

# NIDHUSHAN KANAGARAJA

nidhushank6@gmail.com | +1 (503) 421-2971 | **Open to Relocation**  
Portfolio | LinkedIn | LeetCode | GitHub

## AI/ML RESEARCH

**Time-Aware Episodic Memory for Retrieval Agents** Sep 2025 – Present  
*Independent Research – Long-term Memory Systems for Personal AI*

- Designed time-aware deep retrieval system enabling year-scale conversational recall under bounded LLM context.
- Proposed a Hybrid semantic–lexical–temporal retriever (cite P@3: 0.28–0.33 on single-gold temporal queries).
- Introduced episodic consolidation with manifests to support provenance-aware and auditable memory retrieval.
- Episode hit@3 of 0.88–0.98 on curated temporal queries, for coarse-grained recall with auditable episode boundaries.
- Enabled natural-language temporal queries (e.g., “what did I try last week?”) over long interaction histories.

**Text-Conditioned Procedural Content Generation (Published AAAI 2025)** Jan 2024 – Apr 2025

- Developed a supervised multimodal PCG pipeline using 70K map–description pairs.
- Achieved 85% semantic alignment by training the Five-Dollar Model and discrete Diffusion Model as T2M baselines.
- Created dataset splits with 49K training, 14K test, and 7K validation examples for controlled generative evaluation.
- Evaluated models using BLEU (54.7), ROUGE-L (33.2), METEOR (19.5), SPICE (11.3), and CLIP similarity.

## PROFESSIONAL EXPERIENCE

**Giinius | New York, NY**

*Lead AI/ML Engineer — Serendipity Salon* May 2025 – Present

- Built and deployed an AI-powered reflection and dialogue platform used in live community sessions.
- Architected LLM workflows for guided prompting, structured memory capture, and post-session synthesis.

*AI/ML Engineer — Giinius Educational Directory* May 2024 – May 2025

- Developed ML systems powering content discovery, ranking, and metadata enrichment.
- Integrated DPR with NVIDIA NeMo for metadata retrieval; improved ranking by 18%.
- Reduced inference time from 1.2s to 850ms using pruning, distillation, and quantization.
- Improved LLM response quality via retrieval-aware LoRA adapter routing.

**Software Engineer | TCS - Citibank Project | Chennai, India** Jul 2022 – Jun 2023

- Engineered Python/Shell frameworks for SFTP automation, housekeeping, and outbound pipelines.
- Led testing and delivery of modules supporting compliance-ready financial transactions.

**Machine Learning Engineer Trainee | HighRadius | Chennai, India** Jun 2021 – May 2022

- Developed payment delay prediction model (92% accuracy) and credit risk scoring chatbot.

## PROJECTS

**Mnemo: AI-Powered Diary & Voice Journal System** Nov 2025 – Present

- Designed an LLM-driven diary system converting free-form notes into structured logs using on-device 3B/8B models.
- Constructed multimodal pipeline for audio transcripts and text synthesis using faster-whisper STT and Piper TTS.
- Deployed the system on an Oracle Cloud VM using FastAPI, Nginx, HTTPS, and Docker for low-latency operation.

**SkillBarter: C2C Peer Service Marketplace | NYU Capstone** May 2024 – Dec 2024

- Led a team of 6 to design and develop a peer-to-peer service exchange platform using time-credit economy.

## EDUCATION

**New York University** Sep 2023 – May 2025

M.Sc. in Computer Science – *Merit-Based Scholarship, Co-author: AAAI 2025 Paper* GPA: 3.75/4.0

**SRM University** Jun 2018 – May 2022

B.Tech in Computer Science – *First Class with Distinction* GPA: 3.68/4.0

## SKILLS

**AI & ML Systems:** Supervised/unsupervised learning, deep learning, transformers, multimodal modeling (text/audio), retrieval systems, agentic workflows, model fine-tuning, evaluation and optimization

**ML Frameworks & Tooling:** PyTorch, TensorFlow, numpy, Transformers, LangChain, ONNX

**Production & MLOps:** Python, Docker, Linux, CI/CD, FastAPI, REST APIs