Emil Biju

emilbiju7@gmail.com | emilbiju.github.io | +919840918221 | emilbiju | emilbiju

Education

Indian Institute of Technology Madras (IIT Madras) *2017–2021*
B.Tech (Honors) in Electrical Engineering (**CGPA: 9.70/10**), Minor in Deep Learning Chennai, India

- **Ranked 2nd** out of 53 students in the department.
- Received the top grade (S) in all courses from the Computer Science, Mathematics and Humanities categories.
- B.Tech Thesis: Sample-specific Attention Masks for Model Transparency and Adversarial Detection

Publications

1. **Input-specific Attention Subnetworks for Adversarial Detection** [Paper]
   Emil Biju, A. Sriram, P. Kumar, M. Khapra; Findings of ACL 2022, Dublin, Ireland
2. **Joint Transformer/RNN Architecture for Gesture Typing in Indic Languages** [Paper]
   Emil Biju, A. Sriram, M. Khapra, P. Kumar; COLING 2020, Barcelona, Spain
3. **Perturbation Analysis of Practical Algorithms for the Maximum Scatter TSP** [Paper]
   Emil Biju, S. Raman; ALENEX workshop @ SODA 2022, Virginia, U.S.A.
4. **Vocabulary-constrained Question Generation with Rare Word Masking & Dual Attention** [Paper]
   Emil Biju; 🏆 Best Paper Honorable Mention; ACM CODS-COMAD 2021, Hyderabad, India

Professional Experience

Microsoft R&D *June 2021–Present*
Data & Applied Scientist (Full-time) Bangalore, India

- Working as a researcher at the intersection of data science and cybersecurity, with a focus on OAuth cloud app security for the **last 1.5 years**.
- Developed industry-first ML solutions using NLP, knowledge graphs, anomaly detection and computer vision to model cyber-attack patterns and avert security threats, meeting stringent goals on latency and efficacy.
- Deployed ML models using PySpark that currently analyze terabytes of data and over 10M apps every day.
- Published a **paper** at MLADS 2022, received an early **promotion** and currently filing a **U.S. patent**.

Microsoft R&D *May–July 2020*
Data & Applied Scientist Intern Hyderabad, India

- Developed CNN and Transformer-based deep learning models to analyze multi-spectral satellite images for estimating biomass in agricultural fields and identifying prospective areas for oil exploration.
- Designed a new data structure for `xarray`, an open-source Python package, to support tree-based hierarchical storage of data.

GE Healthcare R&D *May–July 2019*
Data Scientist Intern Bangalore, India

- Used graph-based keyword clustering and topic ranking to analyze text in service records of healthcare machines.
- Set up an automated pipeline to flag common failure patterns and suggest quality improvement opportunities.
- Reduced the time taken to extract insights from service records by 11x and was appreciated by multiple company leaders.

Scholastic Achievements

- **IIT Madras Silver Medal:** Awarded the Dr. Dilip Veeraraghavan Memorial Award by IIT Madras for featuring as the institute's second topper based on overall CGPA and cumulative performance in H category courses.
- **NTSE Scholarship:** Awarded by the Govt. of India based on a nationwide exam with a 0.1% acceptance rate.
- **KVPY Fellowship:** Awarded by IISc based on a nationwide science exam with a 2.5% acceptance rate.
- Awarded the **Branch upgrade** option for exceptional academic performance in the 1st semester at IIT Madras.
- **Samsung-IITM Pravartak Fellowship:** Awarded for research work on interpretability of Transformer models.
- **Best Paper Honorable Mention:** Awarded for my publication at ACM CODS-COMAD 2021 in the YRS track based on my self-guided research work in data-efficient question generation.

Research Projects

Adversarial Detection in Transformers using Attention Subnetworks *Jan–Nov 2021*

B.Tech thesis | Guides: Prof. Mitesh Khapra & Prof. Pratyush Kumar, CS Dept., IITM *Paper* | *Webpage*

- Studied the self-attention framework of Transformer models to improve their interpretability and robustness.
- Demonstrated that Transformers contain input-specific attention subnetworks that are interpretable and can be used to detect adversarial inputs.
- Improved the state-of-the-art accuracy in adversarial detection by 7.5% across 10 NLP tasks and 11 attack types.

IndicSwipe: Decoding Swipe Gesture Inputs to Indic Language Keyboards *Jan–July 2020*

Undergrad research | Guides: Prof. Mitesh Khapra & Prof. Pratyush Kumar, CS Dept., IITM *Paper* | *Webpage*

- Curated a training dataset of swipe gestures for 300k words in 7 Indic languages by using the brain motor control principle of jerk minimization to simulate swipe inputs to a smartphone keyboard.
- Developed a Transformer-LSTM model for accurate swipe decoding and an ELMo-inspired character embedding model for fast/parallelized spelling correction. Achieved state-of-the-art accuracies of 70-95% across 7 languages.

Approximation Algorithms for the Maximum Scatter TSP *Jan–Aug 2021*

Undergrad research | Guide: Prof. Raghavendra Rao B.V., CS Dept., IITM *Paper* | *Webpage*

- Devised 6 discrete approximation algorithms for the NP-hard maximum scatter traveling salesman problem.
- Performed smoothed analysis with various perturbations and edge-cost metrics to benchmark the stability, speed and accuracy of these algorithms. Demonstrated their practical utility using real-world datasets.

Risk propagation in Knowledge Graphs for detecting Attack Campaigns *Sept–Nov 2022*

Cybersecurity research team, Microsoft | Manager: Sudarson Mothilal ***Patent in filing***

- Built a knowledge graph representing OAuth cloud apps and their metadata, developed an algorithm to propagate reputation scores between graph nodes and used k-connectedness to discover attack campaign clusters.
- Discovered & disabled 21 apps from a consent phishing campaign that had been undetected for over 2 years.

Optimized RISC-V CPU implementation *Jan–Aug 2021*

Computer Organization course project, IIT Madras *GitHub*

- Developed a RISC-V CPU with a 5-stage pipeline in Verilog and optimized performance using branch prediction and exception handling. Performed exhaustive verification of CPU functionality on an FPGA board.

Relevant Coursework

Computer Science: Introduction to Programming, Data Structures and Algorithms, Topics in Design and Analysis of Algorithms, Introduction to Automata, Languages and Computation

Data Science/ML: Introduction to Machine Learning, Deep Learning, Natural Language Processing, Advanced Topics in Signal Processing (focused on Computer Vision), Data Mining

Mathematics: Linear Algebra, Probability Foundations, Graph Theory, Series & Matrices, Differential Geometry

Electrical: Computer Organization, Microprocessors, Information Theory, Digital Signal Processing, Digital Systems

Online Courses: Machine Learning (Coursera), Deep Learning: Advanced NLP and RNNs (Udemy), CNNs for Visual Recognition (CS231n - Stanford Online)

Technical Skills

Languages: Python, C, C++, PySpark, SQL, ARM, Verilog
Libraries: TensorFlow, Keras, OpenCV, Numpy, NLTK, Matplotlib, Scikit-learn and other machine learning libraries
Interests: Machine learning, Deep learning, Natural language processing, Optimization algorithms

Leadership roles & Extra-curricular activities

Coordinator, Extra Mural Lectures: Invited top speakers and organized the flagship guest lecture series of IITM.

Learning champ, Microsoft: Organized learning sessions and set up a knowledge-sharing website to create continuous learning opportunities for 1000+ employees across the cybersecurity team.

School Head Boy: Popularly elected by the school community and headed the students' council in high school.

Founder, Passion JEE: Created a blog to mentor engineering aspirants in India & clocked over 6k views to date.

Voluntary Service: Served as a paper reviewer for the MLADS 2021 conference and a mentor for college freshers.

Public Speaker: Featured as the lead emcee/speaker at several prominent events, including welcome speeches to the Vice President of India, top bureaucrats and executives.