

Shams Rupak

New York | 646-619-7873 | shamsrupak@gmail.com | [LinkedIn](#) | [GitHub](#)

Professional Summary

Software Engineer Intern with hands-on experience building AI-powered applications, machine learning pipelines, and full-stack solutions. Skilled in developing scalable systems with Python, JavaScript, and cloud tools, backed by a strong foundation in applied mathematics and data-driven problem solving. Proven track record of integrating AI and automation to streamline workflows and deliver production-ready software. Passionate about creating impactful, user-focused technology through cross-functional collaboration.

Education

Stony Brook University, Stony Brook, NY

Master of Science, Engineering Artificial Intelligence

Aug 2025 - Dec 2026

Stony Brook University, Stony Brook, NY

Bachelor of Science, Applied Mathematics and Statistics

Aug 2021 - May 2025

- **Coursework:** Probability Theory, Numerical Analysis, Operations Research, Financial Mathematics, Data Mining, Object-Oriented Programming, Data Structures

Experience

Outatation

May 2025 - Jul 2025

AI and Automation Extern

Remote

- Engineered AI-powered document processing workflows using Python and NLP techniques, automating classification and data extraction with computer vision, PyMuPDF, and OCR, which reduced manual handling by 60%
- Developed a RAG-powered retrieval system with LlamaIndex to enhance search precision across unstructured mortgage data while evaluating transformer-based LLMs for production-level model optimization
- Synthesized benchmarking results and deployment insights, assessing top open-source AI models for document classification with a focus on trade-offs in speed, context window, and accuracy

Stony Brook University

Jan 2025 - Mar 2025

Biomedical Informatics Bootcamp

- Applied machine learning techniques on biomedical datasets using Python and various visualization libraries to derive actionable insights
- Collaborated with interdisciplinary teams to address real-world healthcare challenges, showcasing a production-oriented mindset in problem-solving and solution deployment

T-Mobile

Jul 2024 - May 2025

Retail Mobile Expert

Nesconset, NY

- Exceeded 100% of monthly performance metrics by integrating AI tools like ChatGPT and Apple Intelligence into sales workflows, streamlining quoting processes and earning recognition for leading tech adoption and operational improvements across the team

Webacy

Sep 2024 - Nov 2024

Web3 Security Data Analytics Extern

Remote

- Applied clustering techniques and unsupervised machine learning to detect anomalous transaction patterns and smart contract vulnerabilities, improving blockchain risk categorization accuracy
- Enhanced data integrity by peer-reviewing and validating labels across blockchain datasets, achieving over 95% reliability and supporting robust analytics in production environments

Projects

Student Buddy: AI Chrome Extension | <https://github.com/ShamsRupak/student-buddy-extension>

Jul 2025 - Present

- Engineered Chrome extension with JavaScript, HTML/CSS, and Gemini API to assist students with real-time academic queries
- Designed intuitive UI/UX and implemented modular architecture for maintainable code and future scalability

NYC Weather Prediction: ML Regression Pipeline

May 2025 - Jul 2025

<https://github.com/ShamsRupak/nyc-weather-prediction>

- Built ML models using XGBoost and Scikit-learn to predict temperature trends from NYC weather datasets
- Performed EDA, feature engineering, and model persistence using joblib for reproducible results

Financial Document Classification with RAG & OCR

May 2025 - Jul 2025

<https://github.com/ShamsRupak/ai-doc-processing-suite>

- Created document pipeline using PyMuPDF and OCR to automate mortgage document classification
- Enabled semantic search across heterogeneous file types using Retrieval-Augmented Generation

Technical Skills

- **Programming:** Python, JavaScript, Java, HTML, CSS, SQL, C, R, MATLAB, VBA
- **Libraries/Tools:** React, Git, Jupyter, LlamaIndex, PyMuPDF, joblib, Chrome Extensions, NumPy, Pandas, Matplotlib, Seaborn, OCR, PyTorch, VS Code, Scikit-learn, XGBoost
- **Competencies:** Statistical Modeling, Machine Learning, Natural Language Processing, Retrieval-Augmented Generation (RAG), Data Visualization, Full Stack Development, AI Integration, UI/UX Design, Semantic Search, Web Scraping