Alex **MATHAI**
Computer Science Researcher

alexmathai.com Google Scholar linkedin.com/in/alex-mathai-403117131
alexmathai98@gmail.com New York, USA

Research Interests

I am passionate about solving challenging research problems. My areas of interest include **LLM Agents**, **LLM training**, **ML for structured data like programming languages and graphs**, **knowledge graphs** and **graph neural networks**.

I am currently a PhD student at Columbia University and am grateful to be advised by Prof. Baishkahi Ray, Prof. Junfeng Yang, and Prof. Suman Jana.

Education

Years	Degree	GPA	Rank
2023-27 (expected)	PhD in Computer Science, Columbia University	4.25 / 4	–
2016-20	B.E. in Computer Science, Birla Institute of Technology and Science Pilani	9.63 / 10	6^{th}/123

Awards and Scholarships

Year(s)	Award
2024-25	**Amazon AI** trusted challenge to help train code LLMs that generate secure code.
2023-25	**Google Cyber NYC** awards to pursue research (with **Google Deepmind**) at the intersection of AI and Security
2022	**IBM Research Distinguished Paper Award** Given to the top 2 papers from IBM Research India and Singapore
2016-20	**Institute Merit Scholarship** Awarded to top 3% students for Exceptional Academic Performance

Research Papers

KGym : A Platform and Dataset to Benchmark LLMs on Linux Kernel Crash Resolution NeurIPS 2024
Alex Mathai, Chenxi Huang, Petros Maniatis, Aleksandr Nogikh et al.

COMEX : A Tool for Generating Customized Source Code Representations ASE 2023
Debeshee Das, Noble Saji Mathews, **Alex Mathai**, Srikanth Tamilselvam et al.

Monolith to Microservices : Representing Application Software through Heterogeneous GNNs IJCAI 2022
Alex Mathai, Sambaran Bandyopadhyay, Utkarsh Desai, and Srikanth Tamilselvam.

Incremental Analysis of Legacy Applications Using Knowledge Graphs for App Modernization CODS-COMAD 2022
Saravanan Krishnan, **Alex Mathai**, Amith Singhee, Atul Kumar, et al.

ReCLive : Real-Time Classification and QoE Inference of Live Video Streaming Services IEEE IWQOS 2021
Sharat Chandra Madanapalli, **Alex Mathai**, Hassan Habibi Gharakheili, and Vijay Sivaraman.

Adversarial Black-Box Attacks On Text Classifiers Using Genetic Optimization Guided By DNNs Arxiv, 2020
Alex Mathai, Shreya Khare, Srikanth Tamilselvam, Senthil Mani

US Patents

Automatically identifying application programming interfaces based on application artifacts, 2023 Patent
A patent to automatically extract API signatures from old monolithic code.

System and Method to explain Candidate Microservices from monolith, 2022 Patent
A patent that uses explainable AI models like **GNN explainer** to justify the microservice recommendations we make when decomposing monoliths.

Heterogeneous graph generation for Application Microservices, 2021 Patent
A patent for the creation of a heterogeneous graph from an application codebase. Clustering is then performed on this graph to create multiple smaller clusters representing microservices.

Industrial and Academic Research Experience

Current – Sept 2023 | **PhD Student, Columbia University | LLM Agents for million line+ codebase debugging, NYC, USA**

- I am currently designing effective LLM Agents that resolve bugs in massive million line+ system software codebases like the Linux kernel. This is in collaboration with **Deepmind :** Petros Maniatis and **Google :** Franjo Ivančić and Aleksandr Nogikh
- We aim to show that LLMs can be effectively used even in challenging situations, i.e., in low-resource languages, in complex codebases, and for tough-to-debug crashes.

LLMs | Agentic Frameworks

Aug 2023 – Aug 2020 | **Research Engineer, IBM Research | Modernizing Monoliths to Microservices , Bangalore, India**

- Worked with Amith Singhee on the knowledge graph (KG) modeling of large monolith applications, and with Srikanth Tamilselvam on KG partitioning to generate candidate microservices.
- Played a pivotal role in (i) **KG construction** using static analysis, (ii) **KG Inference** using graph traversal algorithms and (iii) **KG Partitioning** using graph neural network based representation learning coupled with unsupervised clustering to generate groups (microservice recommendations).

Java | Python | Neo4j | SQL | Pytorch

Dec 2019 – Aug 2019 | **University of New South Wales (UNSW) | Time Series Analysis Research, Sydney, Australia**

- Worked with Vijay Sivaraman and Hassan Habibi on creating deep learning models that can **classify** and isolate **live streaming** network traffic in **real-time** for 5000 Australian homes.
- Played a lead role in crafting and **extracting time-series features** at gigabyte scales. These features were resilient to challenging settings like network congestion and varying bandwidth capacities.

Pytorch | Python | Numpy | Matplotlib

Aug 2019 – May 2019 | **IBM Research Labs | Natural Language Research Engineer, Bangalore, India**

- Generated **adversarial text** using black-box techniques to break NLP models trained for classification.
- Integrated deep learning models with a genetic algorithm for a hybrid AI system.
- Implemented LSTM classifiers, sequence to sequence translation models and language models.

Pytorch | Python | NLTK | Numpy | Matplotlib | FairSeq | TorchText | InferSent | GloVe

Aug 2018 – May 2018 | **Indian Space Research Organization | Computer Vision Research Engineer, Dehradun, India**

- Developed an **Image Segmentation API** for high-resolution satellite images.
- Implemented **U-Nets** for the segmentation of tarred roads and deep water bodies.
- Preprocessed 500 GB of multi-spectral geospatial data for its use in the computer vision model.

Tensorflow | Python | Matplotlib | Shapely | Gdal | Osr | Pandas | Descartes

</> Technical Expertise

Programming	Python, Java
Frameworks	Pytorch, Tensorflow, Flask-restx
Databases	Microsoft SQL Server, Neo4j Graph DB
Operating Systems	Mac OS X, Windows 7/8/10, Linux Redhat, Linux Ubuntu

Leadership Roles

2018-19 **Machine Learning Special Interest Group Coordinator, ACM** BITS Pilani Chapter (India's best ACM chapter 4 years in a row)