Lakshya A Agrawal

Research Fellow, AI4Code Team
Microsoft Research

✉ t-lakagrawal@microsoft.com, lakshya.aagrawal@gmail.com
🏠 Website GitHub in LinkedIn 🎓 Google Scholar

EDUCATION

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 - Present*
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 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

📄 **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 productivity   *2023*
- (News Coverage) Institute Silver Medal for **Best Academic Performance** - CSAM batch 2018-22.   *2022*
- (News Coverage) All Round Performance Medal for **Best Academic & cocurriculars** - CSAM batch 2018-22.   *2022*
- Summer@EPFL 2021 & 2020 **(top 1% of applicants)** scholarship   *2020, 2021*
- Dean of Academics Award for Academic Excellence for years 2020-21 & 2018-19. IIIT-Delhi   *2019, 2021*

SELECTED RESEARCH PROJECTS

Better tokenizers for low-resource languages with existing LLMs   *Oct 2023 - Present*
Collaborator: Priyanshu Gupta, *PROSE Team, Microsoft*

- Motivation: Byte-Pair Encoding results in ∼10x slowdown in processing low-resource languages with LLMs.
- Promising initial results through architecture augmentation of pretrained LLM with expanded tokenizer vocabulary.

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.

Understanding Developer Use of Assertions for Java
Jan 2021 - May 2022

Advisor: Dr. Rahul Purandare, *IIIT-Delhi*

- Built customized AST scanners to mine assertions from large number of GitHub repos and compute code complexity.
- Used n-gram based modeling to cluster assertions. Uncovered positive correlation in code complexity & assertion count.

SELECTED DEVELOPMENT PROJECTS

multilspy : Batteries included LSP client library in Python
Feb 2023 - May 2023

Advisors: Dr. Aditya Kanade, Dr. Navin Goyal, Dr. Shuvendu Lahiri, Dr. Sriram Rajamani

- 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.

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.

Sampark: Data Survey App for Emergent tech users
Sep 2023 - Oct 2023

Collaboration: Ekal Vidyalaya

- Led development of data survey application targeting first-time technology users in remote and rural areas.
- Achieved over **3000 registered users and more than 1500 active users** in rural provinces of Chhattisgarh, India.

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.

CoVideo : Low bandwidth lecture system
May 2020 - Aug 2020

- Developed a system for conducting and recording live lectures, with support for whiteboard tools and slide deck import.
- **Reduced bandwidth requirement by up to 20x** compared to sharing Google Meet lecture recordings.

TALKS

- **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 - ISEC 2024 — *2023*
- 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*
- Administrator, Byld - Institute Software Development Club — *2019 - 2022*
- Student Senate Representative - Computer Science and Applied Mathematics, IIIT-Delhi — *2018 - 2019*

RELEVANT COURSEWORK

Advanced Programming, Algorithm Design & Analysis, Computer Architecture and Operating Systems, Discrete Structures, Data Mining, Linear Algebra, Linear Optimization, Machine Learning, Program Verification, Probability & Statistics, Programming in Haskell, Statistical Inference, Stochastic Processes and Applications, Theory of Computation

SKILLS

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