Michael Paglia

518-852-7364 | michaelpagliadev@gmail.com

mpaglia.com | linkedin.com/in/michaeljpaglia | github.com/michaelpaglia

TECHNICAL SKILLS

Languages: Python, SQL, JavaScript, Java, C++
ML/Data: PyMC, pandas, NumPy, scikit-learn, TensorLy, Matplotlib, Seaborn, Presto, Daiquery
Cloud & Tools: AWS S3, GCP, Docker, Git, PostgreSQL, Flask

EXPERIENCE

Meta — New York, NY
Lead Data Analyst (Generative AI) — *May 2024 – Present*

- Manage multiple data annotation projects end-to-end, collaborating with researchers to enhance **AI systems** and deploy specialized knowledge experts to optimize data quality for **AI agents** powered by **Llama 4** using **Python**, **SQL**, and **Presto/Daiquery**
- Lead and mentored 30+ analysts on scalable data methodologies while developing bespoke **dashboards** and **data pipeline solutions**, improving operational efficiency by 40% and achieving 90% research team approval rating

518 Cares — Troy, NY
Founder — *Mar. 2025 – Present*

- Founded healthcare access non-profit serving 50K+ population in Troy, NY community with **web-based resource database** and **eligibility screening tool** at 518cares.org
- Developed platform aggregating free and low-cost healthcare services with **privacy-first architecture** using **JavaScript** and **PostgreSQL**

The Research Foundation for SUNY (Data Management and Mining Lab) — Albany, NY
National Science Foundation Research Fellow — *Jan. 2024 – May 2024*

- Implemented and tested 10+ **machine learning algorithms** for temporal graph and multi-dictionary signal decomposition using **Python** and **NumPy**, achieving a fit of over 98% to datasets with 20M+ points for data reconstruction
- Integrated custom graph and **time-series dictionary generation** to reduce data retrieval time by nearly 90% using **TensorLy** and **AWS S3**

Regeneron Pharmaceuticals, Inc. — Rensselaer, NY
Software Engineering Intern — *May 2023 – Aug. 2023*

- Developed and deployed a **CRUD application** and **REST API** using **Python**, **Flask**, and **JavaScript** used by IT team members, improving operational efficiency and saving 2 hours of manual effort each week
- Provided **smart contract-based access management** using **Dremio Data Lake**, achieving a 20% improvement in secure data access verification times across an organization of 4K+ employees

EDUCATION

University at Albany — Albany, NY
Bachelor of Science in Computer Science — *May 2024*

- Graduated *summa cum laude* (GPA: 3.85/4.0)
- Member of Phi Beta Kappa

PROJECTS

Premier League Predictions | *PyMC, FBRef API* | https://bit.ly/4p0lw7J — Nov. 2024

- **Bayesian model** using **Weibull hazard functions** and **MCMC sampling** to predict match outcomes with 8% vig compared to major sportsbooks
- Engineered **ML pipeline** with informative priors and normalized tensors implementing time-dependent goal scoring patterns

PySpady | *Python, NumPy, TensorLy, AWS S3* | https://bit.ly/pyspady — Sep. 2023 – May 2024

- **ML-driven tensor decomposition library** with automated model selection for optimal dictionary, model, and optimizer combinations