

Documentation

Achievements

1. Uploading PDFs

- You can upload one PDF at a time or multiple PDFs together.
- Only .pdf files are allowed, and the system double-checks that the file actually is a PDF.
- Large PDFs are supported, but there's a safety limit of 300 pages.
- Everything is handled in memory, so no sensitive data is unnecessarily written to disk.

2. Redacting Content

- You can redact specific words or phrases by entering them in a simple text box.
- You can choose to redact all pages or just specific pages.
- Optionally, you can remove images or logos from the PDF.
- There's a placeholder text option (like "[REDACTED]") to mark redacted areas.
- Redactions happen securely in memory, ensuring sensitive info isn't exposed.

3. Bulk Handling

- Multiple PDFs can be uploaded and processed at once.
- After redaction, all files are packaged into a ZIP file for easy download.
- Each file can also be previewed individually before downloading.
- Previews and redacted files are kept in the session, so you can view or download anytime without re-uploading.

4. Preview & Interface

- Preview the original PDF before any changes.
- Preview the redacted PDF after processing.
- In bulk mode, previews are neatly organized in expandable sections for each file.
- Download options are clear:
 - Single redacted file.
 - Individual files from a bulk upload.
 - Entire ZIP for all bulk files.

5. Security & Logging

- All processing happens in memory; files are not stored permanently.

- The app keeps audit logs recording:
 - Who uploaded the file
 - Which words were redacted
 - Pages processed
 - If images were removed
 - Success or failure of each operation

6. Error Handling

- Invalid files or unsupported PDFs are rejected with clear messages.
- PDFs exceeding the page limit are blocked.
- Page ranges are checked against the actual file to prevent mistakes.
- All failures are logged for auditing.

Screenshots

Logging

The screenshot shows a VS Code editor with a project named 'poc'. The Explorer sidebar on the left shows the file structure, including 'Redaction_poc', 'poc', 'streamlit', 'secrets.toml', 'backend', 'pycache_', 'main.py', 'redactor.py', 'frontend', 'app.py', 'audit.log', and 'pwc_poc'. The 'audit.log' file is selected and open in the editor. It contains a series of log entries with timestamps, log levels (INFO, ERROR, SUCCESS), and detailed messages about file processing, including file names, words, and page ranges. The bottom of the editor shows the 'PROBLEMS' panel with a list of errors, including '400 Bad Request' and 'Error=document closed or No valid PDFs were processed'. The 'TERMINAL' panel shows the output of the application, including the URL 'http://127.0.0.1:8080/' and the status '200 OK'.

```

18 2025-08-21 20:13:08,377 - INFO - SUCCESS: File-National Apprenticeship Training Scheme (NATS).pdf, Words=['I'], Pages=[0], RemoveImages=True
19 2025-08-21 20:14:59,070 - INFO - SUCCESS: File-National Apprenticeship Training Scheme (NATS).pdf, Words=['I'], Pages=[0], RemoveImages=False
20 2025-08-21 20:15:20,577 - INFO - SUCCESS: File-National Apprenticeship Training Scheme (NATS).pdf, Words=['I'], Pages=[0], RemoveImages=True
21 2025-08-21 20:35:16,230 - INFO - SUCCESS [single]: Client=127.0.0.1, File-National Apprenticeship Training Scheme (NATS).pdf, Words=['I'], Pages=ALL
22 2025-08-21 20:35:56,335 - INFO - SUCCESS [single]: Client=127.0.0.1, File-National Apprenticeship Training Scheme (NATS).pdf, Words=['I'], Pages=ALL
23 2025-08-21 20:36:54,130 - INFO - SUCCESS [bulk]: Client=127.0.0.1, File-National Apprenticeship Training Scheme (NATS).pdf, Words=['I'], Pages=ALL,
24 2025-08-21 20:36:54,366 - INFO - SUCCESS [bulk]: Client=127.0.0.1, File-CT20182437032_Application.pdf, Words=['I'], Pages=ALL, RemoveImages=False
25 2025-08-21 20:37:14,200 - INFO - SUCCESS [bulk]: Client=127.0.0.1, File-National Apprenticeship Training Scheme (NATS).pdf, Words=['I'], Pages=ALL,
26 2025-08-21 20:37:14,464 - INFO - SUCCESS [bulk]: Client=127.0.0.1, File-CT20182437032_Application.pdf, Words=['I'], Pages=ALL, RemoveImages=True
27 2025-08-21 20:37:53,583 - ERROR - FAILURE [bulk]: Client=127.0.0.1, File-National Apprenticeship Training Scheme (NATS).pdf, Error=Page index out o
28 2025-08-21 20:37:53,598 - INFO - SUCCESS [bulk]: Client=127.0.0.1, File-CT20182437032_Application.pdf, Words=['I'], Pages=[1], RemoveImages=True
29 2025-08-21 20:38:05,608 - INFO - SUCCESS [bulk]: Client=127.0.0.1, File-National Apprenticeship Training Scheme (NATS).pdf, Words=['I'], Pages=[0],
30 2025-08-21 20:38:05,631 - INFO - SUCCESS [bulk]: Client=127.0.0.1, File-CT20182437032_Application.pdf, Words=['I'], Pages=[0], RemoveImages=True
31 2025-08-21 20:54:11,014 - INFO - SUCCESS [bulk]: Client=127.0.0.1, File-National Apprenticeship Training Scheme (NATS).pdf, Words=['I'], Pages=[0],
32 2025-08-21 20:54:11,026 - INFO - SUCCESS [bulk]: Client=127.0.0.1, File-CT20182437032_Application.pdf, Words=['I'], Pages=[0], RemoveImages=True
33 2025-08-21 21:04:07,204 - INFO - SUCCESS [single]: Client=127.0.0.1, File-National Apprenticeship Training Scheme (NATS).pdf, Words=['I'], Pages=[0]
34 2025-08-21 21:04:54,307 - INFO - SUCCESS [bulk]: Client=127.0.0.1, File-National Apprenticeship Training Scheme (NATS).pdf, Words=['I'], Pages=[0],
35 2025-08-21 21:04:54,348 - INFO - SUCCESS [bulk]: Client=127.0.0.1, File-CT20182437032_Application.pdf, Words=['I'], Pages=[0], RemoveImages=True
36 2025-08-21 21:05:51,743 - ERROR - FAILURE [bulk]: Client=127.0.0.1, File=Eaadhaar xxxxxxxx8613_09102018235117_191143.pdf, Error=document closed or
37 2025-08-21 21:05:51,743 - ERROR - FAILURE [bulk]: Client=127.0.0.1, File=BULK, Error=No valid PDFs were processed
38 2025-08-21 21:06:19,219 - ERROR - FAILURE [bulk]: Client=127.0.0.1, File=Eaadhaar xxxxxxxx8613_09102018235117_191143.pdf, Error=document closed or
39 2025-08-21 21:06:19,372 - INFO - SUCCESS [bulk]: Client=127.0.0.1, File=admit ic 5th sem.pdf, Words=['I'], Pages=[0], RemoveImages=False
40 2025-08-21 21:06:34,057 - ERROR - FAILURE [bulk]: Client=127.0.0.1, File=Eaadhaar xxxxxxxx8613_09102018235117_191143.pdf, Error=document closed or
41 2025-08-21 21:06:34,233 - INFO - SUCCESS [bulk]: Client=127.0.0.1, File=admit ic 5th sem.pdf, Words=['I'], Pages=[0], RemoveImages=True
42

```

```

INFO: 127.0.0.1:49861 - "POST /redact_bulk/ HTTP/1.1" 200 OK
INFO: 127.0.0.1:49929 - "POST /redact_bulk/ HTTP/1.1" 400 Bad Request
INFO: 127.0.0.1:49966 - "POST /redact_bulk/ HTTP/1.1" 200 OK
INFO: 127.0.0.1:49989 - "POST /redact_bulk/ HTTP/1.1" 200 OK

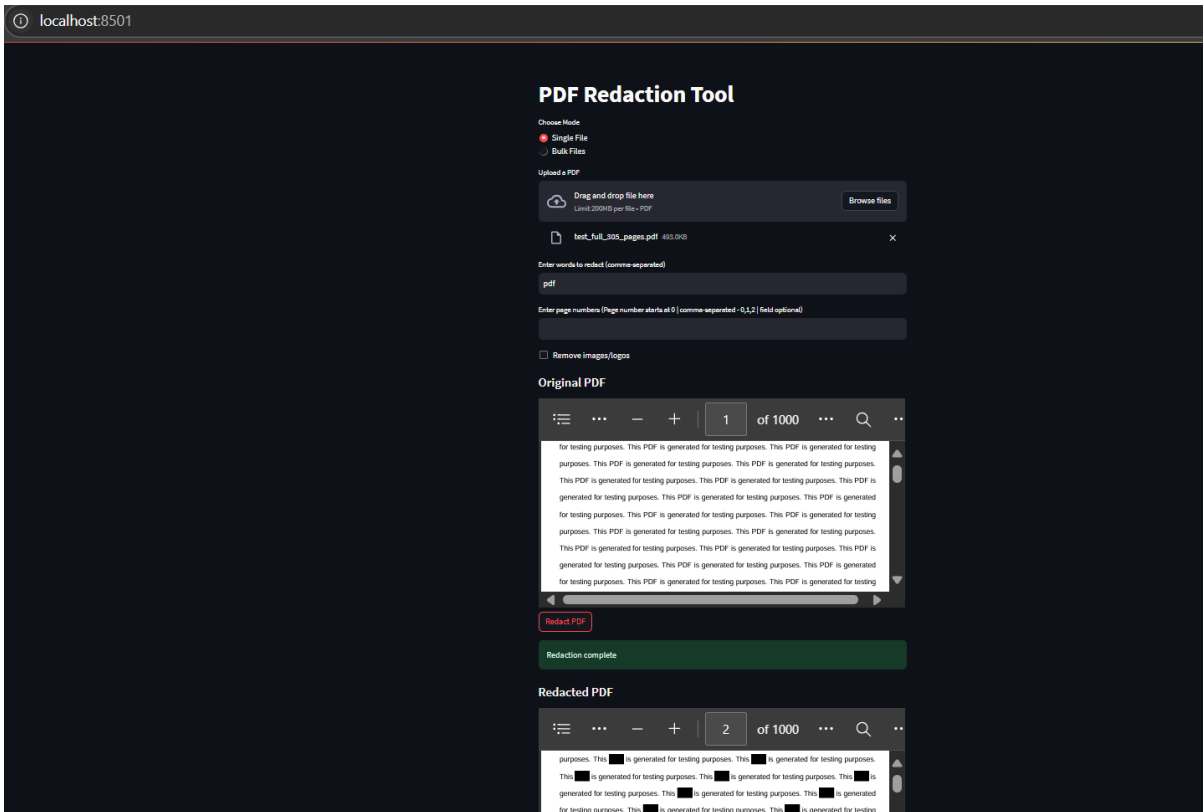
```

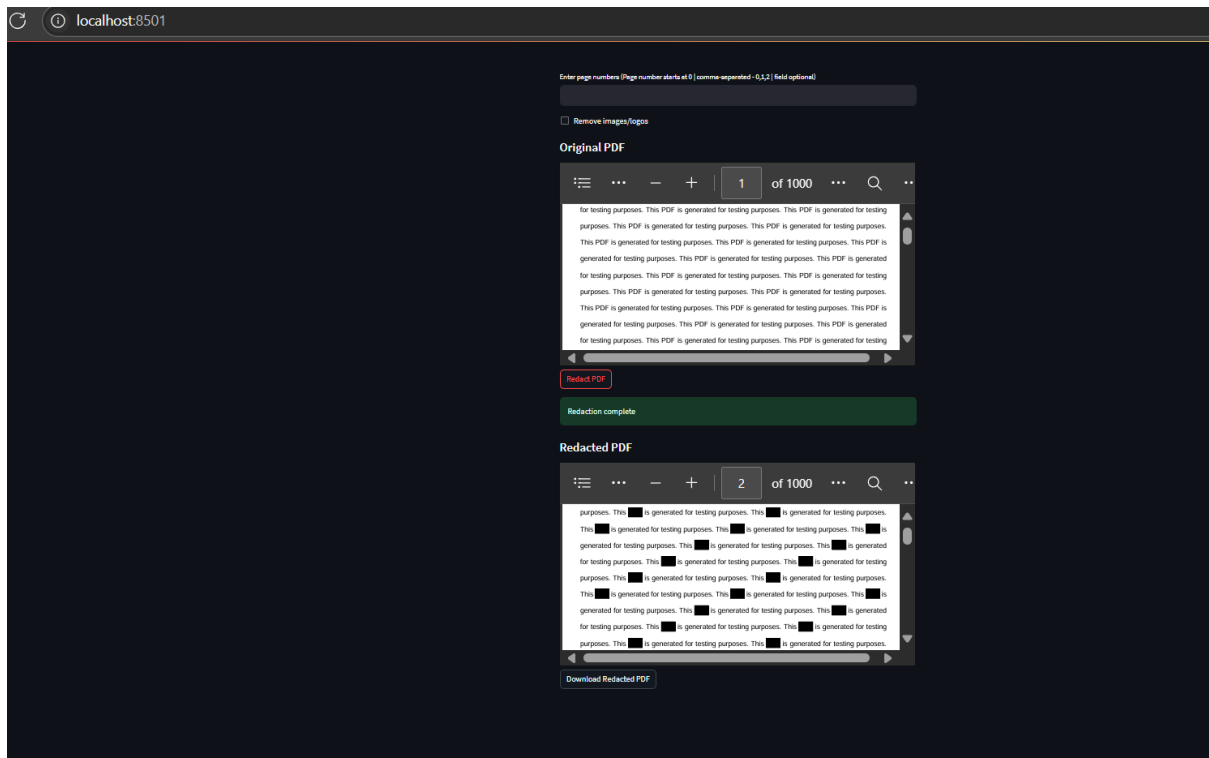
```

32 2025-08-21 20:54:11,026 - INFO - SUCCESS [bulk]: Client=127.0.0.1, File=CT20182437032_Application.pdf, Words=[ 'I' ], Pages=[0], RemoveImages=True
33 2025-08-21 21:04:07,204 - INFO - SUCCESS [single]: Client=127.0.0.1, File=National Apprenticeship Training Scheme (NATS).pdf, Words=[ 'I' ], Pages=[0]
34 2025-08-21 21:04:54,307 - INFO - SUCCESS [bulk]: Client=127.0.0.1, File=National Apprenticeship Training Scheme (NATS).pdf, Words=[ 'I' ], Pages=[0]
35 2025-08-21 21:04:54,348 - INFO - SUCCESS [bulk]: Client=127.0.0.1, File=CT20182437032_Application.pdf, Words=[ 'I' ], Pages=[0], RemoveImages=True
36 2025-08-21 21:05:51,743 - ERROR - FAILURE [bulk]: Client=127.0.0.1, File=EAdhaar xxxxxxxx8613_09102018235117_191143.pdf, Error=document closed or
37 2025-08-21 21:05:51,743 - ERROR - FAILURE [bulk]: Client=127.0.0.1, File=Bulk, Error=No valid PDFs were processed
38 2025-08-21 21:06:19,219 - ERROR - FAILURE [bulk]: Client=127.0.0.1, File=EAdhaar xxxxxxxx8613_09102018235117_191143.pdf, Error=document closed or
39 2025-08-21 21:06:19,372 - INFO - SUCCESS [bulk]: Client=127.0.0.1, File=admit ic 5th sem.pdf, Words=[ 'I' ], Pages=[0], RemoveImages=False
40 2025-08-21 21:06:34,057 - ERROR - FAILURE [bulk]: Client=127.0.0.1, File=EAdhaar xxxxxxxx8613_09102018235117_191143.pdf, Error=document closed or
41 2025-08-21 21:06:34,233 - INFO - SUCCESS [bulk]: Client=127.0.0.1, File=admit ic 5th sem.pdf, Words=[ 'I' ], Pages=[0], RemoveImages=True
42 2025-08-21 22:07:10,623 - ERROR - FAILURE [single]: Client=127.0.0.1, File=test_305_pages.pdf, Error=PDF exceeds 300 pages
43 2025-08-21 22:09:10,037 - INFO - SUCCESS [single]: Client=127.0.0.1, File=test_305_pages.pdf, Words=[ 'I' ], Pages=ALL, RemoveImages=False
44 2025-08-21 22:17:24,473 - INFO - SUCCESS [single]: Client=127.0.0.1, File=test_305_pages.pdf, Words=[ 'I' ], Pages=ALL, RemoveImages=False
45 2025-08-21 22:18:03,139 - ERROR - FAILURE [single]: Client=127.0.0.1, File=test_full_305_pages.pdf, Error=PDF exceeds 300 pages
46 2025-08-21 22:18:50,678 - ERROR - FAILURE [single]: Client=127.0.0.1, File=test_full_305_pages.pdf, Error=PDF exceeds 300 pages
47 2025-08-21 22:19:17,533 - ERROR - FAILURE [single]: Client=127.0.0.1, File=test_full_305_pages.pdf, Error=PDF exceeds 300 pages
48 2025-08-21 22:24:07,206 - ERROR - FAILURE [single]: Client=127.0.0.1, File=test_full_305_pages.pdf, Error=PDF exceeds 300 pages
49 2025-08-21 22:25:54,235 - INFO - SUCCESS [single]: Client=127.0.0.1, File=test_full_305_pages.pdf, Words=[ 'pdf' ], Pages=ALL, RemoveImages=False
50 2025-08-21 22:45:30,558 - INFO - SUCCESS [bulk]: Client=127.0.0.1, File=34.06_09.pdf, Words=[ 'this' ], Pages=ALL, RemoveImages=True
51 2025-08-21 22:45:31,233 - INFO - SUCCESS [bulk]: Client=127.0.0.1, File=drylab.pdf, Words=[ 'this' ], Pages=ALL, RemoveImages=True
52 2025-08-21 22:45:31,449 - INFO - SUCCESS [bulk]: Client=127.0.0.1, File=scientific.pdf, Words=[ 'this' ], Pages=ALL, RemoveImages=True
53 2025-08-21 22:45:31,524 - INFO - SUCCESS [bulk]: Client=127.0.0.1, File=somatosensory.pdf, Words=[ 'this' ], Pages=ALL, RemoveImages=True
54 2025-08-21 22:52:09,094 - ERROR - FAILURE [single]: Client=127.0.0.1, File=test_full_305_pages.pdf, Error=PDF exceeds 300 pages
55

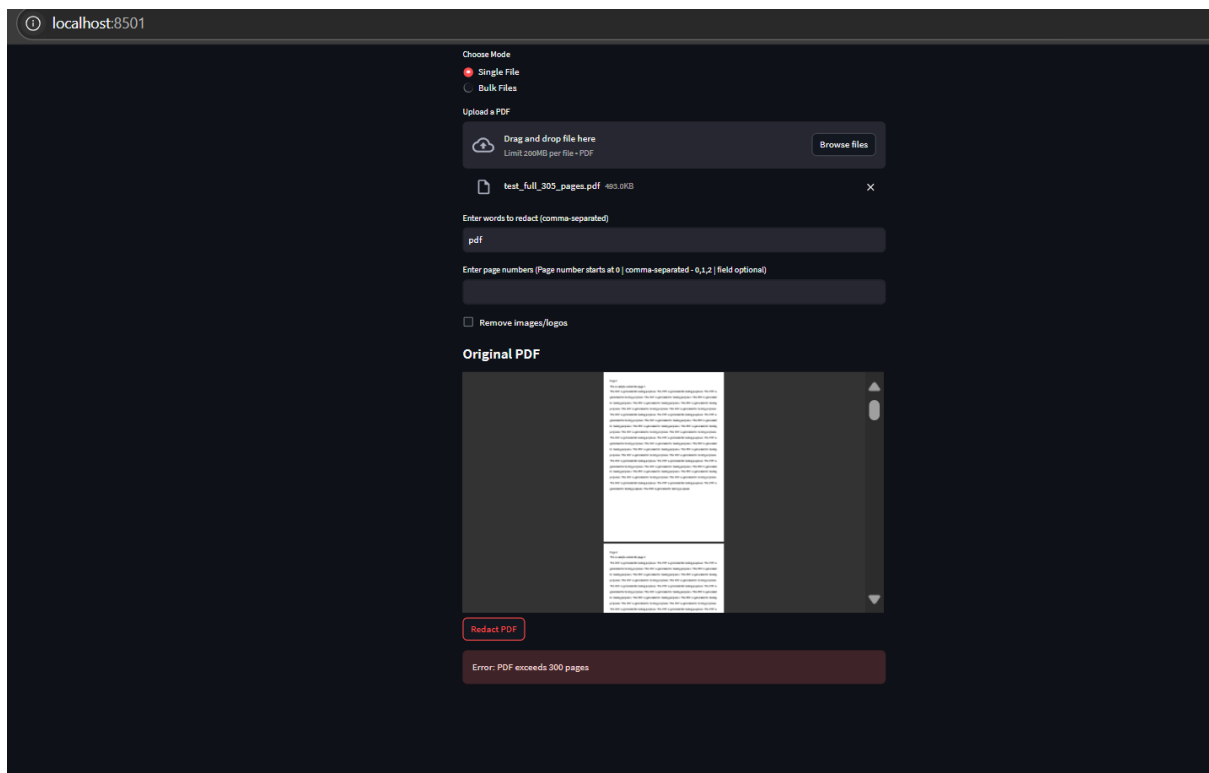
```

Keeping no limits ; testing with 1000 pages full text





Keeping 300 as pages limit as per POC



Bulk files

