Bhavin Jawade, Ph.D

Research Scientist | Netflix
Multimodal LLMs & Post-Training
350 E. Taylor St. Apt 4208•San Jose, CA 95112•+1 716-495-9321•bhavinjawade@gmail.com
LinkedIn - linkedin.com/in/bhavinjawade•**Website -** bhavinjawade.github.io•**Visa -** O1
Github - github.com/bhavinjawade•**Google Scholar** - Bhavin Jawade•**Twitter/X**- x.com/BhavinJawade

Summary

Research Scientist with expertise in multimodal foundation models, LLM post-training, reward modeling, and synthetic data generation. Led research strategy and end-to-end development of AI systems at Netflix, spanning post-training, human annotation, evaluator calibration, LLM-as-a-Judge systems, reward modeling, and RLVR. Published at leading ML venues including NeurIPS, WACV, IJCB, and ICASSP, with best paper awards and a track record of translating novel research into impactful products used at scale.

Education

University at Buffalo, SUNY May 2025
PhD. - Computer Science | *Advisor - Dr. Venu Govindaraju* GPA: 3.9

University at Buffalo, SUNY May 2021
MS. - Computer Science | *Advisor - Dr. Sargur Srihari* GPA: 3.8

Work Experience

Netflix Research | Senior Research Scientist | *Los Gatos, CA* Jan 2025 - Present

- Led research strategy and end-to-end development of multimodal LLM systems for localization, dubbing, and subtitling, translating ambiguous model-behavior and content-quality problems into rigorous research formulations, annotation protocols, reward models, and production systems.
- Built production LLM-as-a-Judge and reward-modeling systems for translation quality evaluation, combining human annotation data, MQM-style quality signals, synthetic data training, on-policy distillation, and RLVR.
- Developed a novel RLVR and verifier-guided post-training pipelines for multimodal translation and dubbing models, using translation-quality and lip-sync judges as reward signals to improve groundedness, fluency, timing, and human acceptance.
- Designed synthetic data generation and filtering pipelines for strong distillation, using judge models to select high-quality training data from noisy machine-generated and human-authored candidates before fine-tuning production LLMs.
- Built OLT-LLM and KNP Agent, near-real-time LLM systems for automated subtitling and key-name/phrase correction, serving 5,000+ hours of Netflix content, improving subtitle timing/text accuracy by 25%+ over prior SOTA, and driving $50M+ in estimated operational savings.
- Developed a multimodal DubScript machine translation system that conditions on video, source dialogue, and lip-motion/viseme signals to generate target-language scripts aligned with original actor lip motion.
- Proposed inverse supervised pretraining for DubScript generation, creating large-scale synthetic supervision by reconstructing source dialogue from viseme-aligned and paraphrased semantic inputs before fine-tuning on filtered expert-written dub scripts.
- Achieved 55% exact-match rate with expert human linguists and 0.2 HTER on DubScript generation, outperforming pretrained frontier models and historical human-written baselines in expert human-in-the-loop evaluation; deployed across 1,000+ hours of Netflix content and 22+ dubbing languages.
- Designed human annotation and evaluator calibration frameworks for subjective translation, dubbing, and lip-sync quality tasks, converting expert linguist and adapter judgments into reliable training, evaluation,

and reward-modeling signals.
- Built Netflix internal Training Recipe Cookbook on top of MS-SWIFT, Open-Instruct, and other training frameworks, providing reusable implementations for SFT, OPD, OPSD, RL, GRPO, RLVR, DAPO, rubric-based RL, Rubric-ARM, and other post-training strategies, with multi-instance distributed training, evaluation harness, and inference optimization across B200, H200, and A100 GPUs.
- Submitted 6+ patent applications, presented research at Netflix AI/ML Summit and ICASSP, and mentored a research intern on video-to-text lip-sync evaluation using lip-reading likelihood as a multimodal alignment signal.

Netflix Research | Research Scientist Intern | *Los Gatos, CA* May 2024 - Dec 2024

- Multimodal (audio-visual) contrastive learning for speech to lip-motion alignment and quality estimation.
- Designed novel loss function "RSMS - Ranking Supervised Multi-Similarity loss" to incorporate positive's ranking prior within the contrastive learning optimization objective.
- Proposed method achieved 11.3% improvement over the baselines in discerning dubbing lip-sync quality on real-world dubs and other audio understanding tasks
- Work published at ICASSP 2025

Yahoo Research | Research Scientist Intern | *New York City, NY* May 2023 - May 2024

- Proposed SCoT, a novel pre-training strategy to improve zero-shot multimodal compositional retrieval and understanding.
- Performed LoRA based Instruction fine-tuning of quantized Falcon-7B, LLama-2, Mistral-8B, Gemma-2B and SOLAR-10B LLM models to generate synthetic textual triplets. Utilized contrastive representations from textual triplets as a proxy supervision.
- Approach outperformed existing methods by ~10.2% on the Fashion-IQ dataset and ~8% on CIRR benchmark in zero-shot retrieval setting while only utilizing image-caption pairs. Qualitative experiments visualizing SCoT's latents with UnCLIP demonstrate its implicit compositional properties.
- Research paper published at WACV 2025 (accepted in Round 1 - 12% acceptance rate)

Adobe Research | Research Scientist Intern | *San Jose, CA* May 2022 - Nov 2022

- Proposed a Large-Scale Vision-Language (multimodal foundational models) contrastive pre-training strategy to learn embeddings that generalize over out-of-distribution tasks.
- Introduced a novel loss formulation incorporating tags, metadata, and clauses to create a semantic hierarchy, enhancing the alignment of vision and language embeddings.
- Performed large-scale distributed pre-training using torch-distributed, web-datasets and gradient-checkpointing. Trained on over 100 million image-text pairs on 16 A100 GPUs.
- Method outperforms Open-AI's CLIP by 6.4% on zero-shot ImageNet retrieval (Top-1) when trained on 50M image-text pairs, 9.2% when trained on 15 million subset and 8.4% when trained on 5 million sub-set

Center for Unified Biometrics and Sensors | Research Assistant | *Buffalo, NY* Jun 2020 – Present

- PhD in Computer Science - Thesis Topic: Multimodal Contrastive Models Through Zero-Shot Training
- IARPA BRIAR Grant | Proposed a uni-modal feature fusion strategies - CoNAN and ProxyFusion CoNAN received **Best Paper Award, IEEE, IJCB 2023, Slovenia,** and **ProxyFusion was presented at NeurIPS 2024**
- Computer vision and vision-language models for retrieval, grounding and alignment. Works presented at WACV 2023 and WACV 2024, Hawaii.
- *NSF National AI Institute Grant* | Working on a novel benchmark and evaluation metric to estimate hallucination in text conditioned image generation diffusion models. DPO trained a SOLAR-10.7B LLM using LoRA (Ranked #2 on Hugging-face openLLMLeaderboard, 35K+ Downloads).
- NSF CITER Grant | Developed a multi-task learning based CNN network trained with Deep metric losses (Contrastive, AdaCos) along with minutiae loss to match contactless fingerprints captured using smartphone camera against legacy fingerprints.

Won Russell Agrusa Research Innovation Award, UB CSE, 2021. Received Best CSE PhD Poster Award (2022) and Best UB SEAS Poster Award (2023)

- NSF DIBBS Grant | Implemented full-stack machine learning framework called MLToolkit where the user can drag and drop the different elements to create and execute machine learning pipelines.

Persistent Systems | Machine Learning Engineer | *Pune, India* Jul 2019 - Dec 2019

- Built OCR based ML application for banks to auto-analyze POS invoices saving ~$200,000 & 100 hrs/week.
- Developed an Invoice Management tool for a leading Supply Chain firm to automate the supplier-distributor financing process and decrease finance approval time by 43%.

IIT Madras, Department of Computer Science | Research Intern | *Remote* Feb 2019 - May 2019
Advised by Dr. Rupesh Nasre, Professor IIT Madras.

- Designed a highly efficient algorithm for computing execution times in hierarchical task scheduling, targeting updates in million-node Directed Acyclic Graphs (DAGs).
- Achieved performance, with the algorithm processing 1,000 updates in a 1,000,000 node DAG in just 0.8 seconds, demonstrating near-linear O(n) complexity

Selected Publications

ProxyFusion: Efficient Face Feature Aggregation Through Sparse Experts Dec 2024
Conference on Neural Information Processing Systems (NeurIPS 2024)

SCOT: Self-Supervised Contrastive Pre-training For Zero-Shot Compositional Retrieval Mar 2025
IEEE/CVF Winter Conference on Applications of Computer Vision (WACV 2025)

NAPReg: Nouns as Proxies Regularization for Semantically Aware Cross-Modal Embeddings Jan 2023
IEEE/CVF Winter Conference on Applications of Computer Vision (WACV 2023) **(Oral)**

CoNAN: Conditional Neural Aggregation Network For Unconstrained Face Feature Fusion Sep 2023
IEEE International Joint Conference on Biometrics (IJCB 2023) **(Best Paper Award, Oral)**

Hear The Flow: Optical Flow-Based Self-Supervised Visual Sound Source Localization Aug 2022
IEEE/CVF Winter Conference on Applications of Computer Vision (WACV 2023) **(Oral)**

Audio-visual Representation Learning for Lip-Sync Estimation Through Ranking Augmented Contrastive Training Sep 2024
IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2025)

J-OPD: Jury On-policy Distillation, Learning from Black-Box LLMs through OPD Apr 2026
Under Review at Top-Tier ML Conference

RealCQA: Scientific Chart Question Answering as a Test-bed for First-Order Logic Feb 2023
IEEE/IAPR International Conference on Document Analysis and Recognition (ICDAR 2023)

DIOR: Dataset for Indoor-Outdoor Reidentification-Long Range 3D/2D Skeleton Gait Collection Pipeline, Semi-Automated Gait Keypoint Labeling and Baseline Evaluation Methods Aug 2023
IEEE International Joint Conference on Biometrics (IJCB 2024)

Liveness Detection Competition-Noncontact-based Fingerprint Algorithms and Systems Sep 2023
IEEE International Joint Conference on Biometrics (IJCB 2023)

GestSpoof: Gesture Based Spatio-Temporal Representation Learning For Robust Fingerprint Presentation Attack Detection. Sep 2023
IEEE International Conference on Automatic Face and Gesture Recognition (FG 2023)

Conditional Neural Aggregation Network For Unconstrained Long Range Person Feature Fusion Nov 2023
IEEE transactions on biometrics, behavior, and identity science (TBIOM) **(Invited Journal Paper)**

Deep Metric Learning for Computer Vision: A Brief Overview Dec 2022
Book Chapter in Handbook of Statistics, Special Issue - Deep Learning ***(Invited Book Chapter)***

RidgeBase: A Cross-Sensor Multi-Finger Contactless Fingerprint Dataset Jun 2022
IEEE/IAPR International Joint Conference on Biometrics (IJCB 2022)

Attribute De-biased Vision Transformer (AD-ViT) for Long-Term Person Re-identification Sep 2022
IEEE International Conference on Advanced Video and Signal-Based Surveillance (AVSS 2022)

Multi Loss Fusion for Matching Smartphone Captured Contactless Finger Images Sep 2021
IEEE International Workshop on Information Forensics and Security (WIFS 2021) ***(Oral)***

Low computation in-device geofencing using hierarchy-based searching for offline usage Nov 2018
IEEE International Conference on Inventive Computation Technologies' (ICICT 18') ***(Oral)***

Research Datasets

RealCQA, Chart Visual Question Answering Dataset 2022

RidgeBase, Contactless Fingerprint Dataset 2021

DIOR, Long-Distance Person ReID Dataset 2023

LivDet 23, Contactless Fingerprint Liveness Dataset 2023

GestSpoof, Spatio-Temporal Fingerprint Liveness Detection Dataset 2023

Achievements and Awards

IEEE Best Paper Award, IEEE IJCB 2023

Graduate Leadership Award, University at Buffalo 2023

Russell Agrusa Research Innovation Award, University at Buffalo 2021

Research Award: Best SEAS PhD Poster Award, University at Buffalo 2022

Research Award: CSE PhD Poster Award, University at Buffalo 2022

Hackathon Winner, Adobe Code Jam 2022

Best Social AI Idea Award: Blackstone Launchpad 2021

Hackathon Winner, Maple Ridge Hackathon (Govt of British Columbia) 2021

Hackathon Winner, WittyHacks Hackathon 2018

Project Award, Hon' Minister of State (Education Minister, Govt of M.P., India) 2018

Service

Area Chair, IEEE International Conference on Acoustics, Speech and Signal Processing, 2026 **(ICASSP)**

Reviewer, IEEE CVF Computer Vision and Pattern Recognition Conference, 2025 (**CVPR**)

Reviewer, International Conference on Machine Learning, 2025 (**ICML**)

Reviewer, Conference on Neural Information Processing Systems, 2024 **(NeurIPS)**

Reviewer, IEEE CVF Computer Vision and Pattern Recognition Conference, 2023 (**CVPR**)

Reviewer, IEEE CVF Winter Conference on Applications of Computer Vision, 2024 **(WACV)**

Reviewer, Association for Computational Linguistics, Rolling Review, 2024 **(ACL)**

Reviewer, International Conference on Learning Representations, 2024 **(ICLR)**

Reviewer, IEEE International Conference on Acoustics, Speech and Signal Processing, 2023 (**ICASSP**)

Reviewer, IEEE Transactions on Biometrics, Behavior, and Identity Science, 2023 (**TBiom**)

Reviewer, IEEE International Joint Conference on Biometrics, 2023 (**IJCB**)

Reviewer, IAPR International Journal on Document Analysis and Recognition, 2022 (**IJDAR**)

Reviewer, IEEE Multimedia, 2023 (**MM**)

Invited Lecture, Large Vision-Language Models, **Clarkson University**, 2023

President, Computer Science Graduate Student Association, 2021-22 **(CSE-GSA, UB)**

Writer, Towards Data Science **(TDS)**

Co-lead, Facebook Developer Circle Indore, 2017-19

Lead organizer, HackIndore, the largest hackathon of Central India, 2019

City Lead, Microsoft Student Partner, 2017-18

Founder and Head, SGSITS' first techno-learning club (#include), 2017-19

Head, Design, Entrepreneurship Cell SGSITS. 2017

Vice-Captain, Aeromodelling club SGSITS, 2016

Skills

Languages and Frameworks:	**Python**, **Pytorch**, TensorFlow, JavaScript, Go (Golang), C++, C, SQL, HTML, CSS, R, Dart, Angular 8, React, Laravel, PHP, Pandas, Dask, PySpark, Numpy, React Native, Android, Java Spring
Databases & Cloud:	SQL Server, MySQL, MongoDB, Redis, Cassandra, Solr, No-SQL, AWS, GCP
Software Tools:	Visual Studio Code, Visual Studio, Android Studio, Jupyter Notebook, Xcode, Cursor