

Rushi Balapure

+91-8329707611 | rushibalapure2003@gmail.com | github.com/Rushi-Balapure | linkedin.com/in/rushi-balapure-75148022a/

Summary

Associate Product Engineer with hands-on experience building enterprise security products across Cloud DSPM, Secure Web Gateway, and Data Loss Prevention. Experienced in backend engineering, ML/LLM systems, security-focused architecture, research, testing, and deployment, with work spanning sensitive-data detection, NER, self-hosted LLMs, and large-scale domain intelligence systems.

Experience

Associate Product Engineer — ProMobi Technologies, Pune

July 2025 – Present

Cloud Data Security Posture Management — Veltar

- Designed and personally implemented a three-layer sensitive-data detection architecture combining deterministic rules, NER models, and a self-hosted LLM to detect structured, semi-structured, and unstructured sensitive information with approximately 90% detection accuracy.
- Built rule-based detectors using regex and identifier-specific validators, including checksum-based validation for structured sensitive data such as payment and identity information.
- Integrated Stanford De-identifier and GLiNER biomedical NER models for entity detection, with a self-hosted Qwen3.5 2B model served through vLLM for semantic detection of sensitive business information.
- Developed file-processing pipelines for Google Drive, Microsoft OneDrive, and Amazon S3, supporting text files, PDFs, spreadsheets, presentations, documents, and media formats using format-specific extraction and OCR.
- Engineered the platform to support at least **5 TB of data per customer**, with current processing throughput of approximately **5 GB/hour**.

Secure Web Gateway — Veltar

- Co-developed an enterprise Secure Web Gateway used across approximately **100 customers and 6,000 managed devices**, enabling administrators to enforce website access policies across managed endpoints.
- Integrated and maintained open-source threat intelligence feeds through scheduled ingestion jobs, building a domain intelligence dataset covering **60M+ known domains**.
- Developed a multi-stage website categorization pipeline combining content scraping, keyword-based classification, ML-based classification, and LLM-assisted categorization for ambiguous domains.
- Supported classification across **45+ website categories** and contributed to a **75% reduction in customer-reported misclassification issues** based on production monitoring.

Secure Web Gateway (DLP) — Veltar

- Contributed to a managed-device DLP capability that prevents sensitive files from being uploaded to cloud services or transmitted through email based on administrator-defined security policies.
- Supported file-extension and regex-based controls for detecting and blocking sensitive data such as exposed AWS credentials on devices managed through Scalefusion MDM.

Skills

Python, Ruby, JavaScript, SQL, Ruby on Rails, FastAPI, Machine Learning, LLMs, RAG, PyTorch, scikit-learn, NLTK, vLLM, Named Entity Recognition (NER), PostgreSQL, MySQL, Cloud DSPM, Data Loss Prevention (DLP), Secure Web Gateway, Sensitive Data Detection, OOP, Operating Systems, Computer Networks, DSA, DBMS

Education

Bachelor of Engineering, Computer Engineering

Pune Institute of Computer Technology
Savitribai Phule Pune University

Open-Source Project

[pdf-2-json-extractor](#)

- Built and published a Python package on PyPI for converting PDFs into structured JSON while preserving document layout and reading context.
- Implemented PDF extraction using PyMuPDF and a research-paper-based layout reconstruction approach to retain relationships between text blocks that are typically lost during plain-text extraction.
- Designed the output specifically for downstream LLM and document-analysis workflows, where layout preservation improves contextual understanding compared with conventional text-only parsing.

Achievements

- [IEEE Author Profile / Publications](#)Published research papers in the domains of Blockchain, Cryptography, and Multi-Party Computation (MPC).
- Delivered a [Lightning Talk at HP Pune](#) on AI-powered security tools.
- Conducted a hands-on [Workshop on LLM Security at PICT Pune](#), covering Generative AI red teaming, blue teaming, and secure AI deployment practices.
- Built and maintain the open-source Python library [pdf-2-json-extractor](#) for layout-aware PDF-to-JSON extraction.
- Secured 1st Position at the Innerve Hackathon hosted by AIT Pune.