

Alex MATHAI

Computer Science Researcher

alexmathai.online  Google Scholar [in linkedin.com/in/alex-mathai-403117131](https://www.linkedin.com/in/alex-mathai-403117131)
[@ alexmathai98@gmail.com](mailto: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** | **PhD Student, Columbia University | LLM Agents for million line+ codebase debugging, NYC, USA**
Sept 2023
- > 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** | **Research Engineer, IBM Research | Modernizing Monoliths to Microservices , BANGALORE, India**
Aug 2020
- > 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** | **University of New South Wales (UNSW) | Time Series Analysis Research, SYDNEY, Australia**
Aug 2019
- > 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** | **IBM Research Labs | Natural Language Research Engineer, BANGALORE, India**
May 2019
- > 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** | **Indian Space Research Organization | Computer Vision Research Engineer, DEHRADUN, India**
May 2018
- > 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)