

AI Document Parsing Engine MVP

Comparable to: AI Document Parsing

An enterprise AI web application that extracts structured data from invoices and receipts.

Estimated MVP cost

\$16,200

Estimated effort

286 hours

To get your own customised cost for an app like this, visit www.chatstack.app and start a ChatStack interview describing how you want it customised.

Key Features

- PDF and image upload
- Asynchronous processing queue with document status and in-app completion notifications
- AI extraction using provider-configured preset invoice and receipt templates
- Mandatory human review and correction before export
- CSV and JSON export
- Monthly subscription billing with document-volume tiers
- Document and extracted-data retention until the customer deletes them or closes the account

Out of Scope for This MVP

- US hosting
- US payment and compliance requirements
- English-only MVP
- Simple defaults should be used for non-essential functionality
- Customers cannot create or edit extraction schemas in the MVP
- Every extracted document must be reviewed before it can be exported

Full Cost Breakdown

Item	Hours	Cost
The system shall allow visitors to create accounts using email-based registration and secure account-creation workflows.	5	\$280
The system shall allow registered users to sign in securely and maintain authenticated sessions for accessing their account and documents.	5	\$280
The system shall allow authenticated customers to upload supported PDF and image documents for processing.	8	\$440
The system shall process uploaded documents asynchronously and display queued, processing, completed, or failed statuses for each document.	9	\$500
The system shall provide an in-app notification when document processing is completed or fails.	5	\$280
The system shall extract invoice and receipt data from uploaded documents using provider-configured extraction templates.	16	\$880
The system shall allow authorized reviewers to view extracted document data and the associated source document before export.	10	\$550
The system shall allow authorized reviewers to correct extracted field values within the configured invoice or receipt template.	8	\$440
The system shall allow customers to export reviewed document data in CSV or JSON format.	7	\$390
The system shall allow customers to view their active subscription and document-volume tier.	4	\$220
The system shall allow customers to manage their subscription, including applicable plan changes and billing management.	12	\$660
The system shall retain customer documents and extracted data while the account remains active and make them available for review and export.	6	\$330

The system shall allow customers to delete documents and associated extracted data.	6	\$330
The system shall allow customers to delete their account and associated data where legally permitted.	9	\$500
The system shall allow internal administrators to configure the invoice and receipt extraction templates available to customers.	12	\$660
Planning	32	\$1,760
Project Management	25	\$1,380
Testing	15	\$830
DevOps	12	\$660
Design & Frontend Build	38	\$2,090
AI Tokens	0	\$430
Security	15	\$830
Risk (10% Contingency)	27	\$1,490
Total	286	\$16,200

Estimate Assumptions

- Estimates cover application development and direct service integrations, but exclude project management, QA overhead, infrastructure operations, and NFR work.
- Supabase Auth, PostgreSQL, object storage, and Row Level Security are available and used as specified.
- A single selected extraction provider and a

single
selected
payment
provider
are implem
ented for
the MVP.

- Invoice
and receipt
templates
use a
bounded
field model
rather than
a fully
general
schema-
builder
platform.

- Reviewers
and adminis
trators use
predefined
roles; no
custom
permission-
manageme
nt interface
is included.

- Account
and
document
deletion
follows a
defined
legal-
retention
policy and
does not
require
complex
legal-hold a
dministratio
n.

- Export
supports
practical
MVP
selection or
eligible-
record

scope without advanced scheduling or recurring exports.

Risks That Could Change This

- Extraction-provider response formats, confidence data, and template capabilities may require additional mapping effort.
- Payment-provider plan changes, webhooks, tax handling, and billing edge cases are not fully defined.
- Large files, multi-page documents, and provider latency may require additional queue and retry engineering.
- Deletion behavior

across
backups,
legal holds,
and third-
party
storage
may
constrain
immediate
hard
deletion.

- Reviewer workload, concurrency rules, and audit requirements could expand the review workflow.

User Stories

US-1 - Visitor

As a visitor,
I want to
create an
account so
that I can
use the
document
extraction
service.

US-2 - Registered User

As a
registered
user, I want
to sign in
securely so
that I can
access my
account
and
documents.

US-3 - Logged In Customer

As a
logged-in

customer, I want to upload PDF and image documents so that they can be processed for data extraction.

**US-4 -
Logged In
Customer**

As a logged-in customer, I want to see the processing status of each uploaded document so that I know when extraction is complete.

**US-5 -
Logged In
Customer**

As a logged-in customer, I want to receive an in-app notification when document processing is complete so that I can review the extracted data promptly.

**US-6 -
Logged In
Customer**

As a logged-in

customer, I want invoice and receipt data extracted using configured templates so that uploaded documents become structured records.

**US-7 -
Logged In
Reviewer**

As a logged-in reviewer, I want to review extracted document data so that I can verify its accuracy before export.

**US-8 -
Logged In
Reviewer**

As a logged-in reviewer, I want to correct extracted values so that the final record accurately represents the source document.

**US-9 -
Logged In
Customer**

As a logged-in customer, I

want to
export
reviewed
document
data as
CSV or
JSON so
that I can
use it in do
wnstream
business
processes.

**US-10 -
Logged In
Customer**

As a
logged-in
customer, I
want to
view my sub
scription
and
document-
volume tier
so that I
understand
my billing
plan.

**US-11 -
Logged In
Customer**

As a
logged-in
customer, I
want to
manage my
subscription
so that I
can
maintain
the plan
appropriate
for my
document
volume.

**US-12 -
Logged In
Customer**

As a
logged-in
customer, I
want my
documents

and
extracted
data
retained
while my
account is
active so
that I can
review and
export
them over
time.

**US-13 -
Logged In
Customer**

As a
logged-in
customer, I
want to
delete
documents
and
extracted
data so
that I can
control
retention of
my
business
information.

**US-14 -
Logged In
Customer**

As a
logged-in
customer, I
want to
delete my
account so
that my
relationship
with the
service and
associated
data can
be closed.

**US-15 -
Internal Ad
administrator**

As an
internal ad
administrator,
I want to

configure
available
invoice and
receipt
extraction
templates
so that
customers
can use
supported
document
formats
consistently.

Require ments

FR-1 - Account R egistration

The system
shall allow
visitors to
create
accounts
using email-
based
registration
and secure
account-
creation
workflows.

FR-2 - Auth entication

The system
shall allow
registered
users to
sign in
securely
and
maintain aut
henticated
sessions
for
accessing
their
account
and
documents.

FR-3 - Document Upload

The system shall allow a uthenticated customers to upload supported PDF and image documents for processing.

FR-4 - Processing Status

The system shall process uploaded documents asynchronously and display queued, processing, completed, or failed statuses for each document.

FR-5 - Notifications

The system shall provide an in-app notification when document processing is completed or fails.

FR-6 - Data Extraction

The system shall extract invoice and receipt data from uploaded documents

using provider-configured extraction templates.

FR-7 - Document Review

The system shall allow authorized reviewers to view extracted document data and the associated source document before export.

FR-8 - Data Correction

The system shall allow authorized reviewers to correct extracted field values within the configured invoice or receipt template.

FR-9 - Data Export

The system shall allow customers to export reviewed document data in CSV or JSON format.

FR-10 - Subscription Visibility

The system shall allow customers to view their active subscription and document-volume tier.

FR-11 - Subscription Management

The system shall allow customers to manage their subscription, including applicable plan changes and billing management.

FR-12 - Data Retention

The system shall retain customer documents and extracted data while the account remains active and make them available for review and export.

FR-13 - Document Deletion

The system shall allow customers to delete documents and associated

extracted
data.

**FR-14 -
Account
Deletion**

The system
shall allow
customers
to delete
their
account
and
associated
data where
legally
permitted.

**FR-15 -
Template A
dministrati
on**

The system
shall allow
internal ad
ministrators
to
configure
the invoice
and receipt
extraction
templates
available to
customers.

**NFR-1 -
Security**

The system
shall
protect auth
entication
credentials,
sessions,
uploaded
documents,
and
extracted
data using
industry-
standard
security
controls.

**NFR-2 - Aut
horization**

The system
shall

enforce
role-based
access so
customers,
reviewers,
and
internal ad
ministrators
can access
only
permitted
functions
and data.

NFR-3 - Performance

The system
shall target
API
response
times
below 200
milliseconds
for routine
synchronou
s requests,
excluding a
synchronou
s
document-
processing
time.

NFR-4 - Scalability

The system
shall use a
managed
database
platform
capable of
scaling with
customer,
document,
and
extraction-
record
volume.

NFR-5 - Reliability

The system
shall
preserve
document-

processing
state and
expose
failed
processing
outcomes
without
silently
losing
uploaded
documents
or
extracted
records.

NFR-6 - Observability

The system shall provide application performance monitoring and operational visibility for key user and processing flows.

NFR-7 - Privacy and Compliance

The system shall support applicable US privacy and compliance obligations for account, document, extracted-data retention, and deletion workflows.

Require

ments Assumptions

- The MVP is primarily a web-based service; mobile app-store distribution is not currently specified.
- No external integrations are included unless explicitly stated, including payment gateways, OCR providers, storage providers, and downstream systems.
- The service supports English-language business documents and common PDF and image formats in the MVP.
- Document processing is asynchronous and may depend on

a provider-selected extraction or OCR service.

- Invoice and receipt templates are configured and maintained by internal administrator s or service providers.

- Customers have access only to their own documents and extracted data; reviewers and administrators receive role-appropriate access.

- Subscriptions are monthly and use predefined document-volume tiers.

- Export supports CSV and JSON for selected or otherwise eligible reviewed records.

- Hard deletion applies

where legally permitted; legal holds, backups, and statutory retention may delay or exclude deletion.

- Routine API performance targets exclude document-upload transfer time and asynchronous extraction duration.

Technical Specification

TS-1 - Web Application

Build a responsive web application using Next.js, React, TypeScript, Tailwind CSS, and shadcn/ui for customer and reviewer workflows.

TS-2 - Authentication

Use Supabase Auth for email registration,

email verification, password recovery, secure sign-in, JWT-based sessions, and session lifecycle management.

TS-3 - Authorization and RBAC

Implement role-based access control for customers, reviewers, and internal administrators, with tenant isolation enforced at the database and API layers.

TS-4 - Backend and API

Implement a REST-oriented backend using Supabase-managed services, PostgreSQL access, and optional edge functions for asynchronous or secured server-side workflows.

TS-5 - Data Storage

Use Supabase managed PostgreSQL for accounts, roles, documents, processing states, extraction records, templates, corrections, notifications, subscriptions, and audit metadata.

TS-6 - Document Storage and Processing

Store uploaded PDFs and images in managed object storage and process them asynchronously through a worker or edge-function workflow with persisted queued, processing, completed, and failed states.

TS-7 - Extraction and Review

Integrate provider-configured invoice and receipt templates with an extraction workflow that presents source documents and extracted fields for reviewer validation and correction.

TS-8 - Customer and Reviewer UI

Provide authenticated screens for document upload, document lists, processing status, in-app notifications, extraction review, corrections, and CSV or JSON export.

TS-9 - Administration UI

Provide a protected Next.js administrative panel for internal administrators to manage invoice and receipt extraction templates and associated configuration.

TS-10 - Subscriptions and Billing

Provide subscription and document-volume tier visibility, plan management, and billing-management integration through a selected payment provider compatible with US web subscriptions.

TS-11 - Hosting and Environments

Host the web application and admin panel on Vercel, use Supabase for managed backend infrastructure, and separate development, staging, and production environments.

TS-12 - CI/CD and Testing

Use GitHub Actions for validation and deployment pipelines with automated unit, integration, and end-to-end tests before production release.

TS-13 - Security

Apply HTTPS with TLS 1.2 or later, secure password storage through Supabase Auth, encrypted storage and transport, least-privilege access, secret management, input validation, and tenant isolation.

TS-14 - Privacy and Data Lifecycle

Implement document and extracted-data retention, customer-initiated deletion, account deletion, and provider-controlled cleanup workflows subject to legal holds, backups, and statutory requirements.

TS-15 - Notifications

Implement persisted in-app notifications for completed and failed document processing, linked to the relevant customer document and status.

TS-16 - Observability and Logging

Implement centralized application and processing logs, error tracking, performance monitoring, audit events for privileged actions, and operational dashboards for key workflows.

TS-17 - Reliability and Operations

Ensure uploaded documents and processing state are durable, expose failures to users, support controlled retries, and prevent duplicate processing or duplicate exports.

Technical Assumptions

- The MVP is a responsive web service; Flutter mobile applications and app-store distribution are excluded.
- Supabase is the baseline managed backend platform, including PostgreSQL, authentication, and row-level security.
- Next.js, React, TypeScript, Tailwind CSS, and shadcn/ui are the baseline web technologies.
- Vercel hosts the web and administrative panels, while GitHub Actions manages CI/CD pipelines.
- Document processing is asynchronous and may use a provider-selected OCR or extraction service; Google Cloud Vision is a reference option.
- The MVP supports English-language business documents and common PDF and image formats.
- Invoice and receipt templates are configured only by internal administrators or service providers.
- Customers can access only their own documents and extracted data; reviewers and administrators receive

role-appropriate access.

- Subscriptions are monthly and use predefined document-volume tiers; the payment provider is not yet fixed.
- Export supports CSV and JSON for selected or otherwise eligible reviewed records.
- Hard deletion applies where legally permitted; backups, legal holds, and statutory retention may delay or exclude deletion.
- Routine API performance targets exclude file-transfer time and asynchronous extraction duration.
- No external integrations are included unless required for authentication, billing, document extraction, hosting, or explicitly approved MVP functionality.

To get your own customised cost for an app like this, visit <https://www.chatstack.app> and start a ChatStack interview.

This is a starting estimate for a custom-built MVP, designed and built from scratch. You own all the code and IP with no platform lock-in. AI Document Parsing is a trademark of its owner; ChatStack / App Developer Studio is not affiliated with, endorsed by, or sponsored by AI Document Parsing. © 2026 ChatStack.

Created with ChatStack (www.chatstack.app)