Sanket Mhatre
https://www.linkedin.com/in/sanket-mhatre-782a081a6/
Email : sanketmhatre7738@gmail.com
Mobile : 7738005469

Experience

- **Microsoft** — US, Remote
  Research Fellow — *Dec 2024 - Present*
  - Researcher contributing to projects for **GitHub Copilot** and **Microsoft Excel**.
  - **Formula Completions in Excel**
    - Fine-tuning LLMs to build code-completion like feature for Excel formulas.
    - Exploring different fine-tuning strategies like **SFT**, **Distillation** and RL(**DPO, GRPO**) on OpenAI's GPT models, Qwen models and Phi models.
    - Preparing **large scale datasets** for different fine-tuning experiments with **40M+ Excel files**(avg token count per file - 10k).
    - Developed an **experimentation framework** to prepare a common code base to try out different context building techniques, different prompts, different evaluators, support different inference mechanisms and prepare training and testing data at scale.
    - Exploring different context building mechanisms to find optimal and token efficient representations of spreadsheets for different Excel tasks.
    - Advisors: Gust Verbruggen, Dr. Gustavo Soares, Dr. Sumit Gulwani
  - **SWE-Sharp-Bench** (Paper, HuggingFace, AIWare 2025)
    - **SWE-Bench** like benchmark for Software Engineering benchmark for **C#** curated from open-source and internal data. Public version was released as SWE-Sharp-Bench on huggingface.
    - Created pipelines to extract, harness and evaluation of benchmark instances supporting different sources.
    - Evaluated various models and agents on different coding benchmarks, developed a system for trajectory analysis to understand agent performance bottlenecks and common failure patterns.
    - Worked on creating a **distributed code execution service** to provision infrastructure and common interface for different agents to schedule runs on different benchmarks in a common reproducible environment at scale. Built tooling for trajectory visualization, agent call stacks and other metrics to help analyse agent performance.
    - Worked very closely with the Visual Studio Code team to help build their C# Coding agent by building specialized evals and trajectory analysis tools.
    - Advisors: Dr. Gustavo Soares, Dr. Emerson Murphy-Hill, Dr. Sumit Gulwani

- **PocketFM** — Bengaluru, India
  SDE-2 — *July 2022 - Nov 2024*
  - Led a small team for R&D for integration and experimentation of various AI applications. Tinkered with various open and closed source LLMs to automate various content generation tasks.
  - Developed a framework to localize long form content using **multi-agent workflows** and knowledge graph based **RAG** systems.
  - Built pipelines for dataset creation for fine-tuning open-source LLMs like LLama-3 for content creation tasks using **TPUs** with GKE+XPK setup with MaxText.
  - Automated Video Creation process by creating an in-house AI powered video editor using **Stable Diffusion image models** and **video diffusion models** and also **fine-tuning LORAs** for specific image generation tasks. Improved the TAT of video generation from 1 day to 30 mins.
  - Built an end to end pipeline using text to speech models from ElevenLabs and Play.ht for automated audio show creation with AI voice-overs and AI background music. Scaled this pipeline to generate **500 hours** of audio content per week. This helped moving 50% of the audio show production to this automated pipeline and drastically reduce production costs for test shows.

- Fine-tuned small transformer models for tasks like **NSFW detection** in user generated content and **spam and profanity detection** for in-app comments. These models reduced 80% of visibility of such entities in the app.
- Built an **event ingestion service** used for near real-time analytics. This service handles 2M+ requests at peak time and also handles validation layer for filtering data for the subsequent ETL pipelines.
- Techstack - Python, Django, FastAPI, PyTorch, Go, MySQL, Kafka, Redis, Celery, RabbitMQ, Airflow, AWS

Publications

• Aayush Kumar* and **Sanket Mhatre***, **"UnWEIRDing LLM Entity Recommendations"** - AAAI 2026, LM4UC Workshop.

• **Sanket Mhatre***, Yasharth Bajpai*, Sumit Gulwani, Emerson Murphy-Hill, Gustavo Soares - "**SWE-Sharp-Bench: A Reproducible Benchmark for C# Software Engineering Tasks". - Best Paper Award** - AIWare 2025

• **S. Mhatre** and A. Masurkar, "**A Hybrid Method for Fake News Detection using Cosine Similarity Scores**," 2021 International Conference on Communication information and Computing Technology (ICCICT), 2021, pp. 1-6, doi: 10.1109/ICCICT50803.2021.9510134. - ICCICT 2021

Education

- **Vidyalankar Institute of Technology** — Mumbai, India
  B.E in Electronics Engineering — *July. 2018 – July 2022*

Skills

- **Languages**: Python, Go, JS/TS, C#, C++
- **Libraries/Frameworks**: Django, FastAPI, PyTorch, JAX, ReactJS
- **Cloud Platform**: AWS, GCP, Azure
- **Databases**: MySQL, Postgres, MongoDB, Cassandra, Scylla