

Agentic Document Extraction



Turn messy, multi-format documents into structured, visually grounded data for your systems. Built for production at enterprise scale.

Enterprises depend on millions of unstructured documents — financial statements, medical forms, contracts, engineering drawings — that traditional OCR and LLM pipelines fail to handle. Without visual context or traceability, automation breaks, accuracy drops, and trust erodes.

Agentic Document Extraction (ADE) combines vision, reasoning, and validation to deliver enterprise-grade document intelligence. Powered by the Document Pre-trained Transformer (DPT-2), ADE interprets both what's written and how it's structured — grounding every extracted element visually and semantically for auditability and confidence.

Agentic Document Extraction delivers

80%

The average time reduction in model deployment

100%

Visual traceability for audit and compliance

50x

Up to 50x Faster time-to-value than legacy OCR or LLM stacks



Zero-Shot Parsing

Works instantly across any document type — no training or templates.



Visual Grounding

Every field is linked to its bounding box for verification and explainability.



Complex Layout Understanding

Handles multi-page tables, charts, handwriting, and embedded figures.



Agentic Orchestration

Plans, decides, and verifies extraction through multi-modal reasoning.



Instant Deployment

Available via API, SDK, or integrations with Snowflake, SAP, Databricks, Google BigQuery, AWS, and more.

“ADE has significantly outperformed other document extractors we’ve used. It has helped us build and Agentic RAG answer engine, based on unique healthcare institutional content, to offer instant, validated support to medical professionals at the point of care.”

”

Dr. Declan Kelly, Founder and CEO, Eolas Medical

Complex Layout Extraction

Zero-shot parsing of diverse document formats (PDFs, scans, tables) without requiring layout-specific training.

Extraction of Tables and Charts

Extracts data accurately from complex visual layouts, figures, and illustrations including hand-drawn and scanned documents.

Visual Grounding

Each text and visual element is grounded with location coordinates to enable answer verification and build trust.

Field Extraction

Field extract lets you quickly extract out the key fields from huge collections of documents for batch processing.

The image displays three screenshots of the Agentic Document Extraction interface. Each screenshot shows a document being processed, with a sidebar on the left containing navigation options like 'Files', 'Parse', 'Extract', 'Chat', and 'View Pricing'. The main area shows the extracted content, including text, tables, and charts. The first screenshot shows a document with a SWOT analysis and a table of financial data. The second screenshot shows a document with a DCF model and a table of financial data. The third screenshot shows a document with a DCF model and a table of financial data.

Integrate Quickly into Custom Pipelines and Applications

```
import os
from pathlib import Path
from landingai_ade import LandingAIADE
from landingai_ade.lib import pydantic_to_json_schema
from pydantic import BaseModel, Field

# Define your schema
class Person(BaseModel):
    name: str = Field(description="Person's name")
    age: int = Field(description="Person's age")
```

Enterprise Ready

Enterprise customers qualify for volume-based pricing. This tier also offers priority rate limits and additional options. Contact us to discuss your processing pipeline, security, technical, and compliance needs.

1

SOC 2 Type II

Independently audited controls for security, availability, and confidentiality.

2

Tenant Isolation

All environments are logically or physically isolated, preventing cross-customer data access.

3

HIPAA + BAA

Certified compliance for healthcare data with signed Business Associate Agreements.

4

Zero Data Retention

Optional zero data retention setting. No documents or outputs stored; all data erased after processing.

5

99%+ Uptime

Redundant, multi-AZ deployments ensure enterprise availability.

6

Trust Center

Real-time status, compliance reports, and security documentation.

7

GDPR Compliant

Data handling meets all EU GDPR requirements.

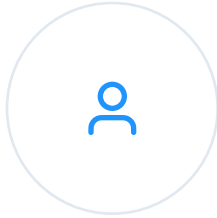
8

Security Controls

End-to-end API security, encrypted data in transit and at rest and more.

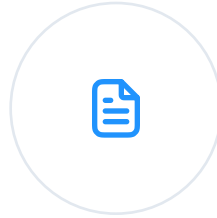
Get Started

Ready to bring intelligent document extraction to your workflows?



Create

Go to va.landing.ai to create your account and start exploring ADE.



Test

Upload your own documents or try sample files to see extraction in action.



Connect

Connect with our team to discuss your workflow and integration needs.



Customize

Tailor your plan for data storage, deployment type, tech stack, and usage volume.

About LandingAI

LandingAI is redefining intelligent document processing with Agentic Document Extraction (ADE), an **API-first Agentic Document Intelligence platform** that turns messy, multi-modal documents and dark data into **structured, auditable** intelligence.



Scan to watch a demo