRAG – RAG based Chatbot

[github.com/SagarChhabriya/LLMs/tree/main/callrag](https://github.com/SagarChhabriya/LLMs/tree/main/callrag)

- Developed a RAG-based chatbot using LlamaIndex and Gemini 1.5 Flash for accurate, context-aware responses, deployed on Streamlit Cloud
- Enhanced user engagement by achieving a 30% faster response time and a 25% improvement in accuracy for user queries

### Salary Predictor – An End-to-End Machine Learning Application

[github.com/SagarChhabriya/ML-Playground/tree/main/01-salary-predictor](https://github.com/SagarChhabriya/ML-Playground/tree/main/01-salary-predictor)

- Developed and deployed a salary prediction web application with linear regression and Streamlit, resulting in a 90% R<sup>2</sup> score and enabling cross-platform compatibility through model serialization
- Enhanced model accuracy by 10–15% through comprehensive data preprocessing, including missing value imputation and feature scaling

### Ragwiki – Wikipedia RAG Assistant

[github.com/SagarChhabriya/RAG/tree/main/apps/01-wikipedia-rag-langchain](https://github.com/SagarChhabriya/RAG/tree/main/apps/01-wikipedia-rag-langchain)

- Developed a Wikipedia RAG chatbot using LangChain, FAISS, and Google Gemini 1.5 to deliver context-aware answers from top-3 relevant Wikipedia articles
- Improved response relevance and user trust by 30% with document caching, vector-based retrieval, and source attribution features in a Streamlit interface

### LangChain Summarizer

[github.com/SagarChhabriya/RAG/tree/main/apps/02-langchain-summarizer](https://github.com/SagarChhabriya/RAG/tree/main/apps/02-langchain-summarizer)

- Built a Streamlit app using LangChain and Google Gemini 2.0 for customizable paper summaries, improving user clarity by 20%
- Integrated streaming API for real-time generation and crafted dynamic prompt templates for accurate context-aware summaries, boosting summary accuracy by 15%

### Real-Time A.I Voice Assistant

[github.com/SagarChhabriya/JarvisCA](https://github.com/SagarChhabriya/JarvisCA)

- Developed a voice-controlled virtual assistant for web searches, Wikipedia lookups, and YouTube music using Python and various libraries
- Improved real-time response efficiency and query accuracy through optimized NLP pipelines and asynchronous API handling.

## Certification

Data Science and AI Bootcamp	—	A four-month Data Science and AI Bootcamp by <a href="#">atomcamp</a> , gaining expertise in Python, Business Intelligence, Machine Learning, Deep Learning, NLP, Computer Vision, LLMs, Generative AI, and MLOps, with hands-on experience in data pipelines and AI agent development
<a href="#">GitHub Foundations Exam</a>	—	Certified in Git and GitHub, covering repositories, commits, branching, markdown, project management, and collaboration features
Excel Fundamentals Skills Track	—	Five-course accredited program by DataCamp covering data visualization, preparation, and analysis with Excel 365

## Achievements

Book Review Competition 2022 Winner  
Sukkur IBA University

Prime Minister Laptop Award  
Sukkur IBA University • PMYL Scheme 2023

Master Trainee - Prevent Violent Extremism  
EDC-Sukkur IBA • NACTA • EU Funded