VINAYAK KUMAR

Machine Learning Research Engineer

+49 1794111524 • vinayaksharan@outlook.com • https://www.linkedin.com/in/vinayak-kumar • https://vinayak-sharan.github.io
• Pforzheim, Germany

Summary

Machine Learning Research Engineer with **7+ years** specializing in **Computer Vision** & **Deep Learning**. Expertise includes GANs, Transformers, Diffusion Models, VAE/VQ-VAE, and Linear Attention architectures, focused on real-time facial animation and gaze correction. Proficient in PyTorch, Core ML, CUDA, ONNX, Python and distributed multi-GPU training. At **Casablanca.AI**, own end-to-end R&D from ideation and rapid prototyping to production deployment on macOS.

Successfully completed my Master's degree while working full-time, demonstrating strong time-management skills.

Experience

CASABLANCA.AI
Pforzheim, Germany

Machine Learning Research Engineer
02/2021 - Present

- Pioneered a **Transformer-based** deep-learning pipeline for real-time **head & gaze correction** in video calls, enabling precise **head and gaze manipulation** to any desired angle, delivering natural eye contact and **establishing the product's core USP.**
- **Reduced inference complexity from quadratic to linear** by integrating linear-attention mechanisms, cutting compute and memory costs on production hardware.
- **Designed and orchestrated a distributed PyTorch training stack** that processed **200 M+ talking-head frames**, accelerating model convergence 3× on multi-GPU clusters.
- **Ported models to Core ML and optimized for Apple Neural Engine (ANE),** lowering per-frame latency from ~30 ms (GPU) to **~15 ms (ANE)** in macOS.
- **Shipped a production macOS desktop app** (PyQt GUI + deep-learning backend), owning the full CI/CD pipeline and App Store deployment.
- **Researched and implemented head- & gaze-redirection techniques** using Adapters, Diffusion Models, optical-flow, and image-warping, boosting accuracy by 12%.
- **Added generative-AI features** such as avatar animation via Diffusion, VAE, and StyleGAN, expanding the product roadmap.

Otto-von-Guericke-Universität
Magdeburg, Germany

Teaching Assistant- Introduction To Deep Learning Course
04/2020 - 07/2020

- Managed assignments, Q&A sessions, and practical workshops.
- Facilitated learning outcomes for ~15 students, covering advanced deep learning concepts.

Sony India/ Ignitarium Technology
Bengaluru

Senior Machine Learning Engineer
12/2018 - 09/2019

- Worked for Sony's Software Architecture Division SARD.
- Developed a Spatio-Temporal Graph Neural Network (ST-GCN) model leveraging OpenPose for recognizing player actions in sports videos, targeting large-scale tennis video analysis.

Hexaware Technologies
Chennai

Machine Learning Engineer
03/2018 - 11/2018

- Intelligent Contract Manager: Developed and deployed machine learning models for classifying various contract types and their respective clauses.
- Web Scraping Financial News: Created a financial news scraping system for sources such as Reuters, MarketWatch, and Yahoo Finance, deploying the data via Flask and MongoDB.
- Implemented invoice image data extraction utilizing SpaCy and Google OCR, enhancing accuracy with a custom attention model (POC stage).
- Built an ERISA document classifier using NLTK and KNN algorithms, deploying it through a Flask-based application.

Infosys Limited
Bengaluru

Senior System Engineer
10/2015 - 03/2018

- Developed key application features for Cisco's Prime Collaboration Deployment (PCD), a tool for managing and upgrading Cisco's CUCM VoIP systems.
- Built backend components for FANUC's Factory Operational Efficiency Dashboard, enabling factory-level, line-level, and machine-level monitoring.

Education

Otto-von-Guericke-Universität
Magdeburg, Germany

M.Sc. in Digital Engineering (specialization in Generative Models)
10/2019 - 03/2025

Education

BVB College of Engineering and Technology

Hubli, India

Bachelor of Engineering

08/2011 - 06/2015

Skills

Machine Learning:

GAN(s) · Diffusion Models · Transformers · Score Matching · GANs · VAEs & VQ-VAE · Markov Chains · Autoregressive Models (PixelCNN/ MADE) · Flow Models (NICE/ RealNVP) · RBMs · BERT · Object Detection (YOLO/SSD) · Resnet · Densenet · EfficientNet · Stereo Vision · SIFT/ Harris descriptor · HOG · Linear/Logistic Regression · Naive Bayes · SVM · KNN · Random Forest · Clustering · Semi-Supervised learning · Reinforcement Learning · Deep Learning · Face Detection · CNN · RNN · CoreML · ONNX

Software Engineering:

Python · Algorithm & DS · Shell Scripting/UNIX · MongoDB · GitHub · SQL(Basics) · Java (Basics) · C++(Basics) · Flask(basics)

Frameworks/ Platforms: PyTorch · Tensorflow · Numpy · OpenCV · PyTorch3D · Sklearn · Pandas · Beutifulsoup · Mediapipe · Google Cloud (GCP)

Publications, Thesis and Academic Projects

Publication & Awards

- **Paper Published**: Winner, **DI2KG Challenge 2020**, published in proceedings of **VLDB Tokyo Conference 2020**..
- **Master's Thesis**: *Head and Gaze Controllable Implicit Motions* – Developed a cross-attention-based model for precise control of head and gaze motions; research confidential, publication planned after a two-year embargo period.
- **Best Team Award**(4-member team): Sony India's Software Architecture Division, recognized for excellence in collaborative project execution.

Master's Thesis @ AILAB OVGU and Casablanca.AI

10/2024 - 02/2025

- My master's thesis builds on Implicit Motion Functions to develop controllable facial animation using efficient Linear Attention, improving the FID to 4.77 (vox1) and speed by 2x.
- The system utilizes antialiasing techniques like BlurPool and PixelShuffle-like upsampling to ensure smooth, artifact-free outputs, while masked loss is applied to key facial regions such as the lips and eyes to enhance detail.
- Facial attributes, including head movement and gaze, are controlled through head and gaze tokens.
- This attention-driven approach mitigates the artifacts seen in keypoint-based and dense motion flow methods, producing more natural and lifelike animations suitable for real-time applications.
- *Note: Casablanca.AI policies prevented publishing these findings, keeping the research confidential.*

Project: Introspection of BERT

Categorizing and explaining prediction errors of image recognition models.

- Visualization of possible biases in BERT using Visual Saliency Maps, LRP, etc.

Project: Deep Hash

Entity Resolution for Unstructured Data.

- Developed a model using triplet loss to measure similarity using image encoding and produce hash codes for the same.

Project: Categorizing and explaining prediction errors

Introspection of State of the Art Models

- ObjectNet is a dataset on which accuracy of SOTA drops around 40-50%. Introspection of same using Visual Saliency Maps, LRP, etc

Entity Resolution for Unstructured Data

Contextual embeddings were explored Entity Resolution.

- It is difficult to identify the same entities (as data is dirty) on different websites.
- Rule-based methods with weak supervision as well as classification with the help of contextual embeddings from BERT were explored for the optimal solution.