Tanmay Chavan

+1 404-250-2684 • tchavan3@gatech.edu • linkedin.com/in/tanmay-chavan01/ • GitHub (Tanmay-Chavan)• chavantanmay.github.io

EDUCATION

Georgia Institute of Technology August 2024 - Present
MS in Computer Science Atlanta, GA
- **Graduate Teaching Assistant** for CS 6601: Intro to Artificial Intelligence
- **CGPA: 4.0/4.0**

Pune Institute of Computer Technology August 2019 - May 2023
Bachelor of Engineering in Computer Engineering Pune, India
- **CGPA: 9.16/10.0**

TECH SKILLS

Programming Languages:	Python, Java, C/C++, Dart, shell
Frameworks:	Tensorflow, PyTorch, Hadoop, Spark, SQL, NLTK, Numpy, Pandas, Scikit, Jenkins, Kafka, Spring Boot, Redis, Qt, Flutter, OpenGL
Interests:	Machine Learning, AI, Web development, App development, Distributed Systems, NLP, DevOps, Computer Networking

WORK EXPERIENCE

***Software Engineer* | Avaya** *Pune, India* | July 2023 – June 2024
Web Development, Cloud Computing, DevOps
- Developed several **backend** components of on-premise and **cloud services** services using **Spring Boot, Azure cloud, Kafka,** and **Redis**. Exposed web services using **REST APIs** and **WebSocket**.
- Reduced **Jenkins** build pipeline runtime from **3 hours** to **40 minutes** by implementing parallelized builds, optimizing scripts and caching partial builds.
- Built transformer-based models to gauge the fatigue of call center agents. Integrated the model with the call routing algorithms to reduce the load of employees with higher fatigue.
- Implemented advanced telemetry functions for Avaya AES using Spring Boot **Actuators** to enable a pay-as-you-go billing system.

***Project Intern* | L3Cube** *Pune, India* | June 2022 – July 2023
Machine Learning, Natural Language Processing
- Investigated the effects of **domain-specific pre-training** in language models.
- Created transformer-based models pre-trained on **40 million tweets** in Marathi and Hindi each, with **SOTA** performance in hate speech detection.
- Built L3Cube-MeCorpus, a collection of large Marathi-English **code-mixed** corpus, labeled datasets for downstream tasks, and SOTA language models.
- **Mentored** a group of students analyzing hate speech patterns in Marathi posts on social media platforms.

***Research Assistant* | Pune Institute of Computer Technology** *Pune, India* | June 2021 – June 2023
Machine Learning, Distributed Systems
- Analyzed the robustness of sentence encoders against **adversarial perturbations** in the Computational Linguistics Lab.
- Worked on building **scalable** recommendation systems in distributed settings on big data at the CAILMD Lab.
- Used Apache **Mahout** and **Hadoop** to build several clustering, classification, and regression machine learning models.

***Student Developer Intern* | Google Summer of Code** *Remote* | May 2021 – August 2021
Computer Graphics
- Contributed to Krita. Implemented a new algorithm for **clipping** vector shapes which accounts for **Bezier curves** in addition to line segments.
- Improved vector shape operation algorithms such as union and intersection of shapes while incorporating the improved Bezier-clipping algorithm.
- Worked in a **Linux** environment on a **C++** codebase with the **Qt** framework. Used building tools such as **CMake** and automated testing tools.
- **Project details:** https://summerofcode.withgoogle.com/projects/#4650352734896128

PROJECTS

PayTTMM - A Bill Splitting App (*Skills: Flutter, Firebase, Dart, App development*)
- Created an app to help split bills amongst a group of people using **Flutter** and **Firebase** (link: https://github.com/Tanmay-Chavan/payTTMM)

Scalable Twitter hate speech detection bot (*Skills: Spark, Hadoop, Mahout, Python, Distributed Computing, ML, NLP*)
- Developed a **scalable** bot to detect hate speech in real time on Twitter using Apache **Hadoop, Mahout** and **Spark**, and the Twitter API.

Snake Xenzia (*Skills: C/C++, OpenGL, Computer Graphics*)
- Created a computer game inspired by the classic Snake Xenzia game found on Nokia 3310 using **OpenGL** and **C++**.

PUBLICATIONS

1. My Boli: Code-mixed Marathi-English Corpora, Pretrained Language Models and Evaluation Benchmarks
   Tanmay Chavan, Omkar Gokhale, Aditya Kane, Shantanu Patankar, Raviraj Joshi, ***IJCNLP-AACL 2023***
2. Spread Love Not Hate: Undermining the Importance of Hateful Pre-training for Hate Speech Detection
   Omkar Gokhale, Aditya Kane, Shantanu Patankar, Tanmay Chavan, Raviraj Joshi, ***ICBINB Workshop at NeurIPS 2022***
3. A Twitter BERT Approach for Offensive Language Detection in Marathi
   Tanmay Chavan, Shantanu Patankar, Aditya Kane, Omkar Gokhale, Raviraj Joshi, ***HASOC Workshop at FIRE 2022***
4. Large Language Models for Multi-label Propaganda Detection
   Tanmay Chavan, Aditya Kane, ***Arabic Natural Language Processing Workshop at EMNLP 2022***