# Daniel Doyon

+1(908) 798-2433 / dannydoyon123@gmail.com / linkedin.com/in/danieldoy0n/ / github.com/danhof1 / Hempstead, NY

## EDUCATION

**HofstraUniversity** Hempstead, NY
*Data Science M.S.- Expected Grad. May 2026* *Cumulative GPA: 3.85*
- Relevantcoursework: Data Warehousing & SQL, Machine Learning, Neural Networks, Regression Methods

*Computer Science & Mathematics B.S.* *Cumulative GPA: 3.5*
- Relevant coursework: Data Science, Text Mining, Probability and Statistics I & II, Linear Algebra, Abstract Algebra, Real Analysis, Calculus I, II, III, Data Structures and Algorithms, Differential Equations, Discrete Mathematics, Software Engineering

## AWARDS, RECOGNITION & SOCIETIES

**Dean's List**: Sept 2021 – Dec 2022, Jan 2024 – May 2025
**2nd Place** – Senior Capstone Competition – Spring 2025
**Societies** – Theta Tau Professional Engineering Fraternity, Pi Mu Epsilon Math Honors Society

## TECHNICAL SKILLS

**Programming Languages**: Python, C++, Java, C#, R, SQL
**AI & Machine Learning**: PyTorch, Cuda, TensorFlow, scikit-learn, LLMs, NLP, Fine-tuning, Multi-label Classification
**Advanced ML Techniques**: Hyperparameter Optimization, Model Ensembling, Vector Embeddings, Dimensionality Reduction, Feature Engineering, Statistical Modeling
**Frameworks/Libraries**: NumPy, Pandas, SpaCy, Matplotlib, Seaborn, Transformers, NLTK
**Tools & Platforms**: Power BI, Git, Docker, Linux (Arch & Debian), MongoDB, Optuna, Firebase, NVIDIA Container Toolkit

## EXPERIENCE

**Teachers Assistant** (C++ & AI) **Feb 2026 - Present**
- Facilitate weekly laboratory sessions for Introduction to C++, guiding students through coding exercises, debugging syntax errors, and reinforcing core programming concepts.
- Provide one-on-one tutoring to support student learning and clarify complex course material.
- Evaluate and grade coursework for both Introduction to C++ and Artificial Intelligence courses, providing detailed feedback on assignments, projects, and exams to ensure academic standards are met.

**Statistical Consulting for Harmony Healthcare** | *R, Data Visualization, Data Management(Team of 3)* Jan 2025 – May 2025
- Built predictive model for 30-day hospital readmission risk on 2,822 patient records in R; identified 6 critical factors (Education, Medication Management, Racial Disparities, Healthcare Utilization, Social Determinants, Clinical Indicators) driving readmission
- Cleaned and standardized 675 clinical variables addressing missing data and text inconsistencies; created visualizations showing relationships between key variables and readmission outcomes to communicate findings to stakeholders

**Math Tutor – Hofstra University** **Nov 2024 – May 2025**
- Mentored 10+ students in mathematics, specializing in algebra, calculus, and pre-calculus concepts
- Create personalized teaching methods and resource management strategies to cater to individual student needs and learning styles

**Research Assistant -** *Computer Vision Research* **May 2024 – August 2024**
- Developed deep learning pipeline for autonomous wound segmentation to support robotic surgical closure system
- Achieved 93.7% Dice score using optimized U-Net architecture with custom preprocessing, hyperparameter tuning, and model ensembling

## PROJECTS

**Master's Thesis (In Progress)-** ***Unlearning for Improved Differentially Private Synthetic Text Generation*** **Sept 2025 – Present**
- Developing a novel pipeline that applies machine unlearning (ReLearn) to Llama-3-8B to remove memorized pre-training data before applying Differential Privacy (DP) fine-tuning.
- Addressing the "privacy-utility gap" in LLMs by forcing models to learn from private datasets rather than relying on contaminated training knowledge
- Utilizing MIMIC-III clinical notes to evaluate the impact of targeted unlearning on Membership Inference Attack (MIA) success rates and downstream classification utility.
- Leveraging high-performance computing (NVIDIA H100 GPUs) to implement DP-SGD with LoRA and evaluate privacy budgets epsilon ranging from 0.5 to 4.

**Senior Design - Gsplat AI Pipeline** | *Python, PyTorch, Docker, Firebase, Flutter, CUDA (Team of 4)* **Dec 2024 – May 2025**
- **2nd Place, Hofstra Spring 2025 CS & E Senior Capstone:** Recognized for innovation in real-time 3D reconstruction using emerging technologies
- Collaborated as Sprint Master for one of three development sprints, coordinating team efforts across frontend, database, mobile, and integration components with effective resource management
- Engineered end-to-end automated, GPU-accelerated pipeline leveraging machine learning algorithms to convert user-supplied images into textured 3D meshes (.ply/.obj) in under 5 minutes
- Developed scalable ML workflows using Python microservices and Docker integration for automation, with Firebase data management, state- tracking and fault-tolerance for production deployment

**Semantic Investment Analysis Platform** | *Python,MongoDB, LLMs,Sentence Transformers(Team of 4)* **Nov 2025 – Dec 2025**
- Engineered a hybrid retrieval system using BM25 and vector similarity to enable high-accuracy semantic search across a data warehouse of 4,600+ ETFs.
- Developed an LLM-powered pipeline for financial research, implementing embedding-based clustering and sentence transformers to automate the summarization of complex investment metadata.
- Architected a NoSQL data warehouse in MongoDB with compound indexes to manage high-frequency financial data, ensuring 100% data collection reliability through custom error handling.

**Home Credit Default Risk Prediction** | *Python, PyTorch, Light-GBM, Optuna* **Sept 2025 – Dec 2025**
- Engineered 50+ features from 7 relational tables (credit history, payment behavior, previous applications) for loan default prediction on highly imbalanced dataset
- Implemented multi-objective Optuna optimization (ROC-AUC + PR-AUC) with Pareto front analysis; compared neural network, LightGBM, and Random Forest architectures