

EDUCATION

King Abdullah University of Science and Technology (KAUST)

Saudi Arabia

PhD Student (Currently Enrolled)

Jun 2025 - Present

King Abdullah University of Science and Technology (KAUST)

Saudi Arabia

MS, Computer Science (with Thesis)

Sep 2023 – Jun 2025

Research focus: Diffusion-based video generation and editing.

- **MatchDiffusion:** Co-authored MatchDiffusion (ICCV 2025), combining joint and disjoint denoising to create visually coherent transitions between distinct scenes.
- **MoCA-Video:** First-authored MoCA-Video (arXiv preprint), using concept-localized latent injection and momentum-based correction for consistent video editing.
- **Coursework:** Machine learning, deep learning for visual computing, deep generative modeling, math foundations of ML, statistics, algorithms, and cybersecurity; additional graduate coursework in private data analysis and computational scientific imaging.

Southern University of Science and Technology (SUSTech)

China

Bachelor of Engineering, Computer Science

Aug 2019 – Aug 2023

- **Final-year project:** Developed rubric-guided LLM report grading; compared zero-shot and few-shot prompting against human scores.

WORK
EXPERIENCE**Qualcomm Mentorship Programme***Qualcomm AI Research (Remote)*

Jun 2025 – present

- Conducting generative AI safety research through a year-long mentorship collaboration.
 - Published the **DiSCO** preprint (under review): a training-free, model- and architecture-agnostic, plug-and-play safeguard using distribution-guided contrastive prompt optimization. Reduced average attack success rate from 23.6% to 2.4% (NudeNet), validated through repeated trials. Demonstrated consistent additional protection on top of existing defenses against state-of-the-art attacks across the evaluated attack–defense pairings, while maintaining or improving generation quality.
 - **Ongoing trajectory research:** Investigating diffusion-based red teaming and defense through joint and disjoint trajectory analysis, using early denoising signals to detect malicious intent, adversarial editing requests, and concealed malicious patches that evade input filters, aiming to enable early safety intervention.

CAMEL-AI – Open-Source Multi-Agent Framework*Software Developer (Remote)*

Jun 2023 – Dec 2023

- Co-developed and deployed task-driven LLM multi-agent workflows for downstream applications, integrating task planning, expert role assignment, role-playing dialogue, and shared environment context to coordinate multi-step problem solving.
- Designed task decomposition and expert coordination workflows: represented subtasks and dependencies in JSON, used topological ordering and Gantt-based plans to schedule execution and pass prerequisite outputs, and paired complementary AI user and assistant roles through compatibility-based assignment and role-specific prompts to guide task-focused expert dialogue.
- Optimized memory and context handling for multi-turn dialogue through structured insight extraction, including topic segmentation, entity recognition, and task-relevant fact extraction. Integrated prior subtask results and dynamic environment updates into subsequent agent conversations to preserve dependencies and relevant context within limited token budgets.

ADDITIONAL RESEARCH PROJECTS

- Developed **CAPTAIN**, a training-free inference-time framework to reduce data-privacy risks from memorized training images in text-to-image diffusion models. Combined frequency-aware initialization with spatially and temporally localized semantic feature injection during denoising. Reduced copy-detection similarity (SSCD) from 0.40 to 0.25 versus BE-Global on Stable Diffusion v1.4 while preserving prompt fidelity; preprint available, manuscript under review.
- Developed **CoLaP**, a multimodal prompt-learning framework for continual adaptation of frozen vision backbones. Combined semantic clustering of VLM-generated captions, contrastive alignment, and teacher–student distillation to improve prompt selection under distribution shifts while retaining image-only inference. Improved ImageNet-O accuracy from 63.24% to 68.50% over L2P, with broader evaluation across five out-of-domain benchmarks; manuscript under review.

ACADEMIC SERVICE AND TEACHING

- Reviewer, BeNTO Workshop, NeurIPS 2026
- Teaching Assistant, Deep Learning in Visual Computing (CSC323), Spring 2026
- Workshop Organizer, CVPR Workshop on Computer Vision for the Environment (CVEU), 2025

TECHNICAL SKILLS

- **Research Areas:** Generative AI safety, text-to-image diffusion, video generation and editing, multimodal learning, continual learning, Artificial Intelligence, Machine Learning, Literature Review, Data Analysis
- **Languages:** Python (primary), C/C++, SQL, R, Bash, LaTeX
- **ML/AI:** PyTorch, Hugging Face, Diffusers, LoRA, GRPO, distributed training, prompt optimization
- **Tools:** Git, Docker, Linux, Weights & Biases, pandas, NumPy, scikit-learn