Emil Biju

emilbiju@stanford.edu | +1 650 507 8666 | emilbiju.github.io | emilbiju | emilbiju

Education

Stanford University Stanford, CA, USA | *2023–2025*
M.S. in Electrical Engineering, Focus Area: Machine Learning/AI
- Coursework/research in machine learning, deep learning, NLP, and signal processing

IIT Madras (Indian Institute of Technology Madras) Chennai, India | *2017–2021*
B.Tech (Honors) in Electrical Engineering (**GPA: 9.70/10**), Minor in Deep Learning
- Department **2nd topper**, highest grade in all courses from the CS and Math departments

Work Experience

Microsoft | Data & Applied Scientist (Full-time, 2 years) Bangalore, India | *2021–2023*
- Pursued machine learning research with a focus on cybersecurity for cloud apps.
- Used NLP, computer vision, graphs, and anomaly detection to model attack patterns and track app behavior leading to the detection of real-world cyberattacks.
- Deployed and owned ML models processing terabytes of data every hour using PySpark, meeting stringent goals on latency and efficacy. **Promoted to Data Scientist 2** role for exceptional work.

Microsoft | Data & Applied Scientist Intern Hyderabad, India | *May–July 2020*
- Developed CNN and Transformer-based deep learning models to analyze multi-spectral satellite images for forecasting biomass in agricultural fields and identifying prospective areas for oil exploration.
- Designed a hierarchical data storage solution for `xarray`, an open-source python library.

GE Healthcare | Data Scientist Intern Bangalore, India | *May–July 2019*
- Used graph-based keyword clustering and topic ranking to analyze service logs of healthcare machines and identify quality improvement opportunities.
- Delivered an end-to-end automated pipeline that reduced time for extracting insights by 11x.

Publications

1. **Input-specific Attention Subnetworks for Adversarial Detection** | ACL 2022 Findings [Paper]
2. **Joint Transformer/RNN Architecture for Gesture Typing in Indic Languages** | COLING 2020 [Paper]
3. **Perturbation Analysis of Practical Algorithms for the Maximum Scatter TSP** | ALENEX 2022 [Paper]
4. **Vocabulary-constrained Question Generation with Rare Word Masking** | CODS-COMAD 2021 [Paper]

** First author of all four publications listed above*

Key Projects

Decoding Swipe Inputs to Smartphone Keyboards | UG Research [Paper][GitHub]
- Built a Transformer-LSTM deep learning model to decode words from swipe inputs to a touch keyboard.
- Curated training datasets in 7 low-resource languages using a mathematical model to simulate human swipe inputs based on the brain's motor control principle.
- Achieved state-of-the-art accuracies of 70-95% across 7 Indian languages.

Adversarial Detection in Deep Transformer Models | UG Thesis [Paper][GitHub]
- Studied the self-attention framework in Transformers to improve their interpretability and robustness.
- Designed an efficient technique to extract critical subnetworks from a Transformer that are interpretable and useful in detecting adversarial inputs.
- Improved the state-of-the-art accuracy in adversarial detection by 7.5% across 10 NLP tasks.

Knowledge Graph for Cyberattack Detection | Microsoft **[Patent pending]**

- Developed a graph representation of 10 million cloud apps and their metadata to discover hidden relationships among malicious apps and improve cyberattack detection.
- Designed a risk propagation algorithm that assigns risk scores to apps and flags malicious clusters.
- This work **uncovered 4 cyberattacks** involving 2k+ apps that have now been disabled by Microsoft.

Deep CNN for detecting Logo Impersonation | Microsoft

- Developed a 3-stage deep CNN to detect impersonated versions of legitimate brand logos using object localization, colorspace/edge analysis, and embedding-based proximity detection.
- Deployed the model efficiently using PySpark to evaluate 30k logos within 10 minutes with 80% recall.

App Governance Copilot using LLMs | Microsoft

- Developed a GPT-3-based chat assistant specialized in cybersecurity to assist security researchers in investigation, hunting, and remediation of threat alerts.
- Used chain-of-thought prompting and external APIs to automate data search and threat mitigation.

Siamese CNN for detecting Signature Forgery | UG Research

- Used deep CNNs and computer vision techniques to generate image embeddings that encode stroke and curve patterns in signatures. Built a Siamese network to fine-tune embeddings and detect forgeries.

Technical Skills

Programming Languages: Python, C, C++, Apache Spark, SQL, HTML, ARM

Libraries: TensorFlow, Keras, PyTorch, OpenCV, NLTK, Numpy, Matplotlib, Sklearn, other ML libraries

Technologies: ML, NLP, Deep Learning, Computer Vision, LLMs, Discrete algorithms

Relevant Coursework

ML: Machine Learning, Deep Learning, Natural Language Processing (NLP), Image Processing, Data Mining

CS: Data Structures & Algorithms, Design & Analysis of Algorithms, Computer Organization, Introduction to Programming, Applied Programming

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

Electrical: Signal Processing, Microprocessors, Internet of Things, Information Theory, Control Engineering

Awards

IIT Madras Silver Medal: Institute 2nd topper based on overall GPA and performance in H.S. courses.

Best Paper Honorable Mention: Awarded for my publication at ACM CODS-COMAD 2021.

JN Tata Scholar: Awarded to the top 100 students from India for pursuing graduate study abroad.

NTSE Scholar: Awarded to the top 750 students by the Govt. of India based on a nationwide exam.

KVPY Fellow: Awarded to the top 1500 students by IISc based on a nationwide exam.

Extra-curricular activities

Blogging: Created a blog to mentor engineering aspirants in India & clocked over 6k views to date.

Learning champ, Microsoft: Curated learning material and organized sessions for 1000+ employees.

School Head Boy: Popularly elected by the school community in high school.

Public Speaking: Featured as the lead emcee/speaker at several prominent events in school and at work.

Academic service: Served as a paper reviewer for the MLADS 2021 conference.