# Sukrit Kumar

+1- (943) 600-2455 | ksukrit2001@gmail.com | Linkedin | Website | GitHub

## EDUCATION

**Georgia Institute of Technology**                                              Atlanta, GA
*M.S. Computer Science with a specialization in **ML, Thesis-track***          *Aug 2024 – Present*

**Birla Institute of Technology and Science (BITS) Pilani**                  Hyderabad,India
*B.E. in Electronics & Communications, Minor in Data Science*                 *Aug 2019 – July 2023*
***CGPA: 9.56/10***

## EXPERIENCE

**Microsoft Research**                                                    July 2023 - July 2024
*Research Fellow*                                                          *Bangalore, India*
- Worked on a framework for root-causing failures in a large-scale container platform at Microsoft
- Created a multi-dimensional action risk metric for LLM workflow automation agents
- Led research to better understand how developers spend their time at Microsoft to improve developer productivity
- Advisors: **Ankur Mallick, Chetan Bansal, Suman Nath**

**ShareChat AI**                                                          Jan 2023 - June 2023
*Data Science Intern*                                                     *Bangalore, India*
- Deployed ranker models leveraging historical user and session data driving retention gains
- Helped build an ML model serving system supporting large-scale MMoE models (>10GB) at <400 ms latency
- Created model monitoring setup and ML Observability to address offline-online ML model performance discrepancies
- Identified and resolved loss of over 70% historical user events
- Worked with **Rishabh Mehrotra** on modeling the user-journey throughout the app

**Arista Networks**                                                       May 2022 - July 2022
*SWE Intern*                                                              *Pune, India*
- Designed and implemented the CloudVision Access Point (AP) Health Dashboard from scratch
- Re-wrote SQL queries leading to upto 3x speed-up across different queries
- Developed novel method to calculate percentile values for distributed time-series data
- First ever intern in India to be awarded the Peer Bonus

**CSIR-CEERI**                                                            May 2021 - July 2021
*Summer Research Intern, Deep Learning*                                   *Pilani, Rajasthan*
- Developed a Dual-attention based implementation of Capsule Networks leading to 1-2% accuracy improvement across datasets
- Achieved 99.47% validation accuracy on MNIST and 92.94% accuracy on FashionMNIST dataset

## ACHIEVEMENTS

**GeorgiaTech SCS GTA Fellowship**                                         Aug 24 - Present
*Awarded to **top 0.5%** students to pursue high-quality research at GeorgiaTech*

**BITS Pilani Merit Scholarship**                                         Jan 20 - July 23
*Awarded institute merit scholarship for **7** consecutive semesters (**Top 1% out of 1100+ Students**)*

## PUBLICATIONS

**Prompt-Propose-Verify**: A Reliable Hand-Object-Interaction Data Generation Framework using Foundational Models
**Sukrit Kumar***, Gurusha Juneja*
International Workshop on AI for Digital Human in AAAI Conference on Artificial Intelligence (AAAI, 2024)

## PROJECTS

**In-Memory Vector Datastore** | *Go Lang*                                April 2024
- Light-weight in-memory vector datastore supporting operations like: Add, Subtract, Multiply, Scale, Division, Resizing
- Supports nearest-neighbor search for finding closest vectors based on two different distance metrics (cosine/euclidean)

**Face-Mask Detection using Deep Learning** | *Python, Keras, Tensorflow, OpenCV*    Dec 2022
- Real-time, light-weight (<6.59 MB) face mask detection on images and video for Edge-IoT devices
- Achieves approximately 98.55% accuracy on single face detection and 94.61 % accuracy in real-world multi-face scenario

**Toxic Comment Classifier** | *Python, Keras, Tensorflow*                March 2021
- LSTM-based RNN model to classify social-media comments on 6 different classes for toxicity
- Deployed a pruned light-weight model (<2.6MB) on the web using tf.js and Flask for real-time inference in the browser

## TECHNICAL SKILLS AND INTERESTS

**Programming Languages**: C, C++, Python, Java, Go
**Libraries**: OpenAI, pandas, NumPy, Matplotlib, Keras, TensorFlow, PyTorch, HuggingFace, LangChain, DSPy
**Tools and Databases**: Git, PostgreSQL, Postman, Jupyter, Docker, K8s, GCP, Pinecone, Weights & Biases (W&B)
**Interests**: Large-scale distributed systems, Systems for ML, Recommendation Systems, Machine Learning