Pranoy Kovuri

kovuripranoy@gmail.com
(979) 324-3647
linkedin.com/in/pranoy-k/
Santa Clara, CA

A technical/product leader with over a decade of experience in leading and developing cutting-edge AI and ML solutions for various domains, such as content safety and health. My passion is to create innovative products that leverage the power of data and algorithms to solve real-world problems and improve people's lives.

Experience (as of August 2024)

Google — Sunnyvale, CA

2024

Tech Lead Software Engineer | Google Cloud Safety Filters (Vertex AI)

- Currently leading Vertex AI Safety, building AI product solutions that serve all of Google's products, including Google Cloud.
- Spearheading development of Vertex AI Safety, creating cutting-edge AI product solutions
- Lead cross-functional teams to implement safety filters serving all Google products
- Drive innovation in AI safety and ethics for Google Cloud platform
- Architect scalable solutions to ensure responsible AI deployment across diverse applications

2023

Lead Software Engineer | AI Content Safety Team

- **Building Generative AI agents for high quality robust translation workflows**
- Launched multilingual models for configurable safety filters on the Vertex AI platform, saving 15 million per month in inference translate API calls:
  - Proposed and implemented Gemini-based translation quality evaluation pipeline.
  - Translated 1M training datasets to 140 different languages using Gemini.
- Developed a streamlined methodology for creating multilingual datasets:
  - Crafted scalable pipeline for multilingual-translation.
  - Handled multiple Cloud escalations.
  - Worked on improving multilingual model performance for content safety by experimenting with different translation and training methods.
  - Acted as the point of contact (POC) for all i18n-related aspects for the team.
- Launched of Terrorist Content Safety model to stop Israel-Hamas War propaganda:
  - Proposed, trained, and productionized in an extremely short timeline of one month, improving precision from 1% to 82%.
  - Trained distilled AutoMUM models, achieving the second unambiguous landing of Model Factory models.
- Designed, implemented, and launched the Responsible AI service
  - Made this a self-manageable, plug-and-play tool for all Generative AI safety initiatives within Google.
  - Implemented everything in a month's time, achieving 1M QPS (currently being used by 10k classifiers)
- Collaboratively Launched GenAI based Hate Speech detection model.
- Helped in creating a **model factory platform** for LLM-powered data curation and model building. This platform enabled us to curate and build models on a much faster & larger scale than ever before. Contributed to creating bulk inference capabilities on this platform.
- Crafted various prompts for Generative AI based safety classifiers and quickly evaluated them.

2022

Critical Code Red Engineer | Bard/Gemini Initiative

- Spearheaded three Googlers (vidhyaprakash@, queenay@ and sasimani@) for post-Bard generative AI-based content safety initiatives.
  - Developed and launched a data sandbox platform.
  - Created and launched an experimental content safety LLM platform, and helped with the effort to build the production platform.
  - Oversaw the setup of an LLM-based manual review queue.
- Initiated an LLM-based content safety initiative and engineered abuse grounding using LLM output overseen by manual reviewers. This increased manual reviewer efficiency by 6x. The initiative was only launched for 10% of traffic for online validation.

2021

- Fast-tracked the team for a rapid generative content safety initiative for Bard, also known as Code RED. This initiative included:
  - Curating 200M labeled entities across 10+ products.
  - Launching experimental LLMs.
  - Piloting an LLM-based content-safety initiative by experimentally proving LLM-based content safety.

Software Engineer | Text Content Safety Team

- Developed an online quality evaluation framework for measuring the utility of newer modeling and algorithm developments. This framework dramatically improved rollouts from O(months) to O(days).
  - Integrated an existing experimental framework for traffic duplication.
  - Enhanced another existing framework for measuring metrics on duplicate traffic.
  - Optimized the evaluation framework to measure metrics on differential entities resulting in a 10x reduction in manual review cost.
- Created a google wide extensible system which fights text obfuscation for malicious intent.
  - Developed a library that unifies, extends and extrapolates existing methods to fight obfuscation from homoglyphs.
  - The system is used by **1Giga QPS** across clients including YouTube, Gmail, Drive, Messaging, Chat, and Meet.
  - Presented Google-wide on how attackers circumvent using Unicode flaws, and acted as a go-to person in my organization for any such escalations.
  - Hosted an intern who found a way to find the level of obfuscation within a string based on structural similarity, OCR-based raster distance, and Hamming distance.

2020

Software Developer | Canonical Protections

- Launched Real time communication spam detection in Google to make Hangouts, text messages, voice, chat safer from spams, scam, phishing and harassment.
- Trained Unicorn, T5 models and advanced SOTA for abuse detection models and Onboarded abuse models to AutoML. Improved spam detection performance from 25P@90R to 98P@90R by onboarding to interpretability tools and training data cleaning.

2019

- Developed Tensorflow Ops for text data augmentation at character level. (substitution, insertion and deletion).

Texas A&M University — College Station, Texas

2019

Summer Research Intern | Reinforcement Learning Engineer

- Developing and testing Q-Learning, DDQN, DDPG and IRL based algorithms for a novel research problem.
- Designing a tailored gym environment for a zebra crossing scenario.

2017-2019

Research Assistant | Natural Language Understanding, Deep learning, Clinical Notes Information Extraction

- Developed unsupervised and semi-supervised neural architectures for joint Relation Extraction and NER.
- Modeled and tested sequential architectures for biomedical and sensor time series data.
- Improved performance of XGBoost models on ICU readmission prediction by incorporating text-based features from clinical notes.

2018

Teaching Assistant | Senior Capstone Design and Natural Language Processing

- Senior Capstone Design: Responsible for supervising the design and development phase for senior year undergraduate projects.
- Natural Language Processing: Responsible for clarifying questions, grading and overseeing projects.

Philips Research North America

2018

Summer Research Intern | HealthTech AI

- Extract and summarize valuable information from manually annotated Chest X-ray based radiology reports into a structured knowledge base for future ailment prediction.
- Employed deep learning techniques for entity and relation extraction from the radiology reports.

Qualcomm — Hyderabad, India

2015-2017 | Software Developer | Wi-Fi Embedded Systems Developer

- Implemented real time features in Wi-Fi subsystems and solved critical customer issues to develop requirements for OEMs.
- Provided onsite software support for OEM's product launches in Qualcomm China.

2014 | Software Intern | Senior Software Developer

- Enhanced and developed software solution for Samsung Group Play application.

Education

Texas A&M University — College Station, Texas

Class of 2019 | Master of Science | Computer Science — GPA: 4.00

National Institute of Technology — Warangal, India

Class of 2015 | Bachelor of Technology | Electronics & Communication Engineering — GPA: 3.80

Skills & Interests

- Recent Skills
  - Languages: Python, C++, Go (Novice)
  - Tools: Colab, VS Code, Perforce and Mercurial equivalent version control
- Old
  - Languages: Python, C++, C, Java, Assembly, Markdown, Ruby, JavaScript, Matlab, GO
  - Libraries: TensorFlow, Pytorch, Pandas, NumPy, SciKit, Keras, OpenCV, PostgreSQL
  - Tools: Google Toolkit, Jupyter, PyCharm, Perforce, Git, Code Collaborator, Code Blocks, Visual Studio
- Interests: Quantum Computation, Maths, Embedded systems, Robotics

Awards & Honors

- 10+ google recognitions (spot bonuses, peer bonuses and kudos), 5 Qual stars in Qualcomm.
- Udacity, Deep Learning Nanodegree Foundation (Jan. 2017 - July 2017).
- Patented a technology for finding adversarial attacks within text based inputs by measuring level of obfuscation, Approved within Google.

Personal & Academic Projects (Github)

April 2019

- Distributed Consensus Raft Algorithm
  - Implemented and open sourced platform for Raft Distributed Consensus algorithm.
  - Tested the system using various functional and unit tests.

May 2018

- Helping Heart Failure Patients survive – MIT LCP Project
  - Designing based search system for selecting Echocardiogram notes of patients for CHF with Sepsis.
  - Creating a pipeline for selecting Echocardiogram notes for Congestive heart failure patients and classifying the notes for various degrees of fluid resuscitation.

Dec. 2017

- The SmartChatBot
  - Developed a system which takes documents as inputs and is capable of answering questions about that document.
  - Worked on End-to-end Memory Networks and Dynamic Memory Network based model using LSTMs and GRUs.

Nov. 2017

- CSE-Dashboard: Social Networking Service Website

- Developed a website for the department of Computer Science and Engineering at Texas A&M. University, for enhancing the communication between student organizations and students.

Oct. 2017

- Implemented classic Machine Learning models using only NumPy
  - Implemented Naive Bayes classifier and KNN classifier for digit recognition.
  - Implemented Logistic Regression, Locally Weighted Logistic Regression, Perceptron, and Gaussian Mixture Models.

Jan. 2017

- Sentiment Analysis on IMDb movie review database
  - Developed and experimented different neural network architectures for classifying sentiments as positive and negative sentiment based on the movie review.

May 2015

- Anti-Forensics of JPEG Image Compression
  - Designed and developed neural network based Anti-forensic techniques in image compression to avoid digital footprints of techniques such as DCT and DWT on JPEG images.