

SAKET ATREYA

inhuman.systems · saket.atreya@students.iiit.ac.in · github.com/saketatreya
Hyderabad, India · +91 7032339992 · US work authorized; no sponsorship required

EDUCATION

International Institute of Information Technology, Hyderabad

B.Tech in Computer Science and Engineering

2023 – May 2027

Expected graduation: May 2027

RESEARCH EXPERIENCE

Lossfunk

Jun 2026 – Present

Research Intern

Bangalore, India

- Developed action-conditioned world models on frozen vision encoders (DINOv3, V-JEPA 2), computing camera motion analytically and learning residual scene dynamics.
- Improved held-out prediction skill from 0.393 to 0.557 over a matched 3M-parameter baseline on 100 ARKitScenes visits (two seeds); analytic transport alone reached 0.482.
- Compared analytic transport, learned baselines and residual models with pre-registered hypotheses, held-out splits and visit-level bootstrap confidence intervals.

SELECTED RESEARCH

Navigation world-model rollout evaluation

Current manuscript: Separating Command Error, Chaining Cost and One-Call Error

- Audited a released family of action-conditioned diffusion world models (49M–1.01B parameters), separating command-composition error, chaining cost and direct prediction error.
- Correcting rollout commands reduced the 1B model's 4 FPS LPIPS by 18.8% (95% CI: 16.9–20.6%) on a prospectively specified confirmation cohort, without retraining.

Readout is not generation: SE(3) in the patch tokens of frozen encoders

NeurReps Workshop, NeurIPS 2026

- A parameter-free rotation warp explained 16–37% of held-out feature change; a fixed-depth plane warp beat a matched learned flow head in all 15 evaluated comparisons.

Spectral Perturbation Bounds for Experience Replay

ICML 2026 Workshop on Decision-Making from Offline Datasets to Online Adaptation

- Derived a bias–variance decomposition for offline decision-making that relates the value of data diversity to the environment's mixing time.

Reading Relatedness and Hierarchy from Frozen Representations

NeurReps Workshop, NeurIPS 2026; revised manuscript

- Distinguished local relatedness, quartet agreement and exact hierarchical recovery across five modalities; tested how readout origin and normalization change apparent structure.

What Transfers through Synthetic Code? Separating Security, Style, and Provenance

Under review, AAAI 2027 (AI Alignment track)

- Investigated generated-code transfer using behavioral evaluations, activation probes and artifact audits; separated security measurements from output-budget and model-lineage confounds.

PROJECTS

Pokémon Showdown RL Agent (write-up)

PyTorch, Stable-Baselines3

- Built a recurrent PPO agent with self-play and explicit opponent-role beliefs and damage-aware state features in partially observed battles.

SKILLS

Python, C, PyTorch, Hugging Face Transformers, vLLM, Stable-Baselines3, Gymnasium, MuJoCo