# Arun George Zachariah

Email: azachariah@nvidia.com | Website: <u>https://arun-george-zachariah.github.io</u>

#### **EDUCATION**

University of Missouri-Columbia	
Ph.D., Computer Science	Spring 2020 – Summer 2022
Dissertation Title: A System for Large-Scale Image and	
Video Retrieval on Everyday Scenes	
Advisor: Dr. Praveen Rao	
University of Missouri-Kansas City	
Interdisciplinary Ph.D., Computer Science (Transferred)	Fall 2018 – Fall 2019
Advisor: Dr. Praveen Rao	
<b>Cochin University of Science and Technology</b>	

Bachelor of Technology, Electrical and Electronics Engineering

2008 - 2012

#### PUBLICATIONS

Nikhil Mehta, Jonathan Lorraine, Steve Masson, Ramanathan Arunachalam, Zaid Pervaiz Bhat, James Lucas, **Arun George Zachariah** – "Improving Hyperparameter Optimization with Checkpointed Model Weights". *arXiv preprint arXiv:2407.01526*, 2024.

Shijia Liao, Shiyi Lan, **Arun George Zachariah** – "EVA-GAN: Enhanced Various Audio Generation via Scalable Generative Adversarial Networks". *arXiv preprint arXiv:2402.00892*, 2024.

Arun Zachariah, Praveen Rao – "Video Retrieval for Everyday Scenes with Common Objects". In the *Annual ACM International Conference on Multimedia Retrieval* (ICMR 2023), Greece, 2023.

Praveen Rao, **Arun Zachariah** – "Enabling Large-Scale Human Genome Sequence Analysis on CloudLab". In IEEE INFOCOM 2022 - IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS), 2022.

**Arun Zachariah**, Praveen Rao, Brian Corn, and Dominique Davison – "Zero Shot Learning for Predicting Energy Usage of Buildings in Sustainable Design". In AAAI Workshop on AI to Accelerate Science and Engineering (AI2ASE), Canada, 2022.

Arun Zachariah, Maha Alrasheed – "Private-Share: A Secure and Privacy-Preserving De-Centralized Framework for Large Scale Data Sharing". In Proc. of the *3rd ACM International Conference on Multimedia in Asia* (ACM MM Asia 2021), Australia, 2021.

Praveen Rao, Arun Zachariah, Deepthi Rao, Peter Tonellato, Wesley Warren and Eduardo Simoes – "Accelerating Variant Calling on Human Genomes Using a Commodity Cluster". In the *30th ACM International Conference on Information and Knowledge Management* (CIKM), Australia, 2021 (Nominated for Best Short Paper Award).

Suveen Angraal, **Arun Zachariah**, Raaisa Raaisa, Rohan Khera, Praveen Rao, Harlan M Krumholz, and John A Spertus – "Evaluation of Internet-based Crowdsourced Fundraising to Cover Healthcare Costs in the United States". In *JAMA Network Open*, 4(1), 2021.

**Arun Zachariah**, Mohamed Gharibi, Praveen Rao – "A Large-Scale Image Retrieval System for Everyday Scenes". In Proc. of the *2nd ACM International Conference on Multimedia in Asia* (ACM MM Asia 2020), Singapore, 2021.

Nouf Alrasheed, Arun Zachariah, Shivika Prasanna, Deepthi Rao, and Praveen Rao – "Deepfakes for Histopathology Images: A Myth or Reality?". In 49th Annual IEEE Applied Imagery Pattern Recognition (AIPR) Workshop 2020: Trusted Computing, Privacy, and Securing Multimedia, Washington, D.C., 2020.

Arun Zachariah, Mohamed Gharibi, Praveen Rao – "QIK: A System for Large-Scale Image Retrieval on Everyday Scenes with Common Objects". In the *10th International Conference on Multimedia Retrieval* (ICMR 2020), Dublin, Ireland, 2020.

Mohamed Gharibi, **Arun Zachariah** and Praveen Rao – "FoodKG: A Tool to Enrich Knowledge Using Machine Learning Techniques". In *Frontiers in Big Data*, Volume 3, 2020.

Daniel E. Lopez Barron, Praveen Rao, Deepthi Rao, Ossama Tawfik, Arun Zachariah – "Large-Scale Storage of Whole Slide Images and Fast Retrieval of Tiles Using DRAM". In 2020 SPIE Defense + Commercial Sensing: Big Data II: Learning, Analytics, and Applications Conference (11395), Anaheim, CA, 2020.

**Arun Zachariah**, Praveen Rao, Anas Katib, Monica Senapati, Kobus Barnard – "A Gossip-Based System for Fast Approximate Score Computation in Multinomial Bayesian Networks". In the *35th IEEE International Conference on Data Engineering* (ICDE), Macau, China, 2019.

#### PATENTS

Parthasarathy Sriram, Farzin Aghdasi, Arun George Zachariah, Varun Praveen – "Automated Media Content Recognition for Understanding Multimedia", *18/673,972* (May 2024) (Under Review)

Steve Masson, Ramanathan Arunachalam, Nikhil Mehta, Farzin Aghdasi, Jonathan Lorraine, Zaid Pervaiz Bhat, **Arun Geroge Zachariah** – "Neural Network Performance Based Search", *18/660,182* (May 2024) (Under Review)

#### DATASETS

Praveen Rao, **Arun Zachariah**, Deepthi Rao, Peter Tonellato, Wesley Warren, Eduardo Simoes, February 27, 2021, "Variant Analysis of Human Genome Sequences for COVID-19 Research", IEEE Dataport, doi: https://dx.doi.org/10.21227/b0ph-s175.

#### **PROFESSIONAL EXPERIENCE**

#### **Software Engineering Manager**

**NVIDIA** Corporation

#### Responsibilities

- Design and build the Metropolis AI Model production pipeline.
- Automate data and asset handling process as a part of the Metropolis AI Model production pipeline.
- Design an AI-assisted data labeling framework by leveraging modern data-centric AI principles,
- Develop automated workflows to generate data sets for AI models.

06/2022 - Present

- Accelerate model adoption by designing a framework to enable users to try TAO models by abstracting away the framework complexities.
- Lead end-to-end planning and execution of TAO releases, ensuring optimal software deployment and release management.
- Exemplify a steadfast commitment to trustworthy AI principles in all aspects of release planning and execution.

# Graduate Research Assistant

Gruduite Research rissistant	
University of Missouri-Columbia	02/2020 - 05/2022
University of Missouri-Kansas City	09/2018 - 01/2020

#### Responsibilities

- Build a comprehensive evaluation framework for benchmarking Blind Learning against other state-of-the-art privacy-preserving deep learning techniques as a part of the industry-sponsored research project with TripleBlind.
- Develop a system to enable democratizing genome sequence analysis on CloudLab as a part of NSF-RAPID (CNS-2034247) and MU Center for Biomedical Informatics.
- Develop a deep learning model to predict energy usage metrics of buildings by considering a large number of building parameters, as a part of the NSF Center for Big Learning (CNS-1747751).
- Develop and test a large-scale image retrieval system as a part of the NSF Center for Big Learning(CNS-1747751).
- Develop a scalable system for storage and fast retrieval of Whole Slide Images, for enabling next generation image analytics, as a part of the NSF I-Crops program.
- High-performance networking with RDMA and DPDK to enable lightning-fast storage and retrieval.
- Collect, process, and analyze data from various sources using complex techniques and procedures as required, to help prepare grant proposals or funding applications.
- Conduct customer discovery interviews, as a part of the NSF I-Crops program, to understand industry requirements and challenges.
- Supervise students working on research projects.
- Perform other duties as assigned by the supervisor.

# Intern, Engineering

Arm Ltd.

#### Responsibilities

- Analyze code to identify bottlenecks across software, micro-architecture, and architecture.
- Perform extensive experiments on real and simulated platforms.
- Develop innovative tools for architectural exploration and performance analysis.
- Autonomously queue and execute benchmarks on a pool of machines from MAAS and AWS with Ansible and Juju.

## **Technology Analyst**

Infosys (Consultant at Apple Inc.)

#### Achievements

- Implemented innovative systems for data collection, storage and management of customer data.
- Built a custom product for Customer person data standardization, matching, de-duplication, linking and feeding to enterprise data warehouse systems.

06/2021 - 08/2021

05/2013 - 08/2018

- Successfully set up Hadoop over multiple nodes and migrated data from Oracle 10g to HBase using Sqoop.
- Designed and implemented a locking and buffering mechanism using an in-memory database, resulting in a 10 times increase in processing bandwidth with the same hardware configuration.
- Transformed regular Normalized RDBMS apps to De-Normalized structure for scaling and improved response times from 200-300ms to 20-30ms.

#### Responsibilities

- Translate business goals and customer needs into prioritized product requirements and use cases.
- Application tuning JVM Performance tuning, Database tuning, Profiling, Capacity planning and health check monitoring during peak business days.
- Develop multiple MapReduce jobs for data cleaning and pre-processing.
- Writing Hive queries to read from HBase.
- Identify new technologies and tools for enhancing product value and increasing team productivity.
- Train, mentor and manage a team of software engineers.

## TECHNICAL SKILLS

- Programming Languages: C, C++, Java, Python, Scala, Gremlin
- J2EE Technologies: Servlets, JSP, Web Services
- Web Technologies: HTML5, Javascript, jQuery, Angular JS
- Database Technologies: SQL, SPARQL, Mongo DB, PostgreSQL, Greenplum, MySQL, TinkerPop, Neo4j
- Software Libraries & Tools: PyTorch, TensorFlow, Keras, Caffe, OpenCV, scikit-learn, SciPy, pandas, NumPy, Hadoop, Spark, Hive, Kafka, HBase, genism, Deepwalk
- Version Control Tools: GIT, SVN
- Build Tools: Ant, Maven, Gradle, sbt

## ACADEMIC SERVICES

#### **Program Committee Member**

• 14 <sup>th</sup> International Conference on Multimedia Retrieval	2024
<ul> <li>5<sup>th</sup> International Workshop on Health Data Management in the Era of AI. (held in conjunction with EDBT/ICDT 2023)</li> </ul>	2022
External Reviewer	
• 40 <sup>th</sup> Internation Conference on Data Engineering	2024
Journal of Engineering Research and Sciences	2024
• 4 <sup>th</sup> IEEE International Conference on Electrical, Computer, and Energy Technologies	2024
• IEEE International Conference on Artificial Intelligence, Computer,	2024
Data Sciences and Applications	
• 3 <sup>rd</sup> IEEE International Conference on Electrical, Computer, and Energy Technologies	2023
• 3 <sup>rd</sup> International Conference on Electrical, Computer Communications and	2023
Mechatronics Engineering.	
<ul> <li>23<sup>rd</sup> IEEE International Conference on Mobile Data Management</li> </ul>	2022
<ul> <li>NeurIPS Workshop on Human and Machine Decisions</li> </ul>	2021
• 23 <sup>rd</sup> International Conference on Big Data Analytics and Knowledge Discovery	2021
• 30 <sup>th</sup> International Conference on Computer Communications and Networks	2021
• 4 <sup>th</sup> International Conference on Connected and Autonomous Driving (MetroCAD)	2021

• Cluster Computing Journal – Springer	2020	
• 8 <sup>th</sup> International Conference on Big data and Cloud Computing		
• 8 <sup>th</sup> IEEE International Conference on Healthcare Informatics	2020	
• 25 <sup>th</sup> International Conference on Pattern Recognition	2020	
• 22 <sup>nd</sup> International Conference on Big Data Analytics and Knowledge Disco	very 2020	
• 21 <sup>st</sup> IEEE International Conference on Mobile Data Management	2020	
• 32 <sup>nd</sup> Conference on Graphics, Patterns and Images SIBGRAPI	2019	
• 7 <sup>th</sup> International Conference on Big data and Cloud Computing	2019	
• 21 <sup>st</sup> International Conference on Big Data Analytics and Knowledge Discov	very 2019	
Mentoship		
Global Mentorship Initiative	2022, 2023, 2024	
KaggleX BIPOC Mentorship Program	2023	
Conference Workshops Conducted		
Upsilon Pi Epsilon - After School Program	2021	
ACM/Code.org Hour of Code at Centro Latino	2021	
ICBCC - Fundamentals of Deep Learning for Computer Vision	2019	
Student Volunteer		
• 35 <sup>th</sup> Conference on Neural Information Processing Systems (NeurIPS)	2021	
• 34 <sup>th</sup> Conference on Neural Information Processing Systems (NeurIPS)	2020	
• 7 <sup>th</sup> International Conference on Big data and Cloud Computing	2019	
Deep Learning Fundamentals for Computer Vision.	2019	
AWARDS AND GRANTS		
• 3rd place at the 38th Annual Research & Creative Activities Forum	2022	
<ul> <li>35<sup>th</sup> NeurIPS Student Volunteer Registration</li> </ul>	2021	
Upsilon Pi Epsilon Scholarship Award	2021	
ACM SIGMM Travel Grant	2021	
VLDB SPEND grant	2021	
MU Graduate Professional Council Research Development Award.	2021	
• 34 <sup>th</sup> NeurIPS Student Volunteer Registration	2020	
UMKC School of Graduate Studies Travel Grant	2019	
Student Activity Fee Committee Travel Grant	2019	
CSEE Balaji Krithikaivasan Memorial Travel Grant	2019	
IEEE Standards Education Grant	2012	