Lakshya A Agrawal

Ph.D. Student, BAIR and Sky Lab
University of California, Berkeley

lakshyaaagrawal@berkeley.edu, lakshya.aagrawal@gmail.com
Website GitHub LinkedIn Google Scholar

Education

University of California, Berkeley *2024 - Present*
Ph.D. in Computer Science
Advisors: Prof. Matei Zaharia and Prof. Dan Klein

IIIT-Delhi *2018 - 2022*
B.Tech in Computer Science and Applied Mathematics **(Department Rank: 1)** GPA: **9.55/10**

Work Experience

Microsoft Research
AI4Code Research Fellow (Prev. Research Intern) *Jan 2022 - Aug 2024*
Advisors: Dr. Aditya Kanade, Dr. Navin Goyal, Dr. Shuvendu Lahiri, Dr. Sriram Rajamani

- Exploring novel evaluations and techniques for improvement in quality and correctness of code generation with LLMs.
- Proposed technique for code generation grounded in repository having certifiably rich correctness properties like valid sequence of method calls (typestates), valid number of arguments, absence of hallucinated dereference symbols, etc.
- **20-25% improvements in compilability of LLM generated code without any modification to the models.**
- Proposed first static analysis based retrieval augmented prompt (RAG) for code generation. **Accepted at NeurIPS '23.**

Very Large Scale Computing Laboratory, EPFL
Summer@EPFL Research Fellow *July 2021 - Dec 2021*
Advisors: Dr. Endri Bezati, Prof. James Larus

- Developed streamblocks-graalvm, a CPU based runtime for CAL dataflow language, based on Truffle/GraalVM.
- Implemented IDE and debugger support (with code stepping) for the CAL programming language over LSP and DAP.

Microsoft
Software Development Intern *May 2021 - July 2021*
Team: Orchestration as a Service, Azure Compute

- Member of team handling Azure-wide orchestrations: running mitigations safely and securely during livesites.
- Feature integrations for critical security components achieving Azure safe deployment practices (SDP) compliance.

Google Summer of Code
Open Source Developer @ INCF *May 2019 - Aug 2019*
Advisors: Dr. Dimiter Prodanov, Dr. Robert Dodier

- Developed Pytranslate, a transpiler in Common Lisp, that translates Maxima symbolic computation code to Python, including support for translating 2D and 3D plot functions. **Included as a part of all Maxima installations** since 2019.

Publications

Why Do Multi-Agent LLM Systems Fail?
Mert Cemri*, Melissa Z. Pan*, Shuyi Yang*, **Lakshya A Agrawal***, Bhavya Chopra, Rishabh Tiwari, Kurt Keutzer, Aditya Parameswaran, Dan Klein, Kannan Ramchandran, Matei Zaharia, Joseph E. Gonzalez, Ion Stoica
Preprint arXiv:2502.20315, 2025.

LangProBe: a Language Programs Benchmark
Shangyin Tan, **Lakshya A Agrawal**, Arnav Singhvi, Liheng Lai, Michael J Ryan, Dan Klein, Omar Khattab, Koushik Sen, Matei Zaharia
Accepted for ACL ARR 2025 February Cycle.

Monitor-Guided Decoding of Code LMs with Static Analysis of Repository Context.
Lakshya A Agrawal, Aditya Kanade, Navin Goyal, Shuvendu Lahiri, Sriram Rajamani.
*Neural Information Processing Systems (**NeurIPS**), 2023*

A SPARQL to Cypher Transpiler: Proposal and Initial Results. Extended Abstract
Lakshya A Agrawal, Nikunj Singhal, Raghava Mutharaju.
*International Conference on Data Science and Management of Data (**CODS-COMAD**), 2022*

A Novel Sentiment Analysis Engine for Preliminary Depression Status Estimation on Social Media.
Sudhir Kumar Suman, Hrithwik Shalu, **Lakshya A Agrawal**, Archit Agrawal, Juned Kadiwala.
Preprint arXiv:2011.14280, 2020.

Awards and Honors

- **First Position at Microsoft Global Hackathon** challenge on AI powered tools for developer productivity *2023*
- (News Coverage) All Round Performance Medal for **Best Academic & cocurriculars** - CSAM batch 2018-22. *2022*

- Summer@EPFL 2021 & 2020 scholarship **(top 1% of applicants)** *2020, 2021*
- Dean of Academics Award for Academic Excellence for years 2020-21 & 2018-19. IIIT-Delhi *2019, 2021*

Selected Research Projects

dspy.GRPO: First OSS multi-module, multi-step GRPO pipeline *March 2025 - Present*

- Developed 1st GRPO pipeline for tuning modular agents, including complex compound AI Systems that compose multiple structured and specialized LM calls and tool invocations.
- Created server-client abstraction, decomposing GRPO policy gradient updates and multi-stage rollouts for efficiency.

Automatic Code Vectorization *Oct 2024 - Present*

- Building LLM based multi-turn code generation pipeline to vectorize C++ code to kernels targetting specific AI chip.
- Developed search based prompt tuner for fast code generation outperforming MIPROv2 (SOTA Prompt Optimizer).
- Successfully post-trained Qwen-2.5-14B-Coder using GRPO for code vectorization with just 40 datapoints.

ANNS-based KV retrieval for long-range attention over large code repos *Aug 2022 - Oct 2022*

- Motivation: Code repositories in enterprise settings can be very large, and not fit in the model context windows.
- Augmented SantaCoder and CodeGen models, to retrieve Key/Value values during attention from large vector stores using FAISS-based ANNS technique. Promising initial results of better code style match and lower symbol hallucinations.

SPARQL to Cypher: A Transpiler for Knowledge Graph Query Languages *Jan 2021 - May 2022*
Advisor: Dr. Raghava Mutharaju, *IIIT-Delhi*

- Work to unify two different knowledge graph data models and their query languages: RDF/SPARQL and PG/Cypher.
- Proposed novel technique based on SMT solvers to infer facts on Cypher query structure from input SPARQL query.

Selected Development Projects

multilspy : Batteries included LSP client library in Python (HackerNews Frontpage) *Jan 2024 - Present*

- Language Server Protocol (LSP) provides a headless interface to interact with IDEs and code analysis tools.
- Provides easy & unified access to various static analyses for code in different supported languages, like type-directed code completion, symbol definition, references, etc. Includes in-built support for static analyses of 4 major languages.
- **Deployed in production by several AI4Code startups**. multilspy makes it very easy to develop IDE tools for LLMs.

Spoon : Java Metaprogramming Library by INRIA *Aug 2020 - Oct 2020*
Advisor: Prof. Martin Monperrus

- OSS Contributions to Spoon (library for analyzing and transforming Java source code) to add new features and fix bugs.
- Implemented TextBlock support (JEP 355) through Eclipse JDT, thus completing Java 15 support in Spoon.

CovidReliefBot : Chatbot for COVID resource information *Apr 2021 - May 2021*

- Developed a chatbot to aid volunteers in resolving COVID related resource requests piloted with **up to 10,000 members.**
- Transcribed information images and aggregate relevant results from various data sources like Twitter, Google Sheets, etc.

Invited Talks

- **AI-Driven Code Performance Optimization for New Accelerators**
  - AMD What's New In Labs Talk Series *Jan 2025*
- **Guiding Language Models of Code with Global Context using Monitors**
  - Microsoft Research RiSE Group, Microsoft Research India Lab, Microsoft DevDiv *July, Aug 2023*
- **CAL Implementation in GraalVM**
  - Very Large Scale Computing Lab, Data Center Systems Lab @ EPFL *Sept 2021*

Professional Responsibilities

- Conference Reviewer - ICLR 2025, ISEC 2024
- Undergraduate Teaching Assistantship @ IIIT-Delhi
  - Machine Learning (CSE343/ECE563) taught by *Dr. Jainendra Shukla* *Monsoon 2021*
  - Theory of Computation (CSE322) taught by *Dr. Debajyoti Bera* *Winter 2021*

Skills

Programming Languages: C#, C++, Common Lisp, Cypher, F*, Java, JavaScript, Python, SPARQL
Tools and Technologies: ANTLR, CodeQL, Docker, FAISS, Flask, Git, GRPO, HuggingFace, JavaFX, LLVM Passes, MongoDB, PostgreSQL, Pytorch, Soot, Spoon, Language Server Protocol, WebRTC, Z3