

# Piyush Vyas

8123914102 | pie.vyas@gmail.com | LinkedIn | GitHub

## EXPERIENCE

---

### Senior Machine Learning Engineer | Numerade Labs, Inc.

June 2023 - Present

- Lead the ML team, overseeing the hiring, mentoring, and guidance of engineers and interns. Responsible for strategic vision and execution of machine learning initiatives, enhancing product capabilities and technological advancements.
- Engineered an AI assistant using **Large Language Models (LLMs) and Python** to generate media-rich outputs such as plots and diagrams. Implemented **Retrieval-Augmented Generation (RAG)** for personalized in-context responses. **Increased retention by 25% and boosted student engagement by 33%.**
- Upgraded the text embedding model to improve recommendations with just half the dimensionality. Replaced 100M+ vectors in the Milvus server and integrated them via a Flask API. **Reduced similarity search time by 40% and server costs by 50%.**
- Prototyped a **AI Avatar/Talking Head system**. Conducted vendor research, initiated RFI processes, scale-tested APIs, and built a webhook for real-time updates to ensure seamless integration and scalability.
- Integrated **Text-to-Speech** models into an advanced Generative AI video pipeline, orchestrating the automated generation of immersive educational content.

### Senior Machine Learning Scientist | Spectrum Labs, Inc.

Oct 2021 - April 2023

- Led the development and enhancement of text and voice chat moderation systems, including pioneering multilingual support and fine-tuning models for voice chat moderation, resulting in an **18% increase in moderation accuracy.**
- Trained proprietary LLMs that became foundational for text classifiers, **increasing PR-AUC by 24%.** Collaborated cross-functionally to fine-tune, evaluate, and deploy these models into production.
- Engineered and optimized machine learning solutions for real-time applications using **ONNX, TensorRT, FastAPI, Docker, and EC2.** Implemented **quantization and distillation** techniques to enhance performance and efficiency.
- Designed a ML toolkit for audio and text models, built on PyTorch and ONNX. The toolkit was crafted for ongoing research and model evolution, quickly becoming an essential daily tool for the ML team.

### Research Intern | Comcast Applied AI

Aug 2021 - Oct 2021

- Worked on the Automatic Speech Recognition engine powering Comcast's X1 Entertainment System. Incorporated the wav2vec2.0 model, **reducing the Word Error Rate by a relative 30%.**

### Research Engineer | Indiana University

May 2020 - Aug 2021

- Led the crowd-sourcing of human speech intelligibility data using **Amazon Mechanical Turk** and conducted experiments to develop and optimize deep learning models for predicting speech intelligibility and speech quality.
- Developed a comprehensive evaluation framework to benchmark model performance using standard metrics such as **Mean Opinion Score (MOS)** and **Short-Time Objective Intelligibility (STOI)**. Published technical reports and academic papers, presenting findings at international conferences.

## PUBLICATIONS

---

R. Tang, K. Kumar, J. Xin, **P. Vyas**, et al., "Temporal Early Exiting for Streaming Speech Commands Recognition," in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing*, pp. 7567-7571, 2022.

**P. Vyas**, A. Kuznetsova, and D. Williamson, "Optimally Encoding Inductive Biases into the Transformer Improves End-to-End Speech Translation" in *Proc. INTERSPEECH*, pp. 2287-2291, 2021. (**Best Student Paper Award**)

Z. Zhang, **P. Vyas**, X. Dong and D. Williamson, "An End-To-End Non-Intrusive Model for Subjective and Objective Real-World Speech Assessment Using a Multi-Task Framework," in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing*, pp. 316-320, 2021. (**Outstanding Student Paper Award**)

## SKILLS

---

Python, SQL, OpenAI/GPT, LangChain, FAISS, Milvus, PyTorch, ONNX, WandB, Hydra, HuggingFace, DeepSpeed, TensorRT, FastAPI, Flask, Docker, AWS, Azure, JAX, TensorFlow, L<sup>A</sup>T<sub>E</sub>X, Snowflake, Swift, Firebase

## EDUCATION

---

### Indiana University Bloomington

M.S. in Intelligent Systems Engineering

2020

### Devi Ahilya Vishwavidyalaya

Bachelor of Commerce

2017