NISHANTH BASAVA

Nashville, TN · 423-987-6206 · nishanthrbasava@gmail.com · github.com/nishanthbasava · linkedin.com/in/nishanth-basava

EDUCATION

Vanderbilt University *Expected May 2028*
B.S. Computer Science; B.A. Mathematics; Minor in Cognitive Studies GPA: 3.81/4.00 (CS: 3.93)
U.S. Presidential Scholar · Chancellor's Scholar (full-tuition, top 1%) · Dean's List

Relevant Coursework: Data Structures, Algorithms, Database Management Systems, Multivariable Calculus & Linear Algebra, Intermediate Software Design, Probability & Statistics, Computer Architectures, Machine Learning & NLP in Healthcare
Organizations: Software Engineer at ChangePlusPlus, VP of Design at VandyHacks, Kappa Theta Pi (Technology Fraternity)

EXPERIENCE

Software Engineering Intern, Frizzle (YC S25) *Sep 2026 – Present*

- Expanded subject coverage in a **LangGraph** worksheet-grading pipeline served via Inngest, splitting history and world-language profiles out of a catch-all "other" category, injecting subject-specific prompting into each grading module and raising agreement with human labels on those worksheets from 83% to 91% (19/23 to 21/23 worksheets)
- Built a **LangSmith** eval loop on 70 labeled worksheets, used it with teacher feedback to iterate on grading prompts, and prototyped confidence-gated escalation to a stronger model for low-confidence grades

Machine Learning Research Intern, Neuroimaging & Brain Dynamics Lab, Vanderbilt *May 2026 – Present*

- Implemented a 458-scan fMRI pipeline using fMRIPrep, denoising, and ROI connectivity on SLURM; found a session-pooling bug that flipped 18/458 QC labels and a $\sim$20$\times$ scanner-vendor bias in outlier flags that would have skewed subject selection
- Built **AutoQA**, an automated QC tool for fMRI scans that triages each scan as include, caution, or exclude, so researchers open only flagged scans instead of reviewing every one
- Trained AutoQA's reviewer by curating an expert-labeled set (293 scans; $\kappa = 0.92$) and distilling a Claude teacher into Qwen2.5-VL-7B with LoRA, raising held-out macro-F1 from 79.6% to 84.1% at 82% lower inference cost

Data Science Intern, Harvard T.H. Chan School of Public Health / Dana-Farber *Sep 2023 – Aug 2025*

- Co-developed **bayesNMF**, an MH-within-Gibbs sampler for Bayesian Poisson NMF that avoids Poisson augmentation (3–30$\times$ faster than Gibbs) and learns rank automatically; applied to 32 PCAWG cancer types; led a 328-exome glioblastoma analysis
- Built and deployed **SigServe**, extending bayesNMF into a full-stack app (React, FastAPI, PostgreSQL, Redis/RQ, Docker, AWS EC2) running the sampler asynchronously via rpy2, with CSV upload, a live job queue, and shareable result pages
- Analyzed 12,700+ SigServe load-test jobs with SQL (CTEs, window functions); 58 tests at 91% coverage in CI (sig-serve.org)

PROJECTS

DiffEx | Explainable probabilistic differential diagnosis *TypeScript, PostgreSQL, Vercel*

- Built an explainable diagnosis engine combining demographic-adjusted priors with a likelihood-ratio knowledge graph (86 conditions, 455 associations), returning ranked ICD-10 diagnoses with per-finding contributions in $\sim$5 ms (diffex.app)
- Evaluated on 23 vignettes (BMJ 2015): 61%/78% top-1/top-3 accuracy vs. 22% top-3 for a common-diagnosis baseline
- Piloted with two VUMC and six Guatemala ER physicians across 250+ encounters; used missed diagnoses to expand the knowledge graph from 72 to 86 conditions; 1st Place, Vanderbilt Hult Prize ($3,000); 2026 USA East Finalist

Perigeon | AI copilot for satellite collision avoidance *Python, FastAPI, ChromaDB, React*

- Built a RAG pipeline that turns predicted satellite collision risks into ranked response plans: embedded a 47-task operations catalog (all-MiniLM-L6-v2, ChromaDB), mapped predictor features to top-10 retrieval queries, and scored three plans in 8 ms
- Developed a retrieval eval harness over 40 rubric-labeled collision scenarios scoring recall@10, MRR, and nDCG@10; benchmarked two embedding models against a BM25 baseline and ablated query construction, raising recall@10 from 0.56 to 0.66
- 1st Place, Xtremis.AI track at 2026 Dry Dock Hackathon ($5,000); team project with a 3D orbital dashboard and live alerts

TECHNICAL SKILLS

Languages: Python, SQL, TypeScript, JavaScript, C++, Java, R, C
ML: PyTorch, Hugging Face (Transformers, PEFT), LoRA fine-tuning, distillation, LLM evaluation, CNNs, YOLO
Frameworks & Tools: FastAPI, Node.js, React, React Query, LangGraph, Inngest, ChromaDB
Data/Infra: Pandas, NumPy, PostgreSQL, MySQL, Redis, Docker, AWS, GitHub Actions, SLURM/HPC, Modal

PUBLICATIONS

J. M. Landy, **N. Basava**, G. Parmigiani. bayesNMF: Fast Bayesian Poisson NMF with Automatically Learned Rank Applied to Mutational Signatures. *Journal of Computational and Graphical Statistics*, Vol. 35 (2026). Google Scholar