



Sahil Kale

Location: Los Angeles, California

Phone: (+91) 7391986368

Email address: sahilrkale@ucla.edu



[LinkedIn](#)



[Google Scholar](#)



[Research Map](#)

Work experience

Mastercard

Software Engineer II | Jan 2026 - Aug 2026 | Pune, India

- Redesigned the entire data architecture for the Click2Pay and Secure Card-on-File systems, serving over **14 million transactions per month**

Software Engineer I | Jun 2024 - Jan 2026 | Pune, India

- Integrated **AI agents into Apache data flows**, enabling automation and concurrency and resulting in a nearly **35% reduction** in transaction settlement and clearance time
- Supported development of distributed **automated billing batches** handling 1 million+ transactions in 6 global regions

KnowledgeVerse AI

AI Research Consultant | Jun 2024 – Aug 2026 | Atlanta, United States (Remote)

- Led 5+ research projects** on creating novel LLM benchmarks focused on trustworthiness and explainability of NLP systems in real-world deployment published in notable venues such as **ACL and NAACL**
- Built **over 80 custom agentic AI workflows** to maximize trust and reliability for critical client applications

Stride.ai

Applied Scientist Intern (NLP) | Sep 2023 – Jun 2024 | Bangalore, India

- Led and architected a **scalable production NLP system (processes 40,000 words in under 2 minutes)** for generating domain-specific FAQs across 17 fields that can use domain jargon and specific constructs

Barclays

Software Engineer Intern | Jun 2023 – Aug 2023 | Pune, India

- Performed SQL partitioning to **reduce query time to less than 89% of original** for post-trade databases
- Implemented automation and **monitoring of over 90 million transactions recorded on 35 distributed servers**

Association for Socially Applicable Research

Data Analyst | Mar 2023 – Jun 2023 | Pune, India

- Designed workflows for data ingestion to build a dashboard and system allowing **easy access to over a million emergency health centres in India**

Education & Training

M.S. Computer Science | University of California, Los Angeles (UCLA) | September 2026 – June 2028

- Masters' degree focused on **NLP and Artificial Intelligence** at the Henry Samueli School of Engineering

Lisbon Machine Learning Summer School (LxMLS 2025) | Instituto Superior Técnico | July 2025

Grade: 4.0/4.0

- Invited with a full **\$1500 Google scholarship** to present novel research on LLM self-knowledge in a PhD school

B.E. Computer Engineering | Pune Institute of Computer Technology | July 2020 – April 2024

Grade: 9.78/10

- Recipient of the Late Ajay Prabhakar Nene Award and fellowship for achieving the **1st rank among 8200 students in the B.E. undergraduate program** in the entirety of the University of Pune

Skills

Artificial Intelligence | Natural Language Processing | Large Language Models | Machine Learning | Agentic AI | Model Fine-Tuning | Model Benchmarking

Programming Languages | C++ | Python | JavaScript | Unix

Database Management | [SQL](#) | [SQL Server](#) | [NoSQL](#)

Data Science and Engineering | [Apache Hadoop](#) | [Apache NiFi](#) | [Apache Spark](#) | [Big Data](#) | [Data Analysis](#)

Web Development | [HTML](#) | [CSS](#) | [FastAPI](#) | [Flask](#) | [JavaScript Libraries](#)

Cloud Computing | [AWS](#) | [Microsoft Azure](#)

Publications

Local Information Access in Marathi: Evaluating LLM-Native Web Retrieval in a Low-Resource Environment

49th International ACM SIGIR Conference 2026, Melbourne - <https://dl.acm.org/doi/abs/10.1145/3805712.3808534>

- Identified challenges in AI-backed information retrieval for under-represented languages isolating failure modes in translated global queries and natively crafted local information needs

KnowRL: Teaching Language Models to Know What They Know

Submitted to NeurIPS 2026, Atlanta - <https://arxiv.org/abs/2510.11407>

- Presented a framework that strengthens a model's internal understanding of its own feasibility boundaries such that with just a small seed set and no external supervision, our method drove gains of 28% in accuracy and 12% in F1

Lie to Me: Knowledge Graphs for Robust Hallucination Self-Detection in LLMs

ICAART Conference 2026, Marbella - <https://www.scitepress.org/PublicationsDetail.aspx?ID=RcMLaWCJ6gk=&t=1>

- Demonstrated that structuring LLM outputs as knowledge graphs significantly improves hallucination self-detection, achieving up to 16% higher accuracy and 20% better F1-score over existing methods

Mirage of Mastery: Memorization Tricks LLMs into Artificially Inflated Self-Knowledge

AAAI IASEAI Conference 2026, UNESCO Paris - <https://ojs.aaai.org/index.php/IASEAI/article/view/43032>

- Showed a noteworthy problem in LLM generalization: models draw confidence from memorized solutions to infer a higher self-knowledge about their reasoning ability, seen as an over 45% inconsistency in feasibility assessments

Line of Duty: Evaluating LLM Self-Knowledge via Consistency in Feasibility Boundaries

NAACL Conference 2025, Albuquerque - <https://aclanthology.org/2025.trustnlp-main.10>

- Obtained intrinsic insights into different types of LLM self-knowledge with a novel methodology and found that even frontier models like GPT-4o and Mistral Large are not sure of their capabilities more than 80% of the time!

TeXpert: A Multi-Level Benchmark for Evaluating LaTeX Code Generation by LLMs

5th Workshop on Scholarly Document Processing @ ACL 2025, Vienna - <https://aclanthology.org/2025.sdp-1.2>

- Proposed TeXpert, our benchmark dataset with natural language prompts for generating LaTeX code focused on components of scientific documents across multiple difficulty levels, showing how poorly LLMs perform

FAQ-Gen: An automated system to generate domain-specific FAQs to aid content comprehension

Journal of Computer-Assisted Linguistic Research, UPV - <https://polipapers.upv.es/index.php/jclr/article/view/21178>

- Proposed an FAQ generation system leveraging custom built and fine-tuned text-to-text transformation models and utilising self-curated algorithms for cognitive ranking of the question-answer pairs

Languages

[English](#) | [Marathi](#) | [Hindi](#) | [German](#) | [Japanese](#)

Certifications

Large Language Model Operations (LLMOps) Specialisation | [Duke University](#) | 2025

Certification in Building LLM-Powered Applications | [Weights and Biases](#) | 2024

Specialisation in Natural Language Processing | [Great Learning](#) | 2023

Machine Learning with Python: Gradient Boosting and Advanced Techniques | [Jovian AI](#) | 2022

Honours and Awards

Zurich GenAI Awards 2026 - Winner | GenAI Zurich | 2026

Awarded as one of the top 3 projects among 600+ submissions in the Impact Achiever category for solutions that use generative AI to deliver meaningful societal benefits through a jury vote

FAO China Youth Hackathon on Urban Agriculture - Winner | UN FAO | 2024

Won first place among 200+ entries from over 50 UN member nations for building an efficient AI-powered audio-based data collection and analysis mechanism to document ground-level farmer challenges

Smart India Hackathon - Winner | Government of India | 2023

Achieved first place from 5000+ entries with a \$1000 prize for building a "Contact Center Knowledge Management Tool" for the Ministry of Commerce and Industry, Government of India in a 36-hour non-stop hackathon

PICT InC Project Competition - Winner | Pune Institute of Computer Technology | 2023

Won first place and a \$500 cash prize from among 200+ entries in an international level project competition sponsored by eQ Technologic for developing a novel AI-powered scheduling assistant

Projects

Sampark Sarita: AI Knowledge Management - <https://github.com/Sahil-R-Kale/Sampark-Sarita-SIH>

- An **agentic AI-powered question-answering tool** built for the Ministry of Commerce and Industry, India
- Reduced **query resolution time by 32.5%** with a dashboard that guides agents to the perfect series of questions and follow-up answers with real-time suggestions and multilingual voice integration

ClashFree: AI-Powered Scheduling Assistant - <https://github.com/Sahil-R-Kale/Clash-Free>

- An automatic, user-friendly schedule builder equipped to solve **scheduling conflicts up to 1-minute granularity**
- Provides **automatic suggestions using a genetic AI algorithm**, drag-and-drop UI for human-in-the-loop validations and a scalable database integration for future reference

Finterest: AI-Powered Financial Planning - <https://github.com/Sahil-R-Kale/Finterest.git>

- A personal financial planner providing AI-powered personalised **portfolio diversification based on 5 parameters of risks**, along with short and long-term investment goal planning
- Provides a precise and customised report using cloud services suggesting **10+ diversified investment options**

Volunteering

Coding Tutor | Young Thinks Foundation | Mar 2023 – May 2023

Helped underprivileged students in rural India improve their analytical and technical skills in Python and C++ through practical mentoring sessions

English Tutor | Read-a-Story Foundation | Sep 2022 – Feb 2023

Conducted 1:1 biweekly tutoring to improve reading and spoken English in the remote regions of India with an outcome of improved word retention by 42%