

DIVYANSHU GOYAL

Sunnyvale, CA ◊ (+1) 404-414-3762 ◊ divyanshu.g25@gmail.com

LinkedIn ◊ GitHub ◊ Google Scholar

SUMMARY

Machine learning researcher specializing in **Large Language Models**, **Vision-Language Models**, and **multimodal reasoning systems**. Focused on novel training methodologies, evaluation frameworks, and alignment techniques for foundation models. Experienced in **model training**, **fine-tuning**, **RLHF**, and **distributed systems**, with a track record of research contributions and deployment of AI systems at scale.

EDUCATION

Georgia Institute of Technology, Atlanta August 2019 – May 2021
M.S. in Computer Science, Specialization in Machine Learning **CGPA: 4.0 / 4.0**

Birla Institute of Technology and Science (BITS), Pilani August 2011 – June 2016
B.E. in Computer Science
M.Sc. in Mathematics **CGPA: 8.75 / 10.0**

PUBLICATIONS & PATENTS

Curiosity-Driven LLM-as-a-judge for Personalized Creative Judgment arXiv 2025

Vanya Bannihatti Kumar, Divyanshu Goyal, Akhil Eppa, Neel Bhandari
Novel curiosity-driven LLM-as-a-judge framework for evaluating creative writing that personalizes assessments to individual judgments.
arxiv.org/abs/2510.05135

Parameter-Efficient Model Serving Systems US Patent App. 2025

Yuanyou Wang, Naveen Vangala, Mayank Anand, Kunal Kumar Jain, Jose Mathew, Eapen Jose, Divyanshu Goyal, et al.
System for efficiently serving fine-tuned models by dynamically loading parameter-efficient layers into pre-loaded base models, minimizing network transfer costs and container startup time.
patents.google.com/patent/US20250217193A1

Brand-Aligned Content Generation with Structured Knowledge US Patent App. 2025

Mayank Anand, Jose Mathew, Divyanshu Goyal
System for transforming unstructured brand information into structured knowledge and generating brand-aligned marketing content using generative models and NLP techniques.
patents.google.com/patent/US20250117829A1

EXPERIENCE

Applied Scientist - Adobe Unified Platform, San Jose June 2021 – Present

Vision-Language Model Training

→ Designed and trained a **13B-parameter vision-language model** from scratch on **2M image-text pairs** across **8× H100 GPUs**, using novel **observable injection techniques** that augment prompts with quantitative image metrics (sharpness, color balance, exposure, lighting distribution).

→ Fine-tuned CLIP on custom vision-language datasets to improve fine-grained attribute recognition, applying contrastive learning and image-text augmentation to boost **F1 accuracy by 20%** in multi-label classification tasks.

→ Built distributed training infrastructure and implemented LoRA-based parameter-efficient fine-tuning methods.

→ Developed **MQCore evaluation benchmark** inspired by DCLM CORE for rigorous assessment of model performance on domain-specific quality tasks.

LLM Alignment & Evaluation Systems

→ Created curiosity-driven LLM-as-a-judge framework for personalized creative evaluation, published in arXiv 2025.

→ Designed LLM-based evaluators for automated quality assessment with strong agreement to expert annotations.

→ Developed advanced prompting and alignment techniques achieving **18%** improvement in human-likeness metrics and **25%** improvement in response adherence.

→ Built human-in-the-loop evaluation frameworks and synthetic data generation pipelines for large-scale model benchmarking.

Creativity & Multimodal Reasoning Research

→ Pioneered creativity modeling using transformer architectures with Bayesian surprise as a quantifiable novelty signal.

→ Achieved **23% improvement in F1 score** and **38% improvement in correlation metrics** for creativity assessment tasks.

→ Leading development of multimodal agent systems capable of complex reasoning and action orchestration for Photoshop image editing workflows.

Production AI Systems & Impact

→ Led development of **Adobe GenStudio for Performance Marketing**, launched October 2024 as a generative AI platform integrating LLMs and diffusion models for enterprise-scale content creation.

→ Platform now serving major enterprise clients including Microsoft, AT&T, and other Fortune 500 companies.

Machine Learning Engineer - Adobe Inc., Bangalore

July 2016 – August 2019

→ Built a Model Management service on top of MLflow and Databricks, enabling end-to-end model lifecycle management including versioned deployment, real-time monitoring, and developer tooling for reproducible model promotion across staging and production environments.

→ Co-engineered Adobe's User Response Prediction Service, a high-throughput ML inference platform serving **3,500 QPS/node** at **0.3–0.6ms app-side latency** (99th percentile), using CPU flame-graph profiling, JMH micro-benchmarking, and lock-free architectures.

→ Designed a GC-free asynchronous event processing pipeline using the LMAX Disruptor ring buffer pattern, decoupling logging and feedback from prediction threads to eliminate back-pressure under peak load.

→ Reduced garbage collection pauses to **~30ms per 20s** for recommendation models (FFM) through ThreadLocal object reuse, custom GC-free JSON serialization, and JVM generation sizing, enabling sustained high throughput with minimal memory overhead.

TECHNICAL SKILLS & RESEARCH EXPERTISE

Research Areas: Large Language Models, Vision-Language Models, Model Fine-tuning & RLHF, Distributed Training, Multimodal Reasoning

ML Frameworks: PyTorch, Hugging Face, DeepSpeed, CUDA, Weights & Biases, LangChain, Docker

Programming: Python, C++, JavaScript