

# Michael Paglia

Troy NY, 12180, United States | 518-852-7364 | [michaelpagliadev@gmail.com](mailto:michaelpagliadev@gmail.com)  
[mpaglia.com](http://mpaglia.com) | [linkedin.com/in/michaeljpaglia](https://www.linkedin.com/in/michaeljpaglia) | [github.com/michaelpaglia](https://github.com/michaelpaglia)

## 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

## EXPERIENCE

---

### Agents Team Lead - Data Labeling Analyst IV (Generative AI)

Oct. 2024 – Present

*Meta*

*Remote*

- Lead, managed, and trained a team of 30+ analysts, providing mentorship on web scraping techniques and data curation methodologies to extract actionable insights for six-figure weekly budget projects
- Developed automated data pipelines that improved synthetic data quality by 40%, achieving a 90% approval rating from research team while optimizing resource allocation for high-value AI initiatives

### Coding Data Labeling Analyst II (Generative AI)

May 2024 – Oct. 2024

*Meta*

*Remote*

- On assignment via Magnit, conducted over 1K+ quality assurance tests weekly on code generated by Llama 3
- Collaborated with project managers to improve vendor performance to yield a 15% increase in labeling accuracy

### Founder and CEO

Nov. 2020 – Present

*M.P. Mercantile, LLC*

*Troy, NY*

- Founded a start-up e-commerce business hosted on 3 major online marketplaces with 1K+ unique customers
- Achieved over six figures annually in gross sales revenue and 25% growth year-over-year

### National Science Foundation Research Fellow

Jan. 2024 – May 2024

*The Research Foundation for SUNY*

*Albany, NY*

- Supervised under the guidance of Dr. Petko Bogdanov in the Data Management and Mining Lab
- Implemented and tested 10+ machine learning algorithms for temporal graph and multi-dictionary signal decomposition, 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%

### STEM Tutor

June 2022 – May 2024

*Paper*

*Remote*

- Coached over 500 students per month on mathematics and computer science using question-based teaching method
- Offered bilingual (English and Spanish) support while working with up to 4 students concurrently

### Software Engineering Intern

May 2023 – Aug. 2023

*Regeneron Pharmaceuticals, Inc.*

*Rensselaer, NY*

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

## PROJECTS

---

### PySpady | *Python, NumPy, TensorLy, Terraform, AWS S3, Git* | <https://bit.ly/pyspady>

Sep. 2023 – May 2024

- Features data compression, missing value imputation, future value prediction, and autocompletion
- Machine learning determines the best combination of dictionaries, models, and optimizers for a given dataset

### Custom Language Interpreter | *Java, Git* | <https://bit.ly/custom-interpret>

Aug. 2022 – Nov. 2022

- Defines a functional programming language with pseudocode-like syntax to introduce people to coding
- Capable of lexical analysis, parsing, and semantic analysis into an interpreted output

## TECHNICAL SKILLS

---

**Languages:** Python, Java, JavaScript, SQL, C++, C#, C

**ML/Data:** pandas, NumPy, TensorLy, Matplotlib, Seaborn

**Web & APIs:** Flask, Django, Node.js, HTML/CSS

**Tools:** Git, Docker, GCP, Dremio, VS Code