

UJWAL DINESHA

Austin, TX | 917-292-2814 | ujwald36@tamu.edu

Google Scholar | [linkedin.com/in/ujwal-dinesha](https://www.linkedin.com/in/ujwal-dinesha) | github.com/flash-36 | tinyurl.com/ujwald

EDUCATION

Texas A&M University

College Station, TX

Ph.D. in Computer Engineering — Reinforcement Learning; Advisor: Dr. Srinivas Shakkottai

Aug 2021 – Jul 2026

- Defended July 2026. Coursework: Reinforcement Learning, Bandit Algorithms, Advanced Convex Optimization, Stochastic Systems, Classical Analysis, Analysis of Algorithms, Queueing Theory

Columbia University

New York, NY

M.S. in Electrical Engineering — Data-Driven Analysis & Computation; GPA: 3.8/4.0

Dec 2020

- Coursework: Neural Networks & Deep Learning Research, NLP, Reinforcement Learning, Sparse & Low-Dimensional Models. Graduate TA for NLP and Genomic Information Science & Technology

National Institute of Engineering

Mysuru, India

B.E. in Electronics and Communication — GPA: 9.04/10.00

Jul 2019

TECHNICAL SKILLS

ML: RLHF & preference optimization (PPO, DPO), inference-time alignment, LLM agents & evaluation harnesses, bandits/RMABs, multi-agent RL, imitation learning, distributed GPU/HPC training & inference

Tools: Python, PyTorch, Hugging Face (Transformers, TRL, PEFT), vLLM, TorchRL, Ray RLLib, LeRobot, Hydra, Weights & Biases, Slurm, Linux, Git, $\LaTeX$

RESEARCH EXPERIENCE

Graduate Research Assistant, Texas A&M University — selected projects; full publication list on page 2.

PITA: Preference-Guided Inference-Time Alignment for LLM Post-Training

arXiv preprint | 2025

- Built an alignment framework that learns a small guidance policy directly from preference feedback to reshape the base LLM's next-token distribution at decode time — eliminating both reward-model training and LLM fine-tuning.
- Raised GSM8K pass@1 of Llama-3-8B-Instruct from 69.1% to 77.1% using a 1B-parameter guidance model, matching a value-guided baseline (Q#) that requires ground-truth reward labels; achieved near-perfect (99.7–99.9%) accuracy on star-graph reasoning where DPO collapses to 0% and REINFORCE stays at chance level.

Agentic Harnesses for ARC-AGI-3

Ph.D. thesis | 2026

- Led the design of ARC-SCOPE, a budget-controlled testbed comparing LLM agentic harnesses (reactive, multi-agent, executable-world-model) on ARC-AGI-3 interactive games under matched interaction, token, and wall-clock budgets; served all agents via vLLM (FP8), from single-GPU Qwen-27B to GLM on an 8×H200 node.
- Built DISH, an event-driven executable-world-model harness targeting sub-30B models: cleared 28–31 of 37 levels at roughly half the token cost per level of the strongest frontier-style baseline ($\approx 20M$ vs. $\approx 43M$ tokens/level).

Risk-Averse Fine-Tuning of Large Language Models (RA-RLHF)

NeurIPS | 2024

- Led the empirical study of risk-averse RLHF: optimized Conditional Value-at-Risk (CVaR) of trajectory rewards via a soft-risk PPO schedule, directly suppressing the rare-but-severe toxic generations that risk-neutral RLHF tolerates.
- Outperformed RLHF, Quark, and DExperts on the worst-quantile prompts of IMDB, Jigsaw, and RealToxicityPrompts generation tasks while preserving perplexity and diversity; scaled from GPT-2 to GPT-J-6B with LoRA + bf16.

DOPL: Direct Online Preference Learning for Restless Bandits

ICLR | 2025

- Co-designed the first online RL algorithm for restless multi-armed bandits driven purely by pairwise preference feedback (Pref-RMAB), with provable $\tilde{O}(\sqrt{T \ln T})$ regret — matching the scalar-reward setting.
- Designed and ran the empirical evaluation: a preference-inference scheme cuts required comparisons per decision epoch from $O(B^2)$ to $B - 1$, achieving near-oracle reward on CPAP therapy-adherence, ARMMAN maternal-health, and app-marketing benchmarks.

INDUSTRY EXPERIENCE

Typeface AI

Palo Alto, CA

Applied Scientist Intern

Jun 2025 – Aug 2025

- Adapted latent-preference inference research to Typeface's marketing-content products, personalizing future LLM generations from users' past edits without retraining the base model.
- Created synthetic user-edit datasets and ran offline experiments to validate the approach and set parameters.
- Implemented the personalization logic in the production codebase and deployed to beta behind a feature flag.

InterDigital

Los Altos, CA

Research Intern, AI Lab

Mar 2021 – Aug 2021; May 2022 – Aug 2022

- Improved learned video codecs for live streaming by adding convolutional-recurrent blocks and attention mechanisms to the inter-frame prediction branch.
- Built an RL-based rate controller for end-to-end neural video compression; internship work led to multiple patent filings (see Patents).

Roche

Little Falls, NJ

Deep Learning Intern

May 2020 – Aug 2020

- Weakly-supervised learning on gigapixel histopathology images for cancer detection (ResNet34 encoder + RNN aggregator); reached 99.3% AUC on an in-house kidney-cancer dataset and presented to the pRED Innovation Center.

ADDITIONAL RESEARCH & PROJECTS

Robot Learning: Dexterous Manipulation & Autonomous Navigation

Texas A&M

- Built an imitation-learning navigation stack for the comma body robot: a VQ-GAN visual tokenizer feeding a GPT-2-style transformer acting as joint policy and world model; the robot autonomously drives laps of the lab from self-collected teleoperation demonstrations.
- Mentored an undergraduate team on manipulation with LeRobot and SO-101 arms, training ACT (ALOHA), Diffusion Policy, and π_0 policy architectures.

Multi-Agent Wireless Video Streaming with Delayed Feedback

INFOCOM | 2024

- Decomposed a constrained multi-agent POMDP for energy-constrained video streaming into independent transmitter-receiver problems via Lagrangian duality and the common-information approach, enabling distributed learning.
- Trained delay-aware LSTM policies over approximate information states, robust to channel variation via domain randomization; at feedback delays of 6+ timesteps, only the delay-aware agent sustained positive QoE reward, with fewer playback stalls and shorter buffers than the delay-oblivious baseline.

EdgeRIC: Real-Time AI Control for NextG Cellular Networks

NSDI | 2024

- Designed DigitalTwin, a full-stack trace-driven emulator for offline RL policy training, and trained the PPO MAC-scheduler μ Apps that beat model-based heuristics by 5–25% in throughput in over-the-air srsRAN experiments, within a real-time RAN controller with sub-millisecond AI-in-the-loop latency.
- Showcased as a poster and live demo at ACM MobiSys and SIGCOMM.

PUBLICATIONS

PITA: Preference-Guided Inference-Time Alignment for LLM Post-Training

arXiv preprint

S. C. Bobbili, U. Dinesha, D. Narasimha, S. Shakkottai

DOPL: Direct Online Preference Learning for Restless Bandits with Preference Feedback

ICLR 2025

G. Xiong, U. Dinesha, D. Mukherjee, J. Li, S. Shakkottai

Risk-Averse Fine-Tuning of Large Language Models

NeurIPS 2024

S. Chaudhary, U. Dinesha, D. Kalathil, S. Shakkottai

A Multi-Agent View of Wireless Video Streaming with Delayed Client-Feedback

IEEE INFOCOM 2024

N. Khan, U. Dinesha, S. Arunachalam, D. Narasimha, V. Subramanian, S. Shakkottai

EdgeRIC: Empowering Real-Time Intelligent Optimization and Control in NextG Networks

USENIX NSDI 2024

W.H. Ko, U. Ghosh, U. Dinesha, R. Wu, S. Shakkottai, D. Bharadia

PATENTS

Variable-Length CSI Feedback via Encoder Compression and Output Masks

WO2025014678A1

A. Malhotra, Y. Deenoo, N. Thangaraj, M.S. Ibrahim, S. Hamidi-Rad, U. Dinesha

Intelligent Control for Cellular Radio Access Networks

US20240267794A1

S. Shakkottai, W.H. Ko, H. Dureppagari, U. Dinesha, S. Ganji, D. Bharadia, R. Wu, U. Ghosh

RL-Based Rate Control for End-to-End Neural Network Based Video Compression

WO2024064329A1

F. Racape, U. Dinesha, H. Choi, M. Ul Haq

Video Compression for Both Machine and Human Consumption (Hybrid Framework)

WO2024049627A1

F. Racape, H. Choi, M. Ul Haq, U. Dinesha

Temporal Attention-Based Neural Networks for Video Compression

WO2023122077A1

V. Mohan, F. Racape, U. Dinesha, H. Choi

SERVICE

Reviewer: NeurIPS 2025

Teaching: Graduate TA at Columbia University (NLP; Genomic Information Science & Technology)