

Aishik Konwer, PhD

Sunnyvale, CA 94087 | 631-747-1244 | akonwer@cs.stonybrook.edu | [aishikkonwer95.github.io](https://github.com/aishikkonwer95)

PROFESSIONAL SUMMARY

Multimodal AI & Alignment Research Scientist (PhD) specializing in vision-language foundation models (VLMs), vision-language-action (VLAs), post-training alignment (SFT, DPO, GRPO), and synthetic video generation. Proven track record in scaling controllable generative models, optimizing grounded spatial reasoning pipelines, and co-authoring peer-reviewed research (CVPR, ICCV, MICCAI). Engineered production-grade alignment and synthetic data curation systems driving +12% model alignment, +6% zero-shot anomaly detection accuracy, and 1.8x acceleration in dataset synthesis. Expert in PyTorch, PEFT/LoRA, RAG, and distributed model fine-tuning.

RESEARCH INTERESTS

Multimodal Foundation Models (VLM/VLA), World Foundation Models, Post-Training Alignment (SFT, DPO, GRPO, RLHF), Generative AI (Video Diffusion, Action Transformers), Spatial Reasoning, Embodied Agents, Video Anomaly Detection, Zero-Shot & Few-Shot Learning, Inference Optimization.

WORK EXPERIENCE

Full-Time Industry Roles

Accenture | Advanced AI Research Scientist May 2025 – Present
Center of Advanced AI, AI Refinery Team

- **Multimodal Generative Agents & Synthesis:** Fine-tuned Cosmos-based video generation models and integrated controllable video generation frameworks with structured prompting, improving domain-adapted synthesis fidelity by 5% and accelerating scenario-specific dataset generation by 1.8x across end-to-end perception, reasoning, and synthesis agents.
- **LLM Post-Training & Scalable Evaluation:** Engineered synthetic data curation pipelines for SFT, DPO, and GRPO-based optimization; applied PEFT/LoRA adaptation for instruction and dialogue fine-tuning while deploying automated LLM-as-a-judge evaluation frameworks to boost model alignment by 12% and speed up iteration benchmarking by 30%.
- **Grounded Perception & Spatial Reasoning:** Architected spatially grounded VLM pipelines combining Chain-of-Thought (CoT) reasoning and open-vocabulary detection, increasing zero-shot anomaly detection accuracy by 6% and boosting mean task accuracy for multi-view VLA models in robotic scenes by 9%.

Cognizant | Programmer Analyst Dec. 2017 – Jul. 2019
Data Warehouse Team

- **Data Infrastructure & Automation:** Automated enterprise ETL data migration pipelines using Shell scripting, SQL, and Informatica, accelerating data transfer efficiency across large-scale warehouse systems.

Research & Industry Internships

GE Healthcare | AI Scientist Intern May 2024 – Aug. 2024
Foundational AI Team

- Architected SAM DPO and prompt-tuning, achieving 84.3–89.7% semi-supervised Dice across CT, ultrasound, and X-ray.

SRI International | Deep Learning Research Intern May 2023 – Aug. 2023
Scene Understanding and Navigation Team

- **Multimodal Remote Sensing:** Designed masked pre-training architectures for hyperspectral (HS), LiDAR, and RGB remote sensing data fusion, boosting multi-modal land-cover segmentation performance by 4%.

- **Annotation-Efficient Pathology AI:** Architected meta-learning cell detection and segmentation algorithms for gigapixel histopathology images under limited exemplars, raising segmentation accuracy by 20%.

EDUCATION

Stony Brook University, New York, USA 2019 – 2025

PhD, Computer Science

Advisor: Dr. Prateek Prasanna | Dissertation: Data-Efficient Multimodal Representation Learning

Institute of Engineering & Management, Kolkata, India 2013 – 2017

BTech, Electronics & Communication Engineering

PUBLICATIONS

-
- **A Konwer**, BP Tandra, AP Shah, “Enhancing Multi-View Vision-Language Models with Scene Graphs and Geometric Priors for Spatial Reasoning”, submitted to **WACV** 2027.
 - E Lai, Y Wang, **A Konwer**, “Concurtree: A Novel Memory Structure For Long-Video Streaming”, submitted to **ICLR** 2027.
 - S Agarwal, **A Konwer**, AP Shah, “Reasoning-Guided Grounding: Elevating Video Anomaly Detection through Multimodal Large Language Models”, submitted to **AAAI** 2027.
 - **A Konwer** et al., “Gaze2Report: Radiology Report Generation via Visual-Gaze Prompt Tuning of LLMs”, **ISBI** 2026. [Oral]
 - **A Konwer**, “Physical AI: The Next Frontier in AI and Robotics to Build Truly Autonomous Machines”, **Preprint** 2026.
 - **A Konwer** et al., “Enhancing SAM with efficient prompting and preference optimization for semi-supervised medical image segmentation”, **CVPR** 2025. [Oral]
 - **A Konwer** et al., “MetaStain: Stain-generalizable Meta-learning for Cell Segmentation and Classification with Limited Exemplars”, **MICCAI** 2024.
 - **A Konwer** et al., “Enhancing Modality-Agnostic Representations via Meta-Learning for Brain Tumor Segmentation”, **ICCV** 2023.
 - **A Konwer** et al., “MagNET: Modality-Agnostic Network for Brain Tumor Segmentation and Characterization with Missing Modalities”, **MICCAI Workshops** 2023.
 - **A Konwer** et al., “Temporal Context Matters: Enhancing Single Image Prediction with Disease Progression Representations”, **CVPR** 2022. [Oral]
 - **A Konwer** et al., “Lesion segmentation and genomic characterization of brain cancer patients from incomplete MR sequences”, Radiological Society of North America (**RSNA**) 2022. [Oral]
 - **A Konwer** et al., “Clinical outcome prediction in COVID-19 using self-supervised vision transformer representations”, **SPIE Medical Imaging** 2022. [Oral]
 - **A Konwer** et al., “Attention-Based Multi-scale Gated Recurrent Encoder with Novel Correlation Loss for COVID-19 Progression Predictions”, **MICCAI** 2021. [Early Accept]
 - **A Konwer** et al., “Predicting COVID-19 Lung Infiltrate Progression on Chest Radiographs Using Spatio-temporal LSTM based Encoder-Decoder Network”, **MIDL** 2021.
 - **A Konwer** et al., “Staff line Removal using Generative Adversarial Networks”, **ICPR** 2018. [Oral]
 - AK Bhunia*, **A Konwer***, A Bhowmick, AK Bhunia, PP Roy, U Pal, “Script Identification in Natural Scene Image and Video Frame using Attention based Convolutional-LSTM Network”, **Pattern Recognition** 2019.
 - S Nag, AK Bhunia, **A Konwer**, AK Bhunia, PP Roy, “Facial Micro-expression Spotting and Recognition Using Time Contrasted Feature with Visual Memory”, **ICASSP** 2019.

- AK Bhunia, AK Bhunia, P Banerjee, **A Konwer**, A Bhowmick, PP Roy, U Pal, “[Word Level Font-to-Font Image Translation using Convolutional Recurrent Generative Adversarial Networks](#)”, **ICPR** 2018.

SKILLS

Core Methods: Vision-Language Models (VLM), LLM Alignment, SFT, DPO, GRPO, RLHF, Embodied AI (VLA), Spatial Reasoning, Video Generation, Diffusion Models, PEFT, LoRA.

Languages & Frameworks: Python, PyTorch, C++, CUDA, vLLM, DeepSpeed, Hugging Face, OpenCV, TensorFlow, SQL.

Infrastructure & Tools: Distributed Training, Slurm, Docker, Kubernetes, AWS, Azure, Git, Linux.

INVITED TALKS & PROFESSIONAL SERVICE

Invited Speaker: NYU (2024), Stanford MedAI (2023), Roche Analytics (2022), Stony Brook (2022), Siemens Healthineers (2022).

Conference Reviewer: CVPR, ICCV, ECCV, NeurIPS, MICCAI, BMVC, MIDL, ISBI.

HONORS & CERTIFICATIONS

- Claude Architect Certification (2026)
- PhD Chairman Fellowship, Stony Brook University (2019)
- CVPR & SPIE Travel Grants (2022)
- SUNY Professional Development Award (2021)