# Suraj Ranganath

suranganath@ucsd.edu | +1(858) 214-8608 | Linkedin | Scholar | Github | Medium | San Diego, CA

## Education

**University of California San Diego** **San Diego, CA**
*M.S. in Data Science* (GPA: 4.0/4.0) *Sept 2025 – Jan 2027*
Courses: ML Systems, Efficient AI, Safety in Generative AI, Scalable Data Systems, Web Mining and Recommender Systems

**Vellore Institute of Technology** **Vellore, India**
*B.Tech. in Computer Science and Engineering* (GPA: 3.98/4.0, *Top 5 in class*) *Aug 2020 – May 2024*
Courses: Machine Learning, Data Structures and Algorithms, Operating Systems, Database Management, Cyber Security

## Experience

**Causal Intelligence Lab, UC San Diego** **Dec 2025 – Present**
*Research Assistant* *San Diego, CA*
- Working with Prof. Huang on extending memory for video diffusion models toward embodied world modeling.
- Training Vision Language Models (VLMs) to reason about the action space to improve performance on long-horizon tasks.

**BalkanID** **May 2023 – Jun 2025**
*Machine Learning Engineer* *Remote*
- Built BalkanID Copilot, an agentic identity security system leveraging LLMs and knowledge graphs, adopted by paying customers & earning over **$200K in revenue** by surfacing actionable risk insights and cutting compliance reporting time, in a 3 member team.
- Designed and productionized a text-to-Cypher stack by LORA fine-tuning and building RAG over a multi-tenant Identity Intelligence Graph (5M+ entities/relations), enabling natural-language querying of IAM policies and entitlements.
- Architected and deployed an MCP server with connector integrations, enabling natural language commands to trigger actions and workflows across the platform and 20+ enterprise services.
- Built multi-tenant ETL pipelines ingesting ∼50GB/tenant from 100+ enterprise sources across 20+ identity & risk data types; normalized, deduped, and entity-resolved feeds into our identity graph to power analytics, compliance, and ML features.
- Conceived and engineered the BalkanID IAM Risk Analyzer, applying graph-based role mining and just-in-time access provisioning to enforce least privilege, reduce identity risk exposure, and accelerate access approval turnaround time by 3X.
- Owned model reliability for Copilot: built evaluation pipeline and production monitoring for drift and behavior regressions, enabling faster iteration on new LLM and retrieval variants through safe rollouts backed by offline evals and live telemetry.
- Co-inventor on **4 U.S. patents** in agentic AI and graph machine learning for identity security - delivering actionable risk insights, cutting manual reviews, and enabling autonomous governance workflows.

**Indian Institute of Technology, Bombay** **Jan 2023 – Jun 2023**
*Research Intern* *Mumbai, India*
- Collaborated with Prof. Mitra, EdTech Lab on multimodal data analysis for spatial problem-solving research, in 6-person team.
- Trained MLP classifier to identify data-backed signatures of 5 spatial thinking strategies, informed by prior qualitative studies.
- Designed and managed multimodal data acquisition (EEG, eye tracking, GSR, affective video analysis), producing a novel dataset of 300+ hours, published in Nature Scientific Data.

## Technical Skills

**Languages**: Python, Cypher, SQL, Java, Go, Javascript, Typescript, React, NoSQL, Matlab, R, C, C++, Bash, Shell, Linux
**Machine Learning**: PyTorch, JAX, Agentic Frameworks, RAG, MCP, LLM Post-Training, LLM Monitoring and Evaluation
**Software and Frameworks**: WandB, FastAPI, Django, Pandas, Matplotlib, NLTK, PySpark, PyTorch Geometric
**Tools and Platforms**: Git, Docker, AWS, Traefik, ArgoCD, GraphQL, Selenium, Tinker, HuggingFace
**Databases**: Neo4J, MySQL, FAISS, Qdrant, Pinecone, PostgreSQL, MongoDB, Redis, Firebase, SQLite3

## Patents and Publications

- **Knowledge Graph-Enhanced AI Copilot Platform for Intelligent Identity Security Governance and Lifecycle Management. (2025).** Ranganath, S., Raghavendra, A., *et al.* U.S. Patent 19/055,635, Published Nov 6, 2025. Patent pending.
- **A Multisensor Dataset of South Asian Post-Graduate Students Working on Mental Rotation Tasks. (2025).** TS, A., Ranganath, S., *et al.* Nature Scientific Data, 12(1), 563.
- **Recommendation and Remediation of Role-Based Access Control Postures for Identities. (2025).** Ranganath, S., Grewal, D., *et al.* U.S. Patent 19/195,207, filed April 30, 2025. Patent pending.
- **Multi-Agent Identity Governance and Administration System. (2025).** Raghavendra, A., Ranganath, S., *et al.* U.S. Patent 19/194,684, Published Nov 6, 2025. Patent pending.
- **Method for Provisioning and De-Provisioning Just-in-Time, Purpose-Based Access for Identities within Applications. (2025).** Wong, N., Deep, A., *et al.* U.S. Patent 19/194,524, Published Nov 6, 2025. Patent pending.

## Projects

**StealthRL: GRPO Post-Training for AI-Text Detector Red-Teaming** *Github* | *Paper* **Dec 2025**
- Built an RL post-training pipeline to learn a paraphrasing policy that generalizes across multiple AI-text detector families while preserving semantic meaning and fluency.
- Implemented GRPO with LoRA adapters on Qwen3-4B-Instruct and a composite reward (detector-ensemble score, semantic similarity, perplexity); ran training and ablation studies on Tinker.

**KV Cache Quantization for Self-Forcing Autoregressive Video Generation** *Github* | *Paper* **Mar 2026**
- Auto-research style system to discover KV-cache quantization strategies for self-forcing autoregressive video generation to enable better consistency in long-horizon generation.
- Comprehensive empirical evaluation: 33 quantization/cache-policy variants across MovieGen and StoryEval.
- Key finding: FlowCache-inspired soft-prune INT4 adaptation reaches 5.42–5.49x compression, reducing peak VRAM from 19.28 GB to 11.7 GB with modest runtime overhead, while staying visually close to the BF16 baseline based on SSIM, and PSNR fidelity checks.