Abhijeet Sahdev

Newark, NJ, USA

☎ +1-862-800-8204   ✉ as4673@njit.edu   💼 abhijeet-sahdev   ⚙ jeet1912   🌐 jeet1912.github.io

EDUCATION

New Jersey Institute Of Technology   Jan 2024 – Dec 2025
*M.S. in Artificial Intelligence, **CGPA: 3.86/4***   *Newark, USA*

Manipal Institute Of Technology   Jun 2017 – Aug 2021
*B.Tech. in Computer Science and Engineering, **CGPA: 3.34/4***   *Udupi, India*
Minor in Computational Mathematics

RESEARCH INTERESTS

Geometric deep learning, graph neural networks, hyperbolic representation learning, diffusion and flow models, clinical machine learning, causal inference, model interpretability, spatial intelligence and 3D human–AI interfaces, retrieval-augmented generative modeling, and high-performance ML systems (CUDA/MPI).

RESEARCH EXPERIENCE

Xu Lab ↗   Sept 2025 – Present
Research Assistant   *Newark, USA*

- Developing a synthetic ICD trajectory framework using rooted-tree ontologies to study geometry–accuracy tradeoffs between Euclidean and hyperbolic embeddings, DDPMs, and rectified-flow generators.
- Proposed a curvature-aware generative risk modeling framework integrating Poincaré-ball ICD embeddings, graph-diffusion encoders, and rectified-flow transport to overcome Euclidean DDPM limitations.
- Reconstructed the MedDiffusion heart-failure cohort with corrected temporal windows and demographic matching, producing the first reproducible benchmark for geometry-aware EHR generation.
- Trained hyperbolic ICD embeddings with HDD-supervised alignment (**Corr = 0.84–0.91**), substantially improving hierarchical fidelity over Euclidean baselines.
- Designed a graph-hyperbolic encoder combining tangent-space attention, diffusion kernels, Einstein pooling, and temporal embeddings to model hierarchical comorbidities and irregular visit intervals.
- Achieved state-of-the-art performance on MIMIC-III (**AUPRC 0.79–0.83 vs. 0.7064** MedDiffusion; **Cohen's $\kappa \approx 0.50$ vs 0.45**), supported by extensive ablations quantifying the roles of curvature, encoder depth, diffusion scope, and decoder expressivity.
- Used scaling laws from synthetic and clinical experiments to guide architectural choices and training strategies for HyperMedDiff-Risk [*Code*], now being prepared as a manuscript and extended technical report.

Immersive Creation Lab ↗   May 2025 – Present
Research Assistant   *Newark, USA*

- Built a browser-native 3D copilot integrating React + react-three-fiber editing, Sketchfab retrieval, and an LLM-driven staging/commit workflow.
- Developed a GLB optimization pipeline (gltf-transform, Draco, texture compression) to accelerate model streaming and caching, reducing 3D asset loading time by **13%**.
- Engineered an Express backend enforcing JSON tool calls, structured error handling, telemetry logging, and traffic rate-limiting.
- Authored a 1.6K-line prompting specification encoding spatial reasoning (collision checks, quadrant heuristics, collective-noun expansions, synonym/fuzzy matching, safe animation code).
- Prototyping GNN-based scene reasoning modules to capture relational and geometric structure for controllable generation.
- Supporting a production-grade UIST submission and longitudinal user studies on controllable 3D interfaces.

SELECTED PROJECTS

Distributed GPT-2   Sept 2025 – Dec 2025
C/C++, MPI, CUDA [Report][Code]

- Implemented a minimal GPT-2 in C++ with custom autograd, CUDA kernels, tokenizer/dataloader, and Adam optimizer to study transformer internals.
- Extended to distributed training with MPI data parallelism, gradient allreduce, and rank-aware batching across multi-GPU nodes.
- Achieved **67×** training and **14×** inference speedups (226s vs. 15,264s CPU baseline) while maintaining model quality.

GIT-base Ablation Study   Jul 2025 – Aug 2025
PyTorch [Report] [Code]

- Curated a 66K+ MIMIC-CXR-JPG dataset using BigQuery + CheXpert with balanced splits and structured mini-reports.
- Benchmarked GIT-base under full fine-tuning vs. LoRA variants, designing memory-efficient training that reduced GPU usage by **50%** and runtime by **67%**.
- Observed **5.8× BLEU** improvement with LoRA text-only adaptation, demonstrating strong domain transfer for medical VLMs.

Sepsis Treatment With Deep RL   **Jun 2025 – Aug 2025**
PyTorch [Report] [Code]

- Implemented a reinforcement learning framework (WD3QNE) for ICU sepsis treatment, replicating published results across 8 experiments with 45 vs. 43 feature sets obtained after sequential backward selection.
- Processed and curated MIMIC-III data (14,505 patients) using hierarchical time-aware imputation and Sepsis-3 cohort selection.
- Trained agents over 100 epochs, achieving structured 5×5 fluid/vasopressor policies aligned with clinician behavior and robust survival outcomes.

Open Source Contribution   **Feb 2021 – Mar 2021**
React Native [Pull Request]

- Contributed a feature addition to `react-native-multiple-select` (550+ stars, 500k+ downloads), enabling display of selected-item counts in production apps.

MANUSCRIPTS AND PREPRINTS

HyperMedDiff-Risk: Hyperbolic Graph Diffusion for Clinical Risk Prediction   **Sept 2025 – In Preparation**
First author with Prof. Mengjia Xu (NJIT) [Manuscript]

- Introduces a geometric generative modeling framework that unifies hyperbolic code embeddings, graph-diffusion signatures, and rectified-flow trajectory modeling for clinical risk prediction.
- Demonstrates that curvature-aware generative structure significantly improves discriminative performance and produces synthetic EHR trajectories with measurable hierarchical preservation.

Sahdev, A. *Statistical Analysis of Sentence Structures through ASCII, Lexical Alignment, and PCA*   **Sept 2024 – March 2025**
[*arXiv:2503.10470*]

- Introduces a statistical framework for assessing sentence-structure balance using ASCII encodings, PCA-based compression, and lexical alignment across diverse text corpora.
- Applies Shapiro–Wilk and Anderson–Darling normality tests to compare human-written and LLM-generated text, highlighting structural differences in lexical usage.

INDUSTRY EXPERIENCE

MyHealthToday ↗   **Dec 2020 – Jun 2021**
Software Engineering Intern [Report]   *Remote*

- Developed cross-platform interfaces in React/React Native and implemented design-compliant test suites for multi-device workflows.
- Standardized DynamoDB schemas and reduced backend processing overhead using typed Redux state management.
- Built and deployed AWS Lambda microservices and RESTful APIs via SAM CLI, reducing client-side computation by **16%**.

TEACHING

New Jersey Institute Of Technology   **Sept 2024 – Dec 2024**
Teaching Assistant — Data Mining & Machine Learning   *Newark, USA*

- Responsible for grading various phases of capstone project and setting up evaluation metrics for assignments of CS Data Mining.
- Invigilated mid-term exams for DS Machine Learning course for 70+ students.

SKILLS

Programming: Python, C/C++, Java, JavaScript, SQL
ML / DL: PyTorch, Geoopt, scikit-learn
High-Performance Computing: CUDA, OpenMPI

Cloud & Infrastructure: AWS, GCP, Docker, Git
Frameworks: React, Three.js (Node.js/Express.js for tooling)
Databases: MySQL, DynamoDB