



Financial Document Processing APIs at Scale

From parsing to intelligent workflows — process millions of documents with production-ready APIs.

Purpose-built for **Developers, CTOs, and Engineering Leads** in FinTech, LegalTech, InsurTech, and HealthTech. We provide the structural reasoning that standard OCR misses.

✓ No Training
Required

✓ Cross-Field
Validation

✓ >99% Validation
Rate

```
POST /v1/extract

// Response 200 OK
{
  "status": "success",
  "validation": {
    "passed": true,
    "confidence": 0.994
  },
  "data": {
    "invoice_id": "INV-2026-001",
    "total_amount": 4500.00,
    "line_items": [
      {
        "desc": "API Usage",
        "qty": 45000,
        "rate": 0.10
      }
    ]
  }
}
```

⚡ 82ms

🛡️ VALIDATION >99%



WHY NOW

Standard OCR Is **Broken** for Complex Documents

Your team spends weeks wrestling with tools that were never built for financial reasoning, multi-page tables, or handwritten fields.



Cloud OCR is Generic

AWS Textract and Google DocAI charge **\$50 per 1,000 pages** for table extraction — and still fail on unseen layouts, merged cells, and multi-page tables.

```
// Textract output
"total": null
"line_items": []
// Table spanning pg 2-3: MISSED
```

✘ Fails on complex layouts



LLMs Hallucinate Data

GPT-4o Vision looks promising in demos but **invents line items**, miscalculates totals, and has no built-in validation. One wrong number in finance = compliance disaster.

```
// GPT-4o extraction
"total": 4,850.00
"sum_of_items": 4,500.00
// $350 mismatch - NO WARNING
```

✘ No cross-field validation



Manual Annotation Tax

Tools like Nanonets force you to **label 50+ documents** to "train your own model" when their base model fails. You're not a data labeling shop — you're a builder.

```
// Nanonets response
"accuracy": 62%
"action_required":
  "Annotate 50 more samples"
```

✘ Weeks of labeling work

→ **HyperAPI: Zero-shot. Validated. Production-ready.**
No training. No hallucinations. No manual annotation.

“ I spent 3 weeks trying to get GPT-4o to extract line items consistently.”



WHAT WE PROVIDE

Engineered for Complex Data

Generic OCR stops at text recognition. HyperAPI begins where others fail — providing structural understanding, validation, and domain-specific reasoning for mission-critical documents.

```
</> schema_definition.json
```

```
"strict_mode": true,  
"validate_totals": true,  
"normalize_dates": true,  
"allow_handwriting": true
```



Validation-Aware Extraction

Automatically enforces document-level constraints. Line-item totals match summaries. Cross-field consistency is guaranteed.



No Training Required

Our vision-language models understand document semantics instantly, adapting to unseen layouts without labeling or fine-tuning.



Multi-Page Tables

Robust line-item capture that spans across pages. We maintain table structure integrity through headers, footers, and page breaks.



Handwriting Support

Process mixed inputs with ease. We handle cursive, printed text, stamps, and noisy scans with high accuracy across 100+ languages.



Configurable Fields

Define your extraction schema. Choose which fields to extract, set types, and apply per-field rules — all via the API.



Data Normalization

Standardized outputs: dates to ISO-8601, currency to ISO-4217, addresses to canonical format. Clean data, every time.



20 APIS

Building Blocks • Composite • Advanced

Production-Ready API Portfolio

● Available ● Waitlist ● Coming Soon



Building Blocks

Core document primitives

Parse

x-Parse

Parse Intent

Split

Layout Analysis

Table Detection

Classify

High-accuracy text extraction, multi-lingual parsing, document splitting, layout detection, and classification.



Composite APIs

Higher-level intelligence

Extract

Extract-Omni

Verify

Redact

Anonymise

Tabular Reasoning

Summarisation

Context Graphs

Structured extraction, validation, redaction, anonymisation, tabular reasoning, and knowledge graphs.



Advanced

Enterprise & specialized

Workflow

Similar-Queries

Sheets

Doc Preprocessing

Finance CLI

Chained pipelines, semantic search, spreadsheet conversion, image enhancement, and CLI tooling.

⚡ All APIs share the same auth, schema config, and billing model

apis.hyperbots.com



Three Steps to Structured Data

From raw document to validated JSON in under 3 seconds.

1

Upload

Send any document

- PDF, TIFF, JPEG, PNG
- Single or multi-page
- Optional: pass your schema

2

Extract & Validate

Vision-Language Model

- Layout-aware parsing
- Cross-field validation
- Auto-normalization

3

Receive JSON

Validated & Normalized

- Clean structured JSON
- Confidence scores
- Ready for your pipeline

extract.py

Python

REQUEST

```
import hyperapi

result = hyperapi.extract(
    file="invoice.pdf",
    schema={
        "vendor": "string",
        "date": "date",
        "total": "currency",
        "line_items": "table"
    }
)
```

RESPONSE 200 OK • 1.8S





```
{
  "vendor": "Amazon Inc.",
  "date": "2026-02-15",
  "total": { "amount": 4500.00, "currency": "USD" },
  "line_items": [
    { "desc": "API Usage", "qty": 45000, "rate": 0.10 }
  ],
  "validation": { "passed": true, "confidence": 0.997 }
}
```







Complete Feature Matrix

● Enterprise Ready





👁️ Extraction & Vision

-  **Handwritten Text Recognition** CORE
Cursive, printed, stamps, annotations on low-quality scans
-  **Complex Table Structures**
Nested headers, merged cells, multi-page tables
-  **100+ Languages** CORE
Latin, CJK, Arabic, Devanagari, and more
-  **Unlimited Page Count**
No page-per-document limits. Process 1 or 1,000 pages.





✂️ Data Normalization

-  **Date Standardization** NEW
Any format → ISO-8601 (YYYY-MM-DD)
-  **Currency Normalization** NEW
Amounts + ISO-4217 currency codes
-  **Vendor & Entity Matching**
"Amzn", "Amazon Inc" → single master entity
-  **Address Canonicalization**
Diverse formats → structured address components

⚙️ Field Configuration

-  **Custom JSON Schema** CONFIG
Define required/optional fields, types, and constraints
-  **Per-Field Validation Rules**
Regex patterns, enum constraints, numerical ranges
-  **Dynamic Field Selection**
Choose exactly which fields to extract per API call
-  **Nested & Repeating Groups**
Line items, sub-tables, and hierarchical data

🛡️ Quality & Operations

-  **Confidence Threshold Billing**
Set minimum scores. Below threshold = no charge.
-  **Human-in-the-Loop (HITL)**
Optional verified workflow for edge cases
-  **Auto-Retry with Credits**
Failed extractions retry free with alt models
-  **Async Batch Processing**
1k-10k pages per job, SLA-based completion



DEEP DIVE

Your Schema. Your Rules. Clean Data Out.

Configure exactly what to extract, and get normalized, pipeline-ready outputs every time.



Field Configuration

Define your extraction schema via the API

```

extraction_config.json Valid

{
  "fields": {
    "vendor_name": {
      "type": "string",
      "required": true,
      "normalize": "entity_match"
    },
    "invoice_date": {
      "type": "date",
      "format": "YYYY-MM-DD"
    },
    "total_amount": {
      "type": "currency",
      "currency_code": "USD",
      "validate": "sum_of_line_items"
    },
    "line_items": {
      "type": "table",
      "columns": ["description", "qty", "unit_price", "amount"],
      "validate_row_totals": true
    },
    "payment_terms": {
      "type": "enum",
      "values": ["Net 30", "Net 60", "Due on Receipt"]
    }
  }
}

```



Data Normalization

Consistent outputs regardless of input format

📅 DATES ISO-8601

Raw Input
 "Feb 15, 2026"
 "15/02/2026"
 "02-15-26"

→

Normalized
 "2026-02-15"

💵 CURRENCY ISO-4217

Raw Input
 "\$4,500.00"
 "USD 4500"

→

Normalized
 {"amount": 4500.00,
 "currency": "USD"}

🏢 VENDOR ENTITIES Entity Match

Raw Input
 "AMZN" / "Amazon Inc."
 "amazon.com"

→

Normalized
 "Amazon Inc."

CONFIG & NORMALIZE



Normalization is built into every extraction — not a separate product.

Enable with `"normalize": true` — zero extra cost on all tiers.

Included in all plans



P2P AUTOMATION

Procure-to-Pay • Same underlying tech powering production workflows

Proven Impact in High-Volume Workflows

STRAIGHT-THROUGH

≥80%

Processed fully autonomously across all customers

VALIDATION RATE

>99%

Confidence-based validation pass rate

TIME REDUCTION

90%

From ~10 min/invoice to <60 seconds

COST SAVINGS

80%

\$9.87 manual → <\$2.00 automated



Series B FinTech

Lending & KYC Platform

Challenge: Processing 20k+ bank statements monthly for underwriting. GPT-4o hallucinated transaction amounts; Textract missed multi-page tables.

Solution: HyperAPI with custom schema for transaction tables, validated totals, and normalized dates. Integrated in 2 days.

99.4%

Extraction accuracy

2 days

Integration time



Mid-Market Insurer

Claims Processing

Challenge: 8k claims/month with handwritten annotations, stamps, and inconsistent form layouts. Nanonets required 50+ labeled samples per form type.

Solution: Zero-shot extraction with HITL fallback for low-confidence handwriting. No training needed.

≥80%

Straight-through rate

0

Training samples needed



Logistics Provider

Bill of Lading Automation

Challenge: 15k BOLs/month from 200+ carrier formats. Manual data entry team of 12. DocAI failed on non-standard layouts.

Solution: Configurable field extraction adapts to each carrier format. Vendor normalization maps all carrier names to master records.

90%

Time reduction

200+

Formats handled

i Case studies represent aggregated P2P customer outcomes using the same underlying HyperAPI technology. Cost benchmarks from Ardent Partners 2024.



USAGE-BASED PRICING

Predictable Costs. Scale on Demand.

Start free, upgrade for volume. No hidden fees. Normalization & field config included in all plans.

BASIC

\$0 / month

Test & prototype. No credit card.

- ✓ 100 pages / month
- ✓ 1 RPM Rate Limit
- ✓ Field Configuration
- ✓ Data Normalization
- ✓ Community Support

Overage: \$1.00 / page
HITL: Not available

Get API Key

MOST POPULAR

PRO

\$0.05 / page

For startups & growth teams.

- ✓ 10,000 pages included
- ✓ 60 RPM Rate Limit
- ✓ Retry Credits
- ✓ Confidence Scoring
- ✓ Email Support (24h)

Overage: **\$0.10 / page**
HITL (Optional): **+\$0.75 / page**

Start Free Trial

ENTERPRISE

\$2k / month

Platform fee + usage. Annual discounts.

- ✓ 40,000 pages included
- ✓ 1,000+ RPM Custom Limit
- ✓ Dedicated Slack Channel
- ✓ SLA & Security Review
- ✓ VPC / Private Link Deploy

Overage: **\$0.05 / page**
HITL (Optional): **+\$0.25 / page**

Contact Sales

ⓘ All plans include field configuration, data normalization, and confidence scoring. Volume discounts available for annual contracts.



Performance Commitments

Enterprise Pool

Dedicated & Always-Warm

P95 Latency **< 2s / page**

Min replicas = 1 (always warm)

- 1,000+ RPM (Custom)
- Private Link / VPC
- SLA Guarantees

Pro Pool

Burstable Shared Capacity

Avg Latency **2-4s / page**

Burst: 2x limit for 10s

- 60 RPM (Burstable)
- Async Batch Support
- Priority Queue

Free Pool

On-Demand Cold Start

Cold Start **~15s first request**

Scale-to-zero, then ~2-4s/page

- 1 RPM Limit
- Shared Resources
- Community Support

Async Batch Processing

High-volume backlog ingestion

BATCH SIZE

1k-10k pages

COMPLETION

SLA-Based

MARGINS

Best Price

Operational Guarantees

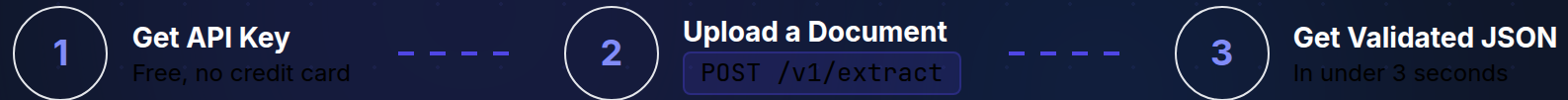
Risk-free integration

- Confidence Threshold Billing**
Below threshold = no charge.
- Auto-Retry Policy**
Failed extractions retry free with alt models.
- Status Page & Monitoring**
Real-time uptime at status.hyperbots.com.




Ready to Build?


Stop wrestling with OCR. Get validated, normalized JSON from any document in minutes.





 [Get Free API Key](#)

 [Try Live Demo](#)

 [Read Docs](#)

 apis.hyperbots.com

 apis.hyperbots.com/docs

 research@hyperbots.com

"Stop parsing strings. Start getting data."