PII Redaction Tool

Overview

This report presents a solution for Redact PII from Text and PDFs. The proposed system is a web-based application that detects and redacts Personally Identifiable Information (PII) from both plain text and PDF files using a hybrid approach (regex patterns, and LLM-based) detection methods.

Key Features:

- Multi-format input support (text, PDF)
- Hybrid PII detection
- Configurable redaction methods
- Professional web interface (UI)

Problem Statement

Organizations need to protect sensitive data before using it with Large Language Models or sharing documents. Manual PII redaction is time-consuming and error-prone, while automated solutions often lack accuracy or flexibility.

Solution Approach

A comprehensive PII redaction tool that combines multiple detection methods:

1. Pattern-based detection for structured PII (emails, phones, SSNs)
2. Named Entity Recognition for contextual PII (names, locations)
3. LLM-enhanced detection for complex or ambiguous cases
4. Configurable redaction with multiple output formats

Why This Approach?

Based on research analysis, this hybrid approach addresses the limitations of single-method solutions:

- Regex alone misses context-dependent PII
- NER models struggle with structured identifiers
- LLMs provide accuracy but require optimization for speed
- Combination leverages strengths while mitigating weaknesses

Core Components

1. Multi-Format Processing

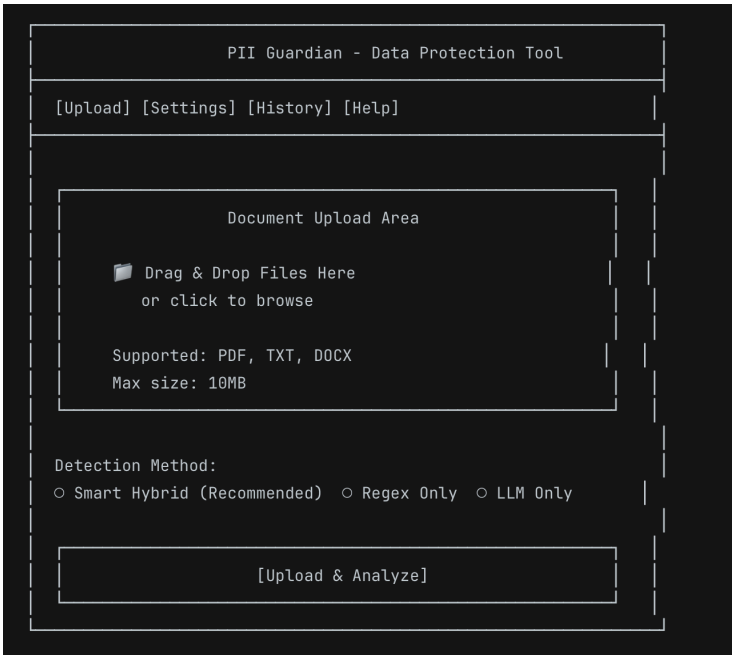
- Text Input: Direct processing through detection engine
- PDF Processing: PyPDF2 for text extraction + Tesseract OCR for scanned documents
- Format Preservation: Maintains original document structure during redaction

2. Redaction Methods

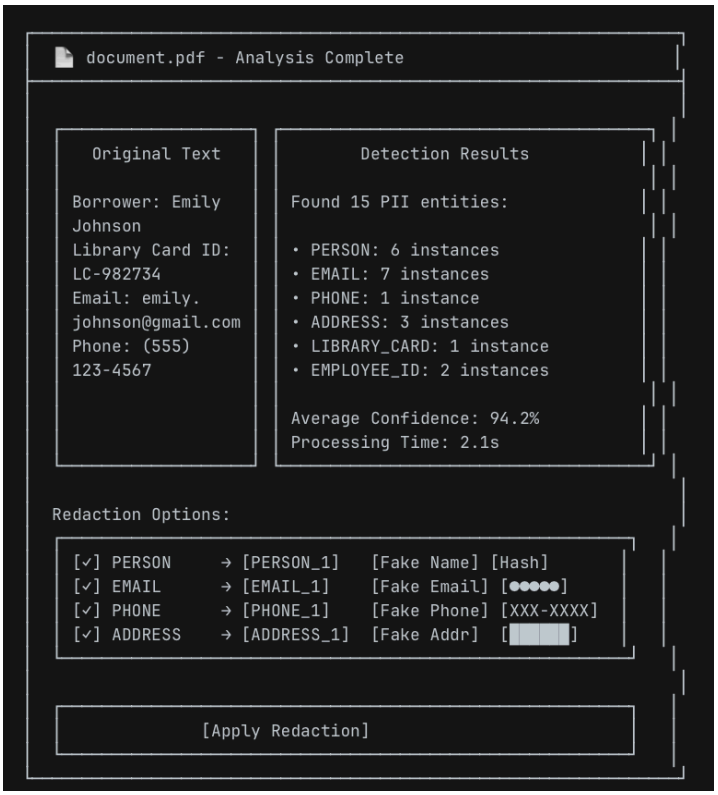
- Placeholder Labels: `[EMAIL_1]`, `[PHONE_1]`, `[PERSON_1]`
- Masking: ▇▇▇▇▇▇▇ or `XXX-XX-XXXX`
- Fake Data: Realistic replacements using Faker library
- Hashing: SHA256 hashes for irreversible anonymizatio

User Interface Design

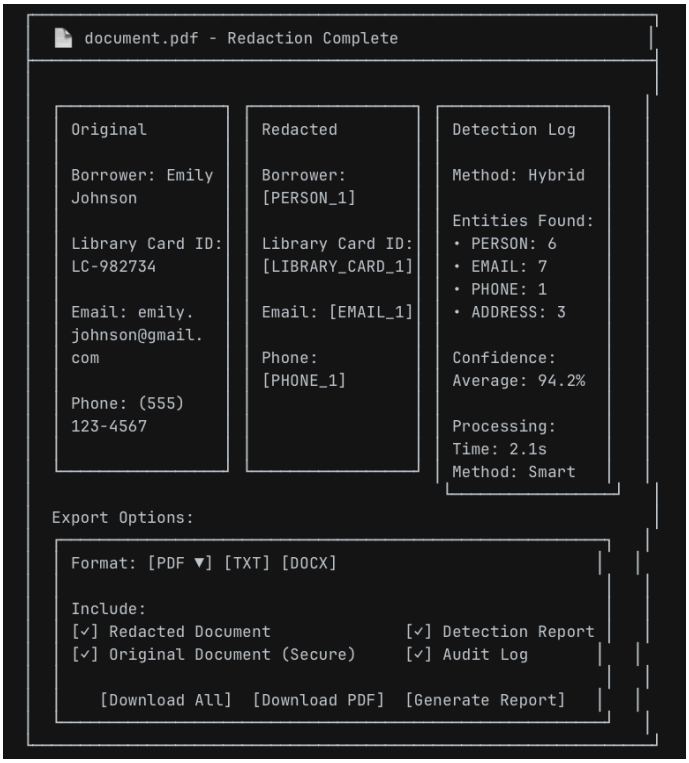
Main Application Interface



Analysis Results View



Export & Results View



Scope of Work

2.1 Core Functionality

File Input Processing

- Plain text input via textarea (up to 50,000 characters)
- PDF file upload with text extraction (up to 10MB)
- Support for both text-based and scanned PDFs
- Real-time processing with progress indicators

PII Detection Engine

- Microsoft Presidio Integration: Pre-trained recognizers for standard PII
- Custom Pattern Recognition: Domain-specific identifiers (Library cards, Employee IDs)
- LLM Enhancement: Context-aware detection for ambiguous cases
- Confidence Scoring: Per-entity confidence levels with explanation

Supported PII Types (12+ Categories):

- Person names (first, last, full names)
- Email addresses (all common formats)
- Phone numbers (US/International formats)
- Physical addresses (street, city, state)
- Social Security Numbers (XXX-XX-XXXX format)
- Credit card numbers (all major brands)
- IP addresses (IPv4/IPv6)
- Bank account numbers
- Library card IDs (LC-XXXXXX pattern)
- Employee IDs (EMP-XXXX pattern)
- Usernames (alphanumeric combinations)
- Office locations and building references

Configurable Redaction Methods

- Placeholder Labels: `[EMAIL_1]`, `[PHONE_1]`, `[PERSON_1]`
- Masking: ███████, `XXX-XX-XXXX`, ●●●●●●●●
- Fake Data Replacement: Realistic substitutions using Faker
- Hash Values: SHA256/MD5 for irreversible anonymization
- Custom Patterns: User-defined replacement strategies

Results Display & Export

- Side-by-side original vs. redacted comparison
- Multiple export formats (PDF, TXT)

2.2 User Interface Features

Modern Web Application

- Responsive design (desktop, tablet, mobile)
- Drag-and-drop file upload interface
- Real-time processing indicators
- Interactive entity highlighting
- Professional corporate styling

User Experience Enhancements

- One-click redaction with smart defaults
- Live preview of redaction results