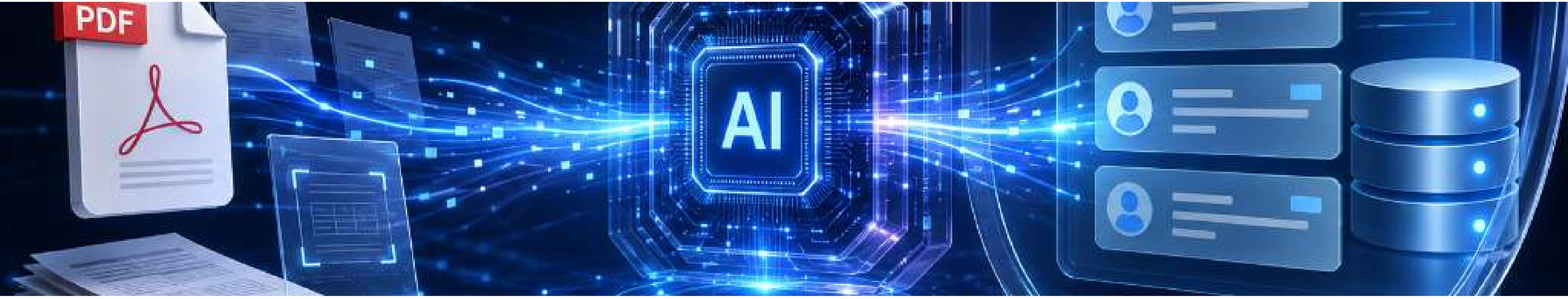
WHITE PAPER｜生成AI × 書類データ化 × CRM連携

PDF・スキャン書類を生成AIで安全にデータ化する方法

個人情報を守るクローズド環境の構築と CRM自動連携・API料金

想定読者：経営者／DX推進担当者／情報システム担当者／営業企画／CRM管理者

対象書類：申込書、契約書、保険証券、本人確認書類、請求書、アンケート、商談資料など



Start Challenge Consulting co., ltd.

発行：2026年8月

Confidential — 社外秘

紙・PDFに残る情報を、生成AIで「使えるデータ」に変える

顧客情報は紙・PDF・画像に分散し、OCR単独では項目の意味を判断できない。生成AIによる項目抽出と、個人情報を前提とした環境設計を同時に満たすことが出発点となる。

現状（As-Is）

01 顧客情報が紙・PDF・画像・メール添付に分散している

02 担当者が書類を目視で確認し、CRMへ手作業で転記している

03 OCRだけでは、項目の意味や文脈まで判断できない

生成AIによる変化

レイアウトが定まらない非定型書類からでも、必要な項目を文脈に基づいて抽出できる

申込書・契約書・保険証券・請求書など、書式が統一されていない書類群を同一の仕組みで処理できる可能性がある。

本資料の前提

個人情報を扱う場合は、通常のAI利用とは異なる環境設計が必要になる

「AIで読み取れるか」ではなく「どの情報を、どこまで外部へ出さずに処理できるか」が設計上の論点となる。

本資料では、PDF・スキャン書類を対象に、クローズド環境での生成AI活用、CRMへの自動連携、必要となる開発費・API料金・体制・スケジュールを一連の設計として整理する。

注記：事実＝書類が紙・PDF・画像に分散している状況／解釈＝OCR単独では項目の意味理解が不足する／提案＝生成AIの項目抽出と閉域設計を併せて検討する

PDF・スキャン書類のCRM登録を阻む5つの壁

障壁は「読み取り精度」だけではない。書類の不統一・OCR誤りに加え、個人情報の外部送信制約とCRM入力規則との不一致が、自動化の到達点を決めている。

壁 1｜書式	壁 2｜記載形式	壁 3｜読み取り精度	壁 4｜個人情報	壁 5｜CRM整合
書類ごとにレイアウトが異なる	**手書き・押印・表・チェック欄が含まれる**	**OCR結果に誤字や項目ずれが発生する**	**個人情報を外部AIへ送信できない**	**AIの出力とCRMの項目・入力規則が一致しない**
帳票ごとの個別対応が前提となり、テンプレート方式では網羅できない。	文字列以外の情報（選択・印影・罫線）の解釈が必要になる。	誤りを検知できないまま後工程へ流れると、修正コストが増大する。	送信可否の判断が社内で確定せず、検討が停滞しやすい。	選択肢・文字数・必須項目の制約に合わせた変換が必要になる。

現行業務（As-Is）｜書類受領からCRM登録までの流れ



■ 手作業（ベージュ） ■ システム処理（グレー） □ 赤枠＝課題箇所（人手依存により処理量に比例して工数が増加）

現行業務では、確認と転記が人手に依存しているため、書類量の増加がそのまま工数の増加につながる。5つの壁のうち、壁1～3は精度設計、壁4は環境設計、壁5は連携設計で扱う領域が異なる。

注記：事実＝5つの障壁は書類データ化の一般的な制約／解釈＝壁の性質により対処領域（精度／環境／連携）が分かれる／提案＝領域ごとに設計を分けて検討する

業務システム化では、ChatGPT画面とAPI連携を分けて考える

画面操作では処理ルールと出力形式が利用者に依存する。CRM登録を業務として成立させるには、ルールをシステム側に固定できるAPI連携が前提となる。

比較項目	ChatGPTへの手動アップロード	APIによるシステム連携
ファイル登録	人がアップロード	システムが自動送信
CRM登録	手作業	APIで自動登録
処理ルール	利用者に依存	システム側で固定
出力形式	会話・文章中心	JSON形式に固定可能
ログ管理	組織設定に依存	自社側でも管理可能
大量処理	不向き	キュー・バッチ処理が可能

PoC段階ではChatGPT画面を利用できても、本番業務ではAPIを利用したサーバー間連携が基本となる。

画面操作を前提とした検証結果は、そのまま本番の処理能力・ログ要件の裏付けにはならない。

検証の目的を「AIが読めるかの確認」に限定するならば画面操作でも足りる。一方、処理件数・監査ログ・出力形式の固定が要件に含まれる場合は、当初からAPI前提で設計する必要がある。

注記：事実＝画面操作とAPI連携では処理主体・出力形式・ログ管理の所在が異なる／解釈＝本番業務要件は画面操作では満たしにくい／提案＝本番前提の検証はAPIで実施する

「クローズド環境」は1種類ではない

クローズド環境は「企業向けクラウド」「ネットワーク制限」「完全オフライン」の3段階に整理できる。どの段階を求めるかで、構成・精度・コストの前提が大きく変わる。

レベル 1｜企業向けクラウド利用

組織アカウントと権限・ログ管理で統制する

- 組織単位のアカウント
- ユーザー・権限管理
- APIキー管理
- 監査ログ
- データ保存期間の設定

レベル 2｜ネットワーク制限型

通信経路と通信先を限定して処理する

- VPC・VNet内で処理
- Private Endpoint
- IPアドレス制限
- 外部通信先の限定
- Azure OpenAIなどの利用

レベル 3｜完全オフライン型

データを外部へ送信せず自社内で完結させる

- インターネットへデータを送信しない
- OCR・個人情報処理・LLMをオンプレミスで実行
- クラウド版OpenAI APIは直接利用しない
- ローカルLLMまたは専用環境を利用

レベルが上がるほど外部送信リスクは下がるが、モデル選択の自由度・精度・運用負荷の制約は強くなる。全社一律で決めず、書類種別と個人情報の重要度ごとに適用レベルを分けることが現実的である。

注記：事実＝クローズド化の手段は統制・通信制限・オフラインの3系統／解釈＝レベル上昇に伴い自由度と運用負荷が変化／提案＝書類種別ごとに適用レベルを設定する

個人情報の重要度に応じて構成を選択する

クラウドAPI型は導入が容易だが原本を外部送信する。完全閉域型は安全性が最も高い一方で精度がモデルに依存する。両者の中間にあるマスキング型が均衡点となる。

方式	AIへ送るデータ	セキュリティ	精度	導入難易度
クラウドAPI型	原本または抽出データ	中〜高	高	低
マスキング型	匿名化後のデータ	高	高	中
完全閉域型	外部送信なし	**非常に高い**	モデル次第	高

推奨方針

社内環境でOCRと個人情報マスキングを実施し、必要最小限の情報だけを生成AI APIへ送信する方式が現実的である。

この方式が均衡点となる理由

個人を特定できる情報を社内に留めたまま、文脈理解が必要な処理のみを高精度モデルに委ねられるため、安全性と精度を同時に確保しやすい。

完全閉域型を選ぶ場合の留意点

利用可能なモデルが限定されるため、精度検証を先行させる必要がある。基盤構築・運用の負荷も増加するため、対象書類を絞って開始する。

注記：事実＝3方式は外部送信データの範囲が異なる／解釈＝安全性・精度・難易度のトレードオフが生じる／提案＝マスキング型を基本構成とし、重要度の高い書類に閉域型を部分適用する

PDF受領からCRM登録までの全体像

全体は12工程で構成される。生成AIが担うのは第7工程のみであり、前後の受付・マスキング・検証・登録を社内側に置くことが安全性と精度の前提となる。

① 受付 社内環境	**1. PDF・画像を受領** メール添付、スキャナ、アップロード画面から取得する	**2. ウイルスチェック** 不正ファイルを入口で遮断・隔離する	**3. 社内ストレージへ一時保存** 原本IDと受付日時を付与する	**4. OCR・画像解析** 文字・表・チェック欄を機械可読化する
② 解析・保護 社内環境	**5. 個人情報を検出** 氏名・住所・番号類を自動判定する	**6. 個人情報をマスキング** 対応表は社内環境のみに保存する	**7. 生成AIで必要項目を抽出** 項目抽出・分類・要約を実行する	**8. JSON形式で結果を返却** CRM項目に対応したキーで受け取る
③ 抽出・登録 AI＋CRM	**9. 業務ルールで検証** 必須・形式・マスタ整合を機械判定する	**10. 担当者が内容を確認** 条件に該当する案件のみ人が確認する	**11. CRM APIで登録** 実データへ復元してから送信する	**12. 一時ファイルを削除** 保存期限に従い自動削除する

■ 手作業・人の関与（ベージュ） ■ システム処理（グレー） ■ 生成AI処理（青） ■ 登録・確定処理（濃紺） □ 赤枠＝閉域設計上の要点

第6工程のマスキングが外部送信の境界線となる。ここを社内環境に置くことで、レベル2相当の構成でも個人情報そのものは外部へ出ない設計が可能になる。

注記：事実＝12工程のうちAI処理は第7工程／解釈＝マスキング工程が外部送信の境界／提案＝境界工程を社内環境に固定して設計する

入口でファイルの安全性と品質を確保する

読み取り不能なファイルや重複ファイルを受付段階で排除することで、後工程の再処理とAPI費用の無駄を抑制できる。受付は精度対策の一部でもある。

受付条件の定義

・対応形式：PDF、JPEG、PNG、Word、Excel
・ファイルサイズ・ページ数の上限設定
・暗号化・パスワード付きPDFへの対応方針
・想定外形式の受付拒否と通知

安全性の確保

・ウイルス・マルウェアチェック
・ファイルハッシュによる二重登録防止
・読み取り不能ファイルの隔離
・隔離時の担当者通知と再提出依頼

追跡可能性の付与

・原本ID・受付日時の付与
・処理ステータスの管理
・一時保存領域と保存期限の設定
・原本とCRM登録結果の紐付け

受付処理の判定フロー



受付時に原本IDを付与しておくことで、CRM登録後に「どの書類のどのページが根拠か」を追跡できる。監査対応と誤登録調査の双方で必要となる要件である。

注記：事実＝受付段階で形式・安全性・重複を判定できる／解釈＝入口の排除が後工程の再処理コストを左右する／提案＝原本IDによる追跡を必須要件とする

OCRと生成AIを組み合わせて精度を高める

OCRは文字の読み取りに強く、生成AIは項目の意味理解に強い。両者は代替関係ではなく、読み取りと解釈を分担させることで精度と検証可能性が両立する。

処理内容	OCR	生成AI
文字の読み取り	得意	可能
手書き文字	製品により対応	画像から判断可能
項目の意味理解	苦手	得意
表・チェック欄の理解	一部可能	得意
要約・分類	不向き	得意
CRM形式への変換	ルールが必要	文脈から判断可能

推奨構成｜読み取りと解釈を分離する

OCR
文字を読み取り、位置情報付きテキストを生成する

▶

生成AI
項目抽出・分類・要約とCRM形式への変換を行う

▶

相互照合
AI抽出値とOCR原文を突き合わせて誤りを検知する

OCR原文が残るため、AI出力の根拠を後から検証できる。

生成AIのみで画像から直接抽出する構成も可能だが、照合対象となる原文が残らないため誤りの検知が難しくなる。検証可能性を重視する場合はOCRを併用する。

注記：事実＝OCRと生成AIは得意領域が異なる／解釈＝併用によりAI出力の照合が可能になる／提案＝読み取りはOCR、解釈は生成AIに分担させる

AIへ送信する前に、個人を特定できる情報を制御する

個人情報を識別子へ置換し、対応表を社内環境に保持することで、AIへ送るデータから個人特定性を除去できる。復元はCRM登録の直前に社内側で行う。

検出対象となる個人情報

- 氏名
- 住所
- 電話番号
- メールアドレス
- 生年月日
- マイナンバー
- 口座番号
- クレジットカード番号
- 保険証番号
- 顧客番号
- 健康・医療情報

※業種固有の識別情報は追加定義する

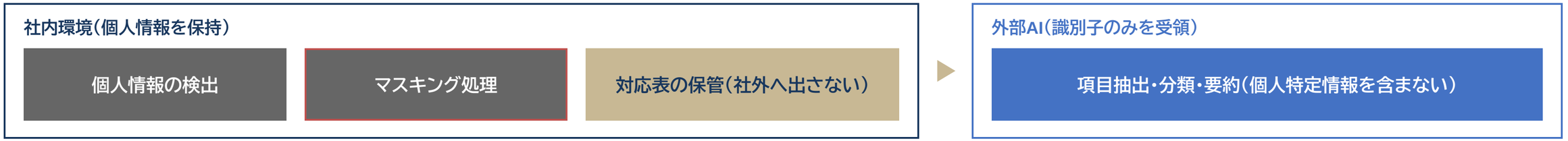
マスキング例

山田太郎 → **[CUSTOMER_NAME_001]**

090-1234-5678 → **[PHONE_001]**

生成AIには識別子のみを送信し、抽出結果を受け取った後、社内環境に保存した対応表を用いてCRM登録直前に実データへ戻す。

社内環境と外部AIの境界



注記：事実＝識別子置換により送信データから個人特定性を除去できる／解釈＝対応表の保管場所が安全性を決定する／提案＝対応表は社内環境に限定し、復元はCRM登録直前に行う

文章ではなく、CRMへ登録できるJSON形式で出力する

出力をCRM項目と同じキー名のJSONに固定し、根拠ページ・信頼度・確認要否を併せて返却させることで、後工程の機械的な検証と人の確認判定が可能になる。

出力例（JSON）

```
{
  "customer_name": "山田太郎",
  "contract_type": "医療保険",
  "application_date": "2026-08-20",
  "monthly_fee": 12500,
  "customer_intent": "保障内容の見直し",
  "source_page": 3,
  "confidence": 0.92,
  "review_required": false
}
```

■ 濃紺＝CRM登録対象項目　■ 青＝検証・確認判定のための付随情報（値は説明用の例示）

設計ポイント

01 CRM項目と同じキー名を設定する
変換処理を減らし、対応漏れを防ぐ

02 日付・金額・選択肢を固定形式にする
表記揺れを排除し、型変換を確定させる

03 不明な項目は推測せず null とする
推測値の混入を防ぎ、確認対象を明確化する

04 根拠となるページ番号・文章を出力する
原本との照合と監査対応を可能にする

05 信頼度と確認要否を付与する
自動登録と人の確認を分岐させる基準にする

06 JSON Schemaで出力形式を制御する
構造の逸脱を仕組みとして防止する

出力を会話文ではなく構造化データに固定することが、CRM連携の前提条件となる。信頼度と根拠ページを持たせるかどうかが、後続の検証設計と人の確認範囲を左右する。

注記：JSON内の値は説明のための例示。事実＝出力形式はスキーマで制御できる／解釈＝付随情報の有無が検証範囲を決める／提案＝信頼度・根拠ページを必須出力とする

生成AIの回答をそのままCRMへ登録しない

AI出力とCRM登録の間に検証層を置き、形式・業務ルール・マスタ整合・原文照合の4段階で判定する。基準を満たさない案件は登録せず保留に回す。

段階 1｜形式検証
構造として成立しているか

・必須項目チェック
・データ型チェック
・日付・電話番号の形式確認

段階 2｜業務ルール検証
業務として妥当な値か

・金額の上下限確認
・郵便番号と住所の整合性
・日付の前後関係の妥当性

段階 3｜マスタ整合検証
CRMに登録できる値か

・CRM選択肢との一致
・商品マスタとの照合
・既存顧客との重複確認

段階 4｜原文照合
根拠と一致しているか

・AI抽出値とOCR原文の照合
・信頼度が基準以下の場合の保留

判定結果の分岐

全段階を通過
CRMへ自動登録する

一部不一致・信頼度不足
担当者の確認画面へ回送する

必須項目欠落・形式不正
登録せず差し戻し、原因を記録する

検証層はAIの精度を補う仕組みであり、精度が向上しても不要にはならない。CRM側の入力規則が変わった際に検証ルールも改定できる構成にしておく必要がある。

注記：事実＝検証項目は形式・業務・マスタ・原文照合に分類できる／解釈＝段階化により保留と差し戻しを区別できる／提案＝検証ルールを改定可能な構成とする

AIの抽出項目とCRMの登録項目を対応させる

マッピングは項目名の対応付けだけでは完結しない。登録ルール（整形・マスタ変換・文字数制限）を項目単位で定義し、CRM側のAPI仕様を事前に確定させる必要がある。

書類上の情報	AI出力項目	CRM項目	登録ルール
契約者氏名	customer_name	Contact.Name	完全一致確認
電話番号	phone	Contact.Phone	数字・ハイフン整形
契約商品	product_name	Policy.ProductId	商品マスタ変換
月額料金	monthly_fee	Policy.Amount	数値型へ変換
要望内容	customer_intent	Opportunity.Needs	500文字以内

事前確認事項｜設計前にCRM側の仕様を確定させる

- CRMに外部登録・更新用APIがあるか
- APIで登録できるオブジェクト・項目の範囲
- 必須項目・文字数・選択肢の制約
- APIの送信上限（件数・頻度）
- CRM側で追加API開発が必要か
- 更新時の上書き可否と履歴の扱い

CRM側に外部登録APIが存在しない場合、CRM側の開発が別途必要となり、全体の工期と人月に直接影響する。要件定義の初期段階で確認すべき最重要事項である。

注記：表中の項目名は設計例。事実＝CRM登録には項目対応と登録ルールの双方が必要／解釈＝API有無が工期を左右する／提案＝要件定義段階でCRM API仕様を確定させる

AIとCRMの間に連携基盤を設ける

AIとCRMを直接つなぐと、認証・変換・再送の責務が分散する。両者の間に連携基盤を置き、送信制御と異常処理を一箇所に集約する構成が保守性を高める。

生成AI処理
JSON形式の抽出結果

▶

連携基盤（Integration Layer）
認証・データ変換・項目マッピング・API送信・再送制御
エラーログ・二重登録防止・API上限管理

▶

CRM
顧客・契約・商談データ

主な連携方法

- REST API
- Webhook
- Power Automate
- Azure Logic Apps
- MuleSoft
- iPaaS
- RPA
- 自社API
- MCPサーバー経由

※CRM製品の対応状況により選択

連携基盤が担う8つの役割

- CRM認証
- データ変換
- CRM項目マッピング
- API送信
- 再送制御
- エラーログ
- 二重登録防止
- API上限管理

※赤系＝異常時の復旧に関わる機能（詳細はP15）

RPAは画面操作に依存するため、CRM側の画面変更で停止しやすい。既存資産の活用を目的とする場合を除き、API連携を基本方式とすることが望ましい。

注記：事実＝連携手段は複数存在し、CRM製品の対応状況で選択肢が変わる／解釈＝責務集約により保守性が向上する／提案＝連携基盤を独立して設ける

正常処理だけでなく、失敗時の復旧方法を設計する

CRM停止・通信断・送信上限は必ず発生する前提で設計する。再送時に同じデータを重複登録しないための冪等性キーが、自動化の信頼性を担保する要件となる。

① 認証・秘密情報管理

誰がCRMへ書き込めるかを統制する

OAuth 2.0・サービスアカウント
個人アカウントに依存させない

APIキーの安全な管理
ソースコードや設定ファイルに直接記載しない

Key Vaultなどによる秘密情報管理
取得権限を限定し、アクセスを記録する

② 異常検知・送信制御

失敗を検知し、送信を制御する

- タイムアウトの設定
- 自動リトライ（回数・間隔の制御）
- Rate Limit対策（送信間隔の調整）
- Dead Letter Queue（失敗データの退避）
- 部分的な登録失敗への対応
- 管理者へのエラー通知

③ 復旧・整合性維持

再送しても状態を壊さない

冪等性キーによる二重登録防止
同一データの再送を検知して重複登録を防ぐ

CRM停止時の再送
復旧後に未送信分を順次送信する

処理状態の記録
原本IDごとに送信状況を追跡できるようにする

失敗時の復旧設計を省略すると、障害発生後に手作業での突合が必要となり、自動化による削減効果が失われる。

注記：事実＝API連携では通信断・上限超過・部分失敗が発生し得る／解釈＝再送設計の有無が運用負荷を決める／提案＝冪等性キーと処理状態記録を必須要件とする

高リスクな登録処理には人の確認を残す

全件自動化ではなく、条件に基づき自動登録と人の確認を分岐させる。確認対象を明文化することで、削減効果とリスク統制を同時に成立させられる。

自動登録できる条件

すべて満たす場合に限り自動処理する

01 必須項目がすべて存在する

02 AIの信頼度が基準以上である

03 CRMマスタと一致する

04 既存顧客情報と矛盾しない

05 金額などが規定範囲内である

06 個人情報の復元に問題がない

人が確認する条件

いずれかに該当する場合は確認画面へ回送する

01 手書き文字を読み取れない

02 同姓同名の顧客が存在する

03 複数の解釈が可能である

04 契約・医療など重要情報を含む

05 AIとOCRの結果が一致しない

06 AIが確認必要と判定した

導入初期は確認対象を広く取り、実績データに基づいて自動登録の範囲を段階的に拡大する運用が現実的である。確認画面では原本と抽出結果を並べて表示し、修正内容を精度改善に反映させる。

注記：事実＝自動化範囲は条件設定により制御できる／解釈＝確認条件の明文化がリスク統制の前提となる／提案＝初期は確認対象を広く設定し、実績に応じて拡大する

クローズド環境に必要な12のセキュリティ管理

管理項目は「通信・データ」「権限・鍵」「証跡・環境」の3領域に整理できる。いずれか一つでも欠けると、閉域構成であっても個人情報の統制は成立しない。

A｜通信とデータの保護

- **01** 通信のTLS暗号化
- **02** 保存データの暗号化
- **03** IPアドレス制限
- **09** 個人情報のマスキング
- **10** 外部通信先のホワイトリスト

B｜権限と鍵の管理

- **04** 最小権限の原則
- **05** サービスアカウントの分離
- **06** APIキーの定期更新
- **07** 一時ファイルの自動削除

C｜証跡と環境の分離

- **08** 操作・処理・変更ログの取得
- **11** 開発・検証・本番環境の分離
- **12** インシデント発生時の緊急停止

緊急停止は「AI送信のみ停止」「CRM登録のみ停止」を個別に実行できる構成としておく。

一時ファイルの自動削除と外部通信先のホワイトリストは、実装後に形骸化しやすい項目である。設定値を運用手順書に明記し、定期的に点検する体制が必要となる。

注記：事実＝管理項目は通信・権限・証跡の3領域に整理できる／解釈＝単一項目の欠落で統制が破綻する／提案＝設定値を運用手順書に明記し定期点検する

「AIの学習に使われない」と「データが保存されない」は異なる

OpenAI APIへ送信したデータは、明示的にオプトインしない限りモデル学習には利用されない。一方、標準設定では不正利用監視ログが最大30日保存される場合がある。両者は別の論点として整理する必要がある。

論点 1｜モデル学習への利用

明示的にオプトインしない限り、API送信データはモデル学習に利用されない

出典：OpenAI「Your data」（developers.openai.com/api/docs/guides/your-data）

論点 2｜データの保存期間

標準設定では、不正利用監視ログが最大30日保存される場合がある

出典：OpenAI「Your data」（developers.openai.com/api/docs/guides/your-data）

確認すべき9つの管理ポイント

AI基盤側の設定

- APIデータのモデル学習利用の有無
- 不正利用監視ログの取り扱い
- 標準環境における保存期間
- store: false の設定
- Zero Data Retention の適用可否

ファイル・検索機能

- Files APIのファイル削除・有効期限
- Vector Storeへの保存内容と保持期間
- アップロード済みファイルの棚卸し

周辺システム側

- CRM側の監査ログの保存内容
- MCP・外部サービス側のデータ管理
- 自社ログに個人情報が残らない設計

社内説明では「学習に使われないこと」と「保存されないこと」を区別して整理する。要件によりZero Data Retentionの適用可否を利用条件とあわせて確認する必要がある。

出典：OpenAI「Your data」developers.openai.com/api/docs/guides/your-data ／ 注記：事実＝学習利用と保存期間は別要件／解釈＝社内説明では両者を分離すべき／提案＝設定値と適用条件を事前確認する

導入費用は「システム開発費」と「月額運用費」に分ける

初期開発費は一度きりだが、月額運用費は処理量に比例して継続的に発生する。OpenAI APIは入力・出力トークン数などで課金されるため、処理量の前提設定が費用見通しを左右する。

初期開発費（一時費用）
構築時に一度発生する

- ・要件定義
- ・セキュリティ設計
- ・OCR・PDF解析
- ・個人情報の検出・マスキング
- ・生成AIの項目抽出
- ・CRM API連携
- ・担当者確認画面
- ・エラー・再送処理
- ・テスト・本番移行

月額運用費（継続費用）
処理量に比例して増減する

- ・OCR利用料
- ・OpenAI API 入力トークン
- ・OpenAI API 出力トークン
- ・PDF画像解析分
- ・File Search・ストレージ
- ・CRM・iPaaS利用料
- ・クラウド基盤費
- ・ログ・監視費
- ・保守運用費

月額費用の計算式
4区分の合計で把握する

月額費用
＝ **AI・OCR利用料**
＋ **クラウド基盤費**
＋ **CRM・連携サービス費**
＋ **保守運用費**

File Searchを利用する場合は、ストレージとツール呼び出しにも料金が発生する。

OpenAI APIは、使用モデルの入力・出力トークン数などで料金が計算される。
月間PDF数・ページ数・解像度・再処理率が未確定の段階では、月額費用は幅を持った試算として扱う。

出典：OpenAI「Pricing」developers.openai.com/api/docs/pricing ／ 注記：事実＝APIはトークン数等で課金される／解釈＝処理量前提が費用を左右する／提案＝初期費と月額費を分離して評価する

業務・AI・セキュリティ・CRMを横断した体制を構築する

必要な役割は7種に整理できる。単一部門では完結しないため、業務側とCRM管理者を初期から参加させ、要件確定の遅延を防ぐ体制設計が前提となる。

役割	主な担当	想定人数
プロジェクトマネージャー	全体管理、要件・進捗・課題管理	1名
業務・CRM担当	業務フロー、CRM項目・登録ルール	1名
AI・OCRエンジニア	OCR、AI抽出、プロンプト、精度改善	1名
API・バックエンド担当	CRM API、認証、変換、再送処理	**1〜2名**
フロントエンド担当	担当者確認・修正画面	1名
クラウド・セキュリティ担当	閉域、権限、暗号化、ログ	1名
テスト・業務検証担当	精度検証、結合・受入試験	**1〜2名**

体制目安（段階別）

技術検証・PoC	**2〜3名**
MVP開発	**3〜5名**
本番構築	**5〜7名**
CRM管理者・業務担当者	**別途参加**

段階が進むにつれて、精度検証とセキュリティ対応の要員比率が高まる。

一人が複数の役割を兼務するため、表に記載した人数の単純合計が必要になるわけではない。

兼務しやすい組み合わせ

AI・OCRエンジニアとAPI担当、PMと業務・CRM担当は兼務可能な場合が多い。

兼務を避けるべき役割

セキュリティ担当と業務検証担当は、統制と検証の独立性を保つため分離する。

注記：人数は標準的な構成を想定した目安（仮定値）。事実＝必要な役割は7種／解釈＝兼務により実人数は圧縮できる／提案＝統制と検証の役割は分離する

標準的な本番構築は5.3〜9.0人月が目安

10工程の合計は5.3〜9.0人月（仮定値）。OCR・PDF処理、個人情報処理、CRM API連携、テスト・本番移行の4工程で全体の約4割を占め、これらの前提条件が総人月を左右する。

開発工程	主な作業	想定人月
要件定義	帳票、業務、CRM項目整理	0.5〜0.8
基本・詳細設計	システム、API、データ、画面設計	0.4〜0.7
OCR・PDF処理	ファイル受付、OCR、画像前処理	**0.6〜1.0**
個人情報処理	検出、マスキング、復元管理	**0.5〜1.0**
AI抽出機能	プロンプト、JSON、精度調整	0.5〜0.8
CRM API連携	認証、変換、登録・更新	**0.6〜1.0**
確認・修正画面	原本確認、修正、承認	0.5〜0.8
基盤・セキュリティ	権限、暗号化、ログ、閉域化	0.5〜0.9
テスト・本番移行	結合、業務検証、本番移行	**0.7〜1.2**
PM・品質管理	進捗、課題、品質、関係者調整	0.5〜0.8
合計	本番構築（セキュリティ・運用込み）	**5.3〜9.0人月**

工程配分の読み方

赤字の4工程（OCR・PDF処理、個人情報処理、CRM API連携、テスト・本番移行）は、帳票の性質とCRM側の仕様により変動幅が大きい。要件定義段階でこの4点の前提を確定させることが、見積精度の向上に直結する。

導入段階	期間	人月	対象範囲
技術検証	1〜1.5か月	0.8〜1.5	OCR・AI精度確認
PoC	1.5〜2.5か月	**1.5〜2.5**	1帳票、簡易連携
MVP	3〜4か月	3.0〜5.0	1〜2帳票、CRM連携
本番構築	4〜6か月	**5.3〜9.0**	セキュリティ・運用込み
大規模展開	6か月以上	10以上	複数帳票・複数CRM

API料金シミュレーションの条件

- 月間PDF数
- 1件あたりのページ数
- PDF画像の解像度・詳細度
- 入力・出力トークン数
- AI再処理率
- OCR処理ページ数
- CRM API送信回数
- File Searchの保存容量・呼び出し回数

人月が増える条件

- 手書き帳票が多い
- 帳票レイアウトが未統一
- CRM側に登録APIがない
- CRM側のAPI開発も必要
- 既存顧客との名寄せが必要
- 完全閉域環境が必要
- 複数部門・複数CRMへ連携する

注記：人月・期間は標準的な構成を想定した目安（仮定値）。事実＝10工程の作業内容／解釈＝4工程が変動要因／提案＝要件定義で帳票性質とCRM仕様を確定させる

安全なPDF・CRM自動連携を実現する10の確認項目

検討開始時に確定すべき項目は「対象書類」「個人情報の取り扱い」「CRM連携条件」「進め方」の4区分に整理できる。この10項目が未確定のままでは、精度も費用も見積もれない。

A｜対象書類の把握

- 01 対象となるPDF・帳票は何種類か
- 02 月間ファイル数・ページ数はどの程度か
- 03 手書き・表・押印は含まれるか

B｜個人情報の取り扱い

- 04 どの個人情報をAIへ送信できるか
- 05 マスキングをどの環境で行うか
- 09 データの保存・削除ルールをどうするか

C｜CRM連携の条件

- 06 CRMへ登録する項目は何項目か
- 07 CRM側に外部登録APIが存在するか
- 08 自動登録と人の確認をどう分けるか

D｜進め方の決定

- 10 PoC・MVP・本番のどこまで実施するか

対象範囲を絞ることで、精度検証と例外処理の負荷を抑えられる。適用範囲の判断が全体工期に最も影響する。

最終メッセージ｜推奨する進め方

標準的な本番構築は、5〜7名程度の体制、4〜6か月、5.3〜9.0人月が一つの目安となる。ただし最初から全帳票・全CRM項目を対象にすると、精度検証と例外処理が膨らむ。

まず「1帳票・1業務・1CRM・限定項目」に絞り、1.5〜2.5人月程度のPoCで精度・API料金・業務削減効果を確認してから本番化する。

STEP 1｜技術検証（0.8〜1.5人月）	STEP 2｜PoC（1.5〜2.5人月）	STEP 3｜MVP（3.0〜5.0人月）	STEP 4｜本番構築（5.3〜9.0人月）

注記：人月・期間・体制は標準的な構成を想定した目安（仮定値）。事実＝確認項目は4区分に整理できる／解釈＝範囲の絞り込みが検証負荷を左右する／提案＝限定範囲のPoCから開始する

お問い合わせ先

対象帳票・個人情報の取り扱い方針・CRMのAPI仕様が確定していない段階でも、要件整理から着手できる。まずは1帳票・1業務に限定したPoCの適用範囲についてご相談ください。

CONTACT

Start Challenge Consulting

担当者　平出

メール　info@startchg.com

本資料に関するご質問、個別の要件整理・API料金試算のご相談を承ります。

Start Challenge
Consulting co., ltd.
挑む力が、世界を照らす

業務・AI・セキュリティ・CRMを横断した
導入設計をご支援します。

ご相談時にお伺いする主な内容

① 対象書類

帳票の種類、月間件数、手書き・押印の有無

② 個人情報の方針

外部送信の可否、求めるクローズド環境のレベル

③ CRM環境

利用製品、外部登録APIの有無、登録項目数

④ 進め方

技術検証・PoC・MVP・本番構築のどこから着手するか