GLENN DALBEY

Data Science & Analytics

319-233-4445 | dalbeyglenn@gmail.com
LinkedIn | GitHub | Portfolio

TECHNICAL SKILLS

Programming

Python (Expert), SQL, TypeScript, JavaScript, R, C++, HTML/CSS

Deep Learning

PyTorch, TensorFlow/Keras, Spatial-Temporal Transformers, GRU/RNN, 3D CNNs, Perceiver IO, Geometric Attention Networks

ML & Data Science

scikit-learn, XGBoost, LightGBM, CatBoost, Pandas, NumPy, Advanced Ensemble Methods, Feature Engineering

Cloud & Infrastructure

AWS (Certified), Multi-GPU Training, Mixed Precision (FP16), PostgreSQL, S3, Weights & Biases

Homelab & Systems

Proxmox VE, pfSense, Ubuntu Server, n8n Automation, RAG Pipelines, 10Gb Networking, VLAN

Web & Deployment

Flask, FastAPI, Streamlit, React/Next.js, Docker, CI/CD, GitHub Actions

Specializations

Trajectory Prediction, 3D Medical Imaging (DICOM/NIfTI), Computer Vision, NLP, Multi-modal AI, Test-Time Augmentation

EDUCATION

MS, Data Science & Analytics

Western Governors University
Aug 2024 - Aug 2025
• Healthcare AI Capstone (Production)
• NSLS & Student Insights Council

BS, Data Analytics

Western Governors University
Mar 2023 - Sep 2024
• NFL Prediction Capstone
• NSLS Member

PROFESSIONAL SUMMARY

Data Scientist with an MS in Data Science who builds things that actually work. I enjoy building and training models on my own multi-GPU homelab, deploying new techniques, and discovering unique ways to solve problems. Whether it's analysis, research, or projections, I dig into how something ticks, figure out why, and extract information that drives better outcomes. My multi-node homelab stack lets me run production applications, train models, and run MLflow, all on my own hardware. 20+ projects on GitHub covering deep learning, medical imaging, NLP, computer vision, and analytics.

FEATURED DATA SCIENCE PROJECTS

NFL Big Data Bowl 2026 - Kaggle Bronze Medal

Deep Learning Player Trajectory Prediction | Top 8% of 1,134 teams

github.com/XxRemsteelexX/NFL-Big-Data-Bowl-2026-

- **Bronze Medal** in prestigious Kaggle competition predicting NFL player trajectories from tracking data
- Conducted systematic exploration of 847+ experiments across 15+ neural network architectures
- Best single model: 6-Layer Spatial-Temporal Transformer achieving 0.547 Public LB score
- Best ensemble: 3-model blend (ST Transformer + CNN + GRU) achieving 0.540 Public LB with architecture diversity
- Engineered 167 features including kinematics, ball-relative positions, temporal patterns, and geometric features with Voronoi tessellation
- Implemented novel geometric attention with spatial distance modulation and frozen encoder fine-tuning
- Utilized multi-GPU training, mixed precision (FP16), and test-time augmentation for +0.005-0.010 improvement

RSNA Intracranial Aneurysm Detection - Kaggle Competition

3D Deep Learning Medical Imaging | 105 Models Trained

github.com/XxRemsteelexX/RSNA-Intracranial-Aneurysm-Detection-Kaggle

- Trained 105 deep learning models (21 architectures × 5 folds) for CT angiography aneurysm detection
- Tested 51 ensemble configurations; best ensemble META_E_top3_weighted achieved AUC 0.8624
- Key discovery: Smaller models (SE-ResNet18) statistically outperform larger models on limited medical data ($r=-0.42$, $p<0.01$)
- Built complete pipeline: DICOM → NIfTI → ROI extraction → Training → Ensemble across 4 GPUs simultaneously
- Multi-label classification across 14 classes with severe class imbalance handling (1.2% to 42.8%)

AS, IT / Programming
Clinton Community College
Jan 2022 - Dec 2022
• Data Analytics Certificate
• Phi Theta Kappa

Diploma, IT / Programming
Clinton Community College
Jan 2021 - Dec 2021
• Phi Theta Kappa

AA, Art Studies
Hawkeye Community College
Sep 2019 - Dec 2020
• Phi Theta Kappa

Graphic Communications
Hawkeye Community College
Sep 1997 - May 1999

CERTIFICATIONS

Industry
• CompTIA Data+ (2024-2027)
• AWS Cloud Practitioner (2024-2027)
• CompTIA A+ (2023-2026)

Udacity Nanodegrees
• Advanced Computer Vision & Deep Learning
• ML DevOps & Model Deployment
• Transformer Models & BERT
• GANs & Convolutional Neural Networks

HONORS
• National Society of Leadership and Success
• Phi Theta Kappa Honor Society
• Eagle Scout (1995)

Apollo Healthcare Connect
Production Multi-modal AI Healthcare Triage System | MS Capstone

apollohealthcareconnect.com

- Built and deployed live production healthcare AI triage system with sub-second response times
- Achieved **93.8% combined multi-modal accuracy** and **98.0% burn classification accuracy**
- Implemented 5-model ensemble architecture combining DistilBERT (NLP) and CNNs (Computer Vision)
- Successfully handled extreme class imbalance (29.7:1 ratio) with advanced loss functions
- Built production pipeline with Flask API, AWS S3 integration, and comprehensive safety protocols

Missing Persons Outlier Detection
Geospatial Crime Pattern Analysis | 41,200 NamUs Cases

github.com/XxRemsteelexX/missing-persons-outlier-detection

- 7 statistical methods + 3 ML models applied to 41,200 cases across 101 years (9,204 county-decade combinations)
- Kenedy County, TX: 46.86σ composite z-score persists after Bayes shrinkage, OLS, Random Forest, and FDR correction
- I-35 corridor: 170% increase in missing persons (193 to 521 cases), structural break at 2020 ($p < 0.001$)
- Spatial autocorrelation confirmed (Moran's I = 0.22, z = 26.03) with LISA hotspot clustering along TX border
- Validated against known serial killers (Ridgway: 4.38σ, Gacy: 1.34σ)
- Live 7-page Streamlit dashboard with geospatial visualization and ARIMA 5-year forecasting

NFL Rookie Wide Receiver Performance Prediction
Advanced ML Analysis with Feature Optimization | BS Capstone

github.com/XxRemsteelexX/NFL_Rookie_WR_1K_Analysis

- Developed predictive model achieving **90.9% ROC AUC** on future data validation for 1000+ yard seasons
- Reduced overfitting gap from 18.5% to 0.4% (97.8% reduction) through feature optimization (46→20 features)
- Implemented temporal validation strategy ensuring model generalization to future NFL seasons
- Created production-ready ensemble model for NFL draft analysis with comprehensive data pipeline (2006-2024)

OE-OS (In Progress)
Distributed AI Orchestration Platform | Python / FastAPI

- Three-tier LLM routing (local Ollama to cheap API to premium models) reducing costs by routing ~80% of requests to free local models
- Triple-layer RAG memory: BM25 over 5M+ chunks, ChromaDB semantic search, Redis session cache with graceful degradation
- 18 MCP-compatible tools and multi-agent sandbox where 4 LLM personas deliberate at zero API cost
- 4,200+ lines of async Python on FastAPI for a private multi-node GPU cluster

AI Homelab & Active Memory Network

Multi-Tier AI Infrastructure | 10Gb Network + RAG Pipeline

glenndalbey.com/infrastructure

- Designed and operate multi-tier AI homelab: **dual RTX 5090** training node + RTX 3090 Ti/3090 secondary node
- Built 256GB unified memory LLM inference cluster (2× Ryzen AI Max+ 395) running Kimi K2, Qwen 3, GLM 4.6
- Implemented automated active-memory pipeline with n8n orchestration, RAG storage, and hot/warm/cold tiering
- Deployed Proxmox VE backbone with pfSense firewall, VLAN segmentation, and 10Gb networking (100TB+ storage)

PROFESSIONAL EXPERIENCE

Freelance Data Science Consultant

Thompson Parking & Mobility Consultants

Current

- Provide data science and analytics consulting services for business intelligence initiatives
- Develop AI-powered Excel analytics platform enabling natural language data queries
- Design custom analytical solutions and machine learning models for client-specific challenges
- Support data-driven decision making through advanced analytics and predictive modeling

Continuous Improvement Leader & Material Flow Specialist

John Deere, Waterloo Works & Ankeny Works

2005-2020, 2021-Present

- CI Department Representative leading process improvement and operational efficiency initiatives
- Developed comprehensive training curriculum for warehouse personnel, improving onboarding efficiency
- Designed and implemented the Zones Project, modernizing material flow training systems
- Led departmental CI mapping initiatives to improve operational efficiency and reduce cycle times
- Optimized material replenishment processes using bin methodology, reducing operational inefficiencies
- Managed supply chain logistics and SAP-integrated inventory management
- Supported engineering teams in workflow re-splits and cycle time analysis for production optimization