

AI PDF Automation Toolkit

AI PDF Automation Toolkit

Filename: AI_PDF_Toolkit.zip

Download Link: https://freepdfreads.com/downloads/AI_PDF_Toolkit.zip (Gate behind email sign-up)

Toolkit Contents

1. GPT-4 PDF Summarizer

- gpt4_summarizer/
 - contract_analyzer.py (Extract key clauses from contracts)
 - research_paper_summarizer.py (Academic PDF summarization)
 - llama2_local/ (Free alternative with HuggingFace)

Code Snippet:

python Copy Download

```
# contract_analyzer.py import openai, fitz
```

```
def analyze_contract(pdf_path):
```

```
    doc = fitz.open(pdf_path)
```

```
    text = "\n".join([page.get_text() for page in doc])
```

```
    response = openai.ChatCompletion.create(
```

```
        model="gpt-4-1106-preview",
```

```
        messages=[{"role": "user", "content": f"Identify 3 critical risks in this contract:\n{text}"}])return response.choices[0].message.content
```

2. Custom ML Model Training

- ml_parsing/
 - spacy_ner_train.ipynb (Train PDF entity recognition models)
 - sample_training_data.json (Texttract-annotated PDF data)
 - pretrained_model/ (SpaCy model for "contract_term" detection)

Sample Training Data:

json Copy Download

```
{ "text": "This Agreement terminates on 2025-12-31.", "entities": [[25, 38, "TERM_DATE"], [52, 61, "CONTRACT_ACTION"]] }
```

3. Compliance Guardrails

- compliance/
 - pii_redactor.py (Presidio-based PII detection/redaction)
 - gdpr_validator.py (LangChain + LLM compliance checks)
 - regulatory_checklist.pdf (GDPR/CCPA rules for PDFs)

Code Snippet:

python Copy Download

```
# gdpr_validator.py from langchain.chains import LLMChain
```

```
from langchain.prompts import PromptTemplate
```

```
template = """Does this text violate GDPR Article 17 Right to Erasure?
```

```
{text}
```

```
Answer: [Yes/No] + 1-sentence reason."""def check_gdpr(text):
```

```
    prompt = PromptTemplate(template=template, input_variables=["text"])
```

```
    chain = LLMChain(llm=llm, prompt=prompt)return chain.run(text)
```

4. Self-Healing OCR

- self_healing_ocr/
 - keras_model.h5 (Pre-trained CNN-LSTM model)
 - ocr_correction.py (Fix scanned PDF errors)
 - sample_scans/ (Test images + ground truth text)

Architecture:

python Copy Download

```
# keras_model.py from tensorflow.keras.models import Sequential
```

```
from tensorflow.keras.layers import Conv2D, LSTM, Dense
```

```
model = Sequential([
```

```
    Conv2D(64, (3,3), activation='relu', input_shape=(128, 128, 1)),
```

```
    LSTM(128),
```

```
    Dense(94, activation='softmax') # 94 printable ASCII chars ])
```

5. Ethical AI Tools

- ethics/
 - bias_audit.ipynb (IBM AIF360 fairness checks)
 - shap_explainer.py (Model decision visualization)
 - gdpr_compliance_guide.pdf

Quick Start Guide

1. Install Dependencies:

bash Copy

```
pip install -r requirements.txt # Includes PyMuPDF, spaCy, langchain, etc.
```

2. Run GPT-4 Summarizer:

bash Copy

```
python gpt4_summarizer/contract_analyzer.py --input legal_doc.pdf
```

3. Train Custom Model:

bash Copy

```
jupyter notebook ml_parsing/spacy_ner_train.ipynb
```

Licensing & Terms

- Code: MIT License (modify/redistribute freely).
- Models: CC BY-NC 4.0 (non-commercial use).
- Data: Synthetic/fake data for testing.

Toolkit Value Proposition

- For Developers: Jumpstart AI-PDF projects with production-ready code.
- For Businesses: Reduce manual doc processing costs by 60-80%.
- For Researchers: Pre-trained models for academic/medical PDF parsing.

Promotion Strategy

1. Blog CTAs:

markdown Copy

```
> *** Ready to Implement AI PDF Automation?*** [Download Our Free Toolkit](#) (Includes GPT-4 scripts, ML models & templates).
```

2. Social Media Teaser:

Copy Download

"Automate contracts, redact PII, and fix OCR errors with AI – no premium software needed. Grab our free toolkit + [freePDFreads.com]"

3. Email Campaign:

html Copy

Run

```
<h2>New: AI PDF Automation Toolkit</h2><p>What's inside:</p><ul><li>GPT-4 contract analyzer</li><li>Self-healing OCR model</li><li>GDPR compliance validator</li></ul><a href="
```

Download Link: freepdfreads.com

