Vijay Murari Tiyyala

vijaymurari.tiyyala@gmail.com | +12405711678 | LinkedIn Profile | Personal Website

Education

Master's in Computer Science - ***Johns Hopkins University,*** *Baltimore* Dec 2023
Focus: Machine Learning, Natural Language Processing, Information Retrieval, Statistics

Bachelor of Technology in Computer Science - ***VR Siddhartha Engineering College,*** *India* Jun 2021

Publications

In Progress/Preprint

1. **BabyData: Exploring the Trade-offs Between Dataset Size and Pre-training Progression: Insights into Fine-tunability and In-context Learning**
   Vijay M. Tiyyala*, Kaiser Sun*, Naomi Saphra, Jessica Forde, Mark Dredze [In Progress]
2. **Training and Aligning Large Language Models with Augmented Clinical Responses: Improving Empathy, Accuracy, and Trust in Healthcare Communication**
   Matthew R. Allen*, **Vijay M. Tiyyala***, Nimit Desai, Karthik Ramesh, Job Shiach, Mark Dredze, Mike Hogarth, John W. Ayers [In Progress]
3. **Evaluating Clinician and AI Chatbot Responses to Clinical Questions Posed in a Health System Using the CREATE TRUST Framework**
   Armaan Johal, Atharva Yeola, **Vijay M. Tiyyala**, ..., Mark Dredze, John W. Ayers [In Progress]

Publications

1. **Kreyòl-MT: Building MT for Latin American, Caribbean, and Colonial African Creole Languages**
   Nathaniel R. Robinson, Raj Dabre, ... **Vijay M. Tiyyala** ... Sanjeev Khudanpur, Stephen D. Richardson, Kenton Murray [*NAACL 2024*]
2. **AnaloBench: Benchmarking the Identification of Abstract and Long-context Analogies**
   Xiao Ye, Andrew Wang, Jacob Choi, Yining Lu, Shreya Sharma, Lingfeng Shen, **Vijay M. Tiyyala**, Nicholas Andrews, Daniel Khashabi [*EMNLP 2024*]
3. **Waldo: Automated Discovery of Adverse Events from Unstructured Self Reports**
   Karan Desai*, **Vijay M. Tiyyala***, ..., Mark Dredze, John W. Ayers [*JAMIA Under Review*]
4. **HIVTrends.org: Public, Real-Time, and Validated HIV Testing Sales Trends from Search Query Surveillance**
   Vijay M. Tiyyala, Atharva Yeola, Karan Desai, Nimit Desai, Mathew R. Allen, Vin Somasundaram, Mark Dredze, Mike Hogarth, Nadir Weibel, Ravi Goyal, Davey M. Smith, John W. Ayers [*JAMIA Under Review*]

Posters

- Poster Presenter, **8th Annual Johns Hopkins Research Symposium on Engineering in Healthcare**
- Presented research on **The CREATE TRUST Framework** for enhancing patient-clinician communication using AI chatbots.
- Engaged with attendees to discuss challenges and future directions for healthcare-focused NLP solutions.

Research Experience

NLP Research Assistant - ***Center for Language and Speech Processing,*** *Full-Time* May 2023 – Present
- Developed an **empathetic medical chatbot** using **LLaMA3**, achieving **88.7%** accuracy on human-annotated test data for healthcare communication.
- Fine-tuned LLMs with **DPO** and **RLHF** for safer medical responses, reducing compute costs by **50%** using **PEFT/QLoRA**.
- Built a **RAG** system for Tobacco Watcher using **Apache Solr Cloud**, achieving **500ms** latency for **100+ concurrent users**.
- Implemented distributed training on **A100 GPU clusters**, optimizing performance with **PyTorch/SLURM**.
- Migrated classifiers to **XLM-RoBERTa** for multilingual capabilities, improving accuracy by **2.5%**.

Applied ML Researcher - ***Qualcomm Institute, UC San Diego,*** *Part-Time* Jan 2024 – Nov 2024
- Led development of **Adverse AI**, achieving **97.5%** accuracy in adverse event detection from medical texts.
- Created **HIVTrends.org**, a health surveillance platform achieving **0.87 R^2** in trend prediction.
- Developed ML pipelines using **Ridge**, **Lasso**, and **XGBoost** for health trend analysis.
- Implemented comprehensive anomaly detection for model validation and monitoring.

NLP Research Engineer - ***Johns Hopkins University,*** *Part-Time* Jan 2023 – Jun 2023
- Improved machine translation accuracy to **86%** for medical terminologies in low-resource languages.
- Analyzed **15,000+** compound words, enhancing English translation quality.
- Designed and implemented a **300+** language translation pipeline with compound splitting algorithms.

Industry Experience

Machine Learning Engineer - ***Hanwha Qcells,*** *Full-Time* Aug 2024 – Present

- Built end-to-end ML systems processing **5M+ images daily** across multiple production lines for quality control.
- Developed ML models for defect detection, junction box classification, and pyramid segmentation.
- Created **Flask**-based feedback dashboard enabling targeted data collection and model retraining.
- Automated ML workflows with **Apache Airflow** and **CI/CD** pipelines.
- Implemented production monitoring using **Metabase** and **Spotfire**.

AI Engineer - ***BotDojo,*** *Full-Time* Mar 2024 – May 2024

- Developed a **RAG**-based no-code platform for creating custom chatbots with user data.
- Built AI evaluation systems and node components using **TypeScript**.
- Integrated chatbots with Teams and Slack, ensuring robust functionality.

Business Technology Analyst - ***Deloitte USI,*** *Full-Time* Jul 2021 – Jun 2022

- Developed data integration solutions reducing processing time by **20%** through SQL optimization.
- Created analytical dashboards and reports using **PowerBI**.
- Improved client retention by **30%** through enhanced analytics solutions.

Volunteer Experience

- Reviewer EMNLP 2023
- Course Assistant for Machine Translation course taught by Prof. Philipp Koehn

Projects

Cannabis Use Detection in Clinical EMR - ***Python, PyTorch, Git***

- Trained NLP models such as BERT, RoBERTa, and ClinicalBERT to increase detection accuracy of cannabis use in EHRs by 97%.
- Achieved 92% accuracy in distinguishing medicinal and recreational cannabis use from unstructured text, enhancing data quality.
- Collaborated with clinical researchers to validate model outputs, ensuring compliance with HIPAA and high data fidelity.

Adverse AI: Automated Discovery of Adverse Event Reports from Unstructured Text - ***Python, PyTorch, Git***

- Led the development of 'Adverse AI', achieving 97.5% accuracy in identifying adverse events from diverse text sources including medical reports and social media.
- Automated extraction and analysis of adverse event data by training models like BERT and RoBERTa, reducing manual review time by 90%.
- Open-sourced the tool to enable widespread adoption and continuous improvement by the healthcare community.

HIVTrends.org: Real-Time HIV Testing Trends from Search Query Surveillance - ***Python, PyTorch, Git***

- Developed a real-time HIV testing trends platform using search query data, achieving an **Adjusted R^2 of 0.87** in predicting testing patterns.
- Engineered data pipelines to preprocess and align search query data with HIV testing kit sales, enhancing model training efficiency.
- Implemented **Ridge**, **Lasso**, and **XGBoost** models to accurately predict HIV testing trends, improving public health surveillance.
- Validated model predictions with anomaly detection, correlating spikes with major HIV awareness events, thereby improving predictive reliability.

SAMOYEDS - ***Python, PyTorch, HuggingFace, Git, Flask, HTML/CSS, JavaScript***

- Led the design and development of the SAMOYEDS application, a policy simulation tool using LLMs focusing on public health.
- Enabled SAMOYEDS to simulate diverse human personas, predicting public health policy responses with **76% accuracy**, enhancing policymaker decision-making.

Technical Skills

- ML/NLP: PyTorch, TensorFlow, HuggingFace, LangChain, PEFT, LoRA, Deepspeed, MLflow
- Infrastructure: Docker, AWS, Azure, GCP, Kubernetes, Airflow, SLURM
- Languages: Python, Java, R, C++, C, TypeScript
- Data Systems: SQL, PostgreSQL, Apache Solr, Spark, Hadoop, Elasticsearch
- Development: Git, CI/CD, Linux, Shell Scripting, HTML/CSS