Azariah Jebin

AI / ML Engineer · Computer Vision · Graph Neural Networks · Clinical AI
azariahjebin@gmail.com | +918300183081 | Tirunelveli, Tamil Nadu, India | LinkedIn | Github

SUMMARY

AI/ML Engineer with an **M.S. in Artificial Intelligence from UT Austin (GPA 3.7)**. Author of NeuroGraph, a spatially informed GNN framework predicting pediatric brain tumor molecular biomarkers (GFAP, H3K27M, INI1, ALK1, Synaptophysin) directly from H&E whole slide images, achieving GFAP AUC 0.939 and mean AUC 0.761 across 5 biomarkers on 182 real-world clinical cases (IRB: STUDY00007710), in collaboration with Dell Medical School. Background spans 3 years of production data engineering at Capgemini (concurrent with MSAI) and active freelance ML work. Committed to building AI that functions reliably under messy, resource-constrained, real-world clinical conditions.

SKILLS

Languages: Python, SQL
AI/ML & Deep Learning: PyTorch, PyTorch Geometric, GNNs, EdgeConv (DGCNN), CNNs, U-Net, EfficientNet, YOLOv5, Transformers, Multi-task & Multi-label Learning, Scikit-learn
Computer Vision: WSI Processing, Digital Pathology, Object Detection, Semantic Segmentation, Monocular Depth Estimation, Feature Extraction (UNI2-h foundation model)
Generative AI / LLMs: LangChain, LiteLLM, OpenAI APIs, Prompt Engineering, Google ADK
Data Engineering & MLOps: GCP BigQuery, Apache Airflow, Neo4j, Ab Initio, ETL Pipelines, TACC HPC, NVIDIA A100, Streamlit

EXPERIENCE

Computer Vision Engineer (Freelance) | Independent | Remote ***Jun 2026 – Present***

- Building a multimodal tomato disease diagnosis system for a postdoctoral researcher. Fusing macroscopic images, microscopic images, and clinical metadata across 5 progressive PyTorch architectures (ConvNeXt-Tiny achieving best macro-baseline performance).
- Writing all pipeline code from scratch: dataset curation, custom transforms, training loops, evaluation, to build production PyTorch proficiency without LLM assistance.

Graduate Researcher | The University of Texas at Austin, USA | Advisor: Dr. Ying Ding ***Jun 2025 – Present***

- Active lab member post-graduation, extending **NeuroGraph** into a multimodal framework fusing gigapixel WSIs, tumor coordinates, and patient demographics via a joint embedding layer.
- Designing a multi-task learning objective to simultaneously decode molecular biomarkers, genomic signatures, DNA methylation profiles, and histologic tumor type, unifying currently separate diagnostic workflows.
- Proposed a novel extension integrating a biomedical LLM branch with the GNN via cross-attention fusion, enabling structured clinical explanation generation from molecular predictions, directly addressing the interpretability gap in current computational pathology models.

Analyst | Capgemini | Bengaluru, Karnataka, India ***Feb 2023 – Dec 2025***

Completed MSAI at UT Austin simultaneously while working full-time.

- Built SemantIQ: an **LLM-powered NL-to-Cypher/SQL** metadata retrieval assistant using LangChain and Gemini, with a Streamlit frontend for Neo4j graph query visualization.
- Engineered enterprise-scale data migration pipelines from legacy **Teradata SQL to GCP BigQuery**, orchestrated via Apache **Airflow DAGs** in Cloud Composer, and large-scale Dev/UAT validation using the Pelican tool.
- Maintained production ETL pipelines (Ab Initio, SQL) under live SLA constraints, developing rigorous instincts for data quality and incident resolution.

PUBLICATIONS & PREPRINTS

NeuroGraph: Predicting Pediatric Brain Tumor Biomarkers from H&E WSIs Using GNNs | Master's Thesis, UT Austin
Advisors: Dr. Ying Ding, Dr. Leqi Liu · Clinical: Dr. Chandra Krishnan, Dell Medical School ***Dec 2025***

- Identified the core limitation of standard MIL approaches, patch-independence discards tissue spatial topology and designed a spatially-aware GNN pipeline to address it.
- Built a dense patch-level graph (400–6,000 nodes per slide) using Delaunay Triangulation (pruned at 4,000 μm threshold); each node encoded by UNI2-h foundation model (1,024-dim features), preserving physical tumor microenvironment at full resolution.
- Applied EdgeConv (Dynamic Graph CNN) layers to compute explicit feature differences between neighboring tissue patches, enabling the model to reason about tissue gradients and infiltrative edge patterns invisible to standard CNNs.
- Engineered a custom Masked Binary Cross-Entropy Loss to handle real-world label sparsity, GFAP available for only 54% of cases, H3K27M for 28% enabling simultaneous learning across partially annotated slides without discarding data.
- Results on 182-case pediatric cohort (5-fold CV, patient-level split): GFAP AUC 0.939 / F1 0.896 / Acc 85%;

Synaptophysin AUC 0.777; H3K27M AUC 0.699; ALK1 AUC 0.736; INI1 AUC 0.656 (specificity >89% under extreme class imbalance); Mean AUC 0.761, Mean Acc 86.6%.
- Integrated GNNExplainer to generate biomarker-specific tissue heatmaps; built a Clinical Interpretability Dashboard validated qualitatively by the collaborating neuropathologist, AT/RT rhabdoid morphology correctly highlighted for INI1.

NEXUS: A Multi-Agent Architectural Framework for Autonomous Enterprise Workflows | Preprint | Preprints.org | DOI
- Authored a system design position paper proposing a transition from human-default processing structures to decentralized, **AI-native agentic environments**.
- Architected a distributed coordination framework utilizing **Google Agent Development Kit (ADK)** and **LangChain** to orchestrate domain-specific reasoning nodes.
- Introduced "Truth Score Engine", a weighted, penalty-adjusted consensus metric ($Ts = [\Sigma wi \cdot Ci] \cdot \Pi(1-Hpenalty)$) for confidence-aware autonomous decision routing, utilizing **evidential deep learning** principles to quantify epistemic uncertainty across multi-agent consensus loops.

Mitigating Dataset Artifacts in NLI: A Sensitivity Analysis of Ensemble-Based Debiasing
- Analyzed empirical dataset shortcuts by training a hypothesis-only baseline using **ELECTRA-Small**, revealing a baseline accuracy exploitation rate of **60.12%** on the SNLI corpus.
- Engineered with an ensemble-based training pipeline utilizing **custom penalty loss terms** to structurally prevent the model from optimizing spurious lexical overlap artifacts.
- Conducted a comprehensive sensitivity analysis over the debiasing strength parameter (α), quantifying the mathematical trade-off between leaderboard accuracy and out-of-domain robust generalization.

Real-Time Video Analytics and Object Classification via Cascaded YOLOv5 and EfficientNet Architectures (*Technical Report, UT Austin)*
- Designed a multi-stage, cascading computer vision pipeline optimizing spatial bounding box localization (**YOLOv5**) coupled with a deep feature extractor classification backbone (**EfficientNet**).
- Structured custom end-to-end data pipelines involving geometric and photometric image augmentations to ensure model robustness under highly variable background noise.
- Evaluated inference performance bottlenecks, focusing on frame-per-second (**FPS**) metrics and memory footprint constraints for downstream edge-device compatibility.

PROJECTS

Autonomous Driving Planner – SuperTuxKart AI | *MSAI Project* | *UT Austin*
- **Engineered deep learning architectures** including **MLP, CNN, and Transformers** to predict vehicle trajectories within the SuperTuxKart simulation environment.
- **Metric-Driven Optimization:** Improved trajectory accuracy by **40%** through the implementation of custom loss functions and rigorous evaluation of lateral and longitudinal errors.
- **Simulation Deployment:** Validated model robustness in **PySuperTuxKart**, ensuring high-performance navigation under diverse track conditions and corner cases.

Multi-Task Road Scene Segmentation & Depth Estimation | *MSAI Project* | *UT Austin*
- **Developed a Residual U-Net** architecture for simultaneous road segmentation and monocular depth estimation in synthetic driving environments.
- **Performance Engineering:** Utilized **multi-task loss functions** and advanced data augmentation to boost real-time inference speeds for downstream planning tasks.

The Amazing Bible App *(Personal Project* | Google Play Store*)*
- Designed and built an **AI-native mobile application** with intelligent search, bookmarks, and prayer tracking.
- Utilized **LLMs as development accelerators**, applying structured **prompt engineering to design architecture, generate scaffolding code, and iterate on UI/UX.**

CERTIFICATIONS

- **Biomedical Researchers Certification** (CITI Program, Valid until 2028), **Computer Vision with OpenCV & Python** (Coursera)

EDUCATION

Master of Science in Artificial Intelligence, — *Jan 2024 – Dec 2025*
University of Texas at Austin, Texas, USA — Grade: 3.7/4.0
Bachelor of Engineering, Computer Science Engineering — *Aug 2018 - Jul 2022*
SCAD College of Engineering and Technology, Tamil Nadu, India — Grade: 8.25