immoGenio

No. 01
März 2026

OCR

KI-Belegerfassung in der Hausverwaltung: Mistral Vision, Konfidenz pro Feld und der Mobile-Capture-Flow

Wie eine ehrliche KI-Belegerfassung Lieferant, Betrag und IBAN aus Foto extrahiert — mit Feld-Konfidenz, IBAN-Kreuzvalidierung und Mobile-Capture für unterwegs.

AUTOR
ImmoGenio

VERÖFFENTLICHT
5. März 2026

ONLINE
www.immogenio.de/blog

— AUF EINEN BLICK

Inhalt

Eine Buchhalterin öffnet montags den Posteingang, und vor ihr liegt ein Stapel aus Handwerker-Rechnungen, Versicherungspolicen, Heizkostenabrechnungen, Quittungen aus dem Baumarkt und einer Überweisungsbestätigung, die jemand zur Post mitgegeben hat. Drei bis fünf Minuten pro Beleg sind realistisch, wenn Lieferant, Rechnungsnummer, Datum, Bruttobetrag, Steuerkennzeichen, IBAN und Verwendungszweck sauber in die Buchhaltung übernommen werden sollen — inklusive Sichtkontrolle gegen die Stammdaten. Bei achtzig Belegen pro Woche und drei Mandaten verschwindet ein ganzer Arbeitstag in einer Tätigkeit, die niemand wirklich gerne macht und die niemand bemerkt, solange sie funktioniert.

Parallel dazu fährt der Hausmeister gegen halb elf zum Baumarkt, kauft drei Dichtungen und eine Spülkastengarnitur, bekommt eine Quittung aus Thermopapier, schiebt sie in die Brusttasche der Arbeitsjacke und vergisst sie dort bis Freitag. Dann ist sie blass und der Beleg landet zerknittert in einem A4-Umschlag, der irgendwann in der Verwaltung ankommt und dort wieder im Posteingang landet — also wieder bei der Buchhalterin von oben.

Genau an dieser Stelle setzt eine moderne KI-Belegerfassung an. Nicht als Versprechen, alles zu automatisieren, sondern als ehrliches Werkzeug, das die mechanische Arbeit übernimmt und den Menschen genau dort hält, wo er nicht ersetzbar ist: bei der Entscheidung, ob ein Beleg sachlich richtig, rechnerisch richtig und freigegeben ist.

Was OCR realistisch leistet — und was nicht

OCR steht für Optical Character Recognition. Klassische Engines wie Tesseract oder ABBYY FineReader leisten genau das, was der Name sagt: Sie wandeln Pixel in Zeichen um. Was diese Engines nicht leisten, ist semantisches Verständnis. Tesseract weiß nicht, dass „Robert Bosch GmbH" ein Lieferant ist und „DE89 3704 0044 0532 0130 00" eine deutsche IBAN. Es liefert eine Textwand, aus der nachgelagerte Regeln, Templates oder ein Layoutparser semantische Felder extrahieren müssen — und scheitert zuverlässig an Layouts, die nicht im Template enthalten sind.

Eine zweite Generation sind Vision-Sprachmodelle, also multimodale Large Language Models, die Bilder und Text gemeinsam verarbeiten. ImmoGenio setzt produktiv das Modell `pixtral-large-latest` von Mistral ein — ein 124-Milliarden-Parameter-Modell mit dediziertem Vision-Encoder, das eine Beleg-Seite direkt als Bild aufnimmt und gegen ein vorgegebenes JSON-Schema extrahiert. Vergleichbare Optionen sind Claude-Sonnet-Vision, Gemini Pro Vision oder GPT-4o, jedes mit eigenen Stärken bei Layout, Sprache und Halluzinationsneigung. Die Wahl fiel auf Mistral aus zwei Gründen: Modell-Hosting in der EU (Mistral La Plateforme, Region Frankreich) und ein klares „Daten werden nicht zum Training verwendet" im B2B-Vertrag.

Was Vision-LLMs gegenüber klassischem OCR ändern, ist der Sprung vom Zeichen zum Feld. Das Modell beantwortet nicht „Was steht da?", sondern „Was ist der Lieferantenname?" und „Was ist der Bruttobetrag in Euro?" — und liefert die Antworten in einem strikt typisierten JSON gegen ein Schema, das Felder wie `lieferant`, `lieferant_iban`, `rechnungsnummer`, `datum`, `betrag_brutto`, `betrag_netto`, `mwst_satz`, `verwendungszweck`, `kaltmiete`, `kaution_betrag` umfasst. Was es nicht ändert: Die Verantwortung für die Richtigkeit der Buchung bleibt beim Menschen.

Warum Konfidenz pro Feld zentral ist

Viele OCR-Lösungen melden einen Konfidenz-Score auf Beleg-Ebene — „Beleg mit 91 Prozent Sicherheit erkannt". Diese Zahl ist im besten Fall nutzlos und im schlechteren Fall gefährlich. Ein Beleg, bei dem der Lieferantenname mit 0,98 Sicherheit erkannt wurde, das Rechnungsdatum mit 0,95 und die IBAN mit 0,71, ist genau dort gefährlich, wo es weh tut: an der IBAN. Ein gemittelter Score von 0,88 verschleiert das.

ImmoGenio führt deshalb für jedes extrahierte Feld eine separate Konfidenzzahl mit, die in der Spalte `extraktion_konfidenz` als JSON-Map abgelegt wird:

```
{
  "lieferant": 0.98,
  "rechnungsnummer": 0.94,
  "datum": 0.97,
  "betrag_brutto": 0.93,
  "lieferant_iban": 0.82,
  "verwendungszweck": 0.71
}
```

Auf dieser Map sitzen Auto-Approve-Thresholds, die produktiv pro Feldtyp kalibriert sind. Für die IBAN liegt der Threshold bei 0,95 — alles darunter geht zwingend in das Human-in-the-Loop-Review. Für die Kaltmiete liegt der Threshold bei 0,90, für den Kautionsbetrag wieder bei 0,95, weil hier ein Tippfehler eine Mietsicherheit auf das falsche Konto schickt. Diese Zahlen sind keine Marketing-Zahlen, sondern das Ergebnis aus tausenden manuell nachgereviewten Belegen pro Tenant.

Das Prinzip dahinter: Konservative Defaults, transparente Schwellen, kein „die KI ist sich sicher genug" über Felder hinweg. Wer einen Score auf Beleg-Ebene meldet, redet von Statistik. Wer Schwellen pro Feld kalibriert, redet von Risikomanagement.

Kreuzvalidierung gegen Stammdaten und IBAN-Whitelist

Eine Konfidenz von 0,98 für einen Lieferantennamen ist immer noch nur eine Meinung des Modells. Die zweite Sicherheitsebene ist die Kreuzvalidierung gegen die eigenen Stammdaten. ImmoGenio pflegt pro Tenant eine Tabelle `lieferanten_bankverbindungen`, in der jede zulässige IBAN je Lieferant mit einem Vier-Augen-Bestätigungs-Flag steht. Erkennt das Vision-Modell die IBAN „DE89 3704 0044 0532 0130 00" mit hoher Konfidenz, prüft die Pipeline:

1. Existiert dieser Lieferant in den Stammdaten?
2. Ist die erkannte IBAN für diesen Lieferanten freigegeben (also auf der Whitelist)?
3. Wenn nicht: Existiert die IBAN für einen anderen Lieferanten — also liegt möglicherweise ein gefälschter Beleg vor?

Punkt drei ist nicht akademisch. Genau diese Vorgehensweise — eine echte Lieferantenrechnung wird abgefangen, die IBAN gegen die eines Betrügers ausgetauscht, der Beleg geht ungeöffnet in die Verwaltung — ist der häufigste Vektor bei Rechnungsbetrug in Hausverwaltungen. Die Pipeline ist deshalb eng verzahnt mit dem In-Flight-Swap-Schutz für E-Rechnungen, der den gleichen Vergleich für ZUGFeRD- und XRechnung-Belege durchführt. Eine extrahierte IBAN, die nicht auf der Whitelist steht, blockiert die Auto-Approve-Strecke unabhängig von der Konfidenz und landet im Review mit einem expliziten Hinweis „IBAN nicht in Lieferanten-Bankverbindungen freigegeben".

Staging-Workflow: ocr_pending, ocr_ready, geprueft

Das Datenmodell hinter der Belegerfassung ist bewusst einfach. Migration 047 hat die Spalten `staging_status`, `extraktion_raw`, `extraktion_normalisiert` und `extraktion_konfidenz` zur Beleg-Tabelle hinzugefügt. Ein Beleg durchläuft drei Zustände:

- **`ocr_pending`** — Der Beleg ist hochgeladen, EXIF-bereinigt und liegt im Object Storage. Eine BullMQ-Queue arbeitet die Extraktion ab, der Status bleibt bis zum Abschluss auf `pending`.
- **`ocr_ready`** — Die Extraktion ist erfolgt, die JSON-Antwort ist gegen das Schema validiert, die Konfidenzwerte stehen, die Kreuzvalidierung gegen Stammdaten ist abgeschlossen. Der Beleg ist sichtbar in der KI-Posteingang-Liste, mit Markierungen pro Feld: grün für „über Threshold und whitelisted", gelb für „Review erforderlich", rot für „Validierungsfehler".
- **`geprueft`** — Ein Mensch hat den Beleg gesehen, die Felder bestätigt oder korrigiert und die Buchung freigegeben. Erst jetzt greift die nachgelagerte Buchungsregel-Engine, die aus Lieferant, Verwendungszweck und Kostenkategorie das Sachkonto und die Kostenstelle ableitet.

Auto-Approve ist ausdrücklich kein „skip Review", sondern nur ein „grüne Felder werden vorausgewählt". Auch ein vollständig grüner Beleg muss bestätigt werden. Die einzige Abkürzung ist visuell: Statt jedes Feld einzeln zu prüfen, scrollt der Bearbeiter durch eine Liste vorausgefüllter Werte und drückt einmalig auf „Bestätigen", sofern keine Markierung gelb oder rot ist.

Mobile-Capture: Rückkamera, Full-Screen, Session-Queue

Die zweite Hälfte des Workflows spielt nicht im Büro, sondern unterwegs — und genau dort scheitert klassische Belegerfassung. Wer einen Hausmeister bittet, Quittungen zu sammeln, sie zu kopieren, einzuscannen und per E-Mail zu schicken, bekommt verzögert oder gar nicht. Wer ihn bittet, eine Quittung zu fotografieren und auf „Senden" zu drücken, bekommt sie noch im Baumarkt-Parkplatz.

ImmoGenio bietet dafür unter `/buchhaltung/tagesgeschaeft/ki-posteingang/foto` eine dedizierte Full-Screen-Seite — kein modaler Dialog, kein Mehrschritt-Wizard. Drei Eigenschaften sind technisch entscheidend:

1. `capture="environment"` **am Datei-Input** — Mobile Browser interpretieren dieses Attribut als „öffne direkt die Rückkamera". Kein Auswahldialog, kein Wechsel zwischen Foto- und Galerie-Modus, kein versehentlicher Front-Selfie-Beleg. Auf dem ersten Tippen ist die Kamera offen.
2. **Session-Queue mit Live-Status** — Mehrere Belege in Folge werden lokal in eine Queue gelegt und parallel hochgeladen. Der Hausmeister sieht „Beleg 1 — extrahiert", „Beleg 2 — wird verarbeitet", „Beleg 3 — hochgeladen". Das nimmt die Unsicherheit, ob der Upload geklappt hat, und reduziert das doppelte Fotografieren.
3. **EXIF-Strip vor dem Upload** — Bevor das Foto die Browser-Sandbox verlässt, werden EXIF-Metadaten entfernt. Geokoordinaten, Geräte-ID, Kamera-Seriennummer haben in einem Buchungsbeleg nichts zu suchen, sind aber standardmäßig in jedem Smartphone-Foto enthalten. Das Strippen ist ein einfacher Canvas-Re-Encode in JPEG mit Qualität 0,82.

Der gleiche Mobile-First-Gedanke steckt hinter dem Offline-First-Übergabeprotokoll für Hausmeister — ein Hausmeister vor Ort hat selten Lust auf zwei verschiedene Apps. Die Capture-Seite ist deshalb als Teil der gleichen Mobile-Surface konzipiert, mit identischer Auth, identischen Touch-Targets und identischem Offline-Verhalten.

DSGVO: keine US-Provider ohne TIA, keine Trainings-Verwendung

Wer KI in der Buchhaltung einsetzt, fasst zwangsläufig personenbezogene Daten an. Auf einem Beleg stehen Lieferantenname, Bankverbindung, eventuell Bearbeitername, eventuell Kundennummer, in Mietabrechnungen sogar Mieternamen und Adressen. Damit greifen Art. 5, Art. 6, Art. 28 und — sobald algorithmische Entscheidungen über Menschen getroffen werden — Art. 22 DSGVO.

Die produktive Architektur stützt sich auf drei Punkte:

- **Modell-Hosting in der EU**: Mistral La Plateforme betreibt die Inferenz in Frankreich. Damit entfällt die Notwendigkeit eines Transfer Impact Assessments für US-Cloud-Provider und das Hin und Her um die Angemessenheit der EU-US-Adäquanzentscheidung. Wer mit OpenAI oder Anthropic-Hosting in den USA arbeiten möchte, muss eine TIA dokumentieren und die Standardvertragsklauseln samt zusätzlicher Maßnahmen umsetzen — möglich, aber für eine Hausverwaltung mit fünfzig Wohneinheiten ein Mehraufwand, den die EU-Variante einspart.
- **Vertraglich ausgeschlossene Trainings-Verwendung**: Der Auftragsverarbeitungsvertrag mit Mistral schließt aus, dass übermittelte Belege zum Training zukünftiger Modelle verwendet werden. Das ist nicht selbstverständlich und gehört in jeden Vergleich.
- **Datenminimierung nach Art. 5 Abs. 1 lit. c DSGVO**: EXIF-Strip vor Upload, kein Speichern von Zwischenständen, automatische Löschung der Modell-Roh-Antwort nach 30 Tagen (`extraktion_raw` wird genullt, `extraktion_normalisiert` bleibt). Was nicht erhoben werden muss, wird nicht erhoben.

Art. 22 DSGVO — keine ausschließlich automatisierten Entscheidungen mit rechtlicher Wirkung — ist über die HITL-Pflicht implizit umgesetzt. Eine Buchung wird nie ohne menschliche Bestätigung erzeugt, eine Zahlung schon gar nicht.

Praxisbeispiel: Quittung im Baumarkt-Parkplatz

Der Hausmeister kauft um 10:47 Uhr drei Spülkastendichtungen und eine Eckventil-Garnitur, bekommt eine Bon-Quittung über 18,73 Euro, geht zum Auto, öffnet auf dem Smartphone die Capture-Seite und tippt auf „Beleg fotografieren". Die Rückkamera öffnet, er hält die Quittung gegen das Armaturenbrett und drückt ab. Drei Sekunden später ist das Foto hochgeladen, weitere acht Sekunden später ist die Extraktion durch:

- Lieferant: „OBI Bau- und Heimwerkermärkte" — 0,97
- Datum: 2026-04-29 — 0,99
- Betrag brutto: 18,73 EUR — 0,96

- USt-Satz: 19 Prozent — 0,94
- Verwendungszweck: „Sanitär-Verbrauchsmaterial" — 0,71 (semantische Zuordnung, nicht Texterkennung)

Der Lieferant ist in den Stammdaten, die Quittung hat keine IBAN (Barzahlung), die Konfidenzen liegen über den Auto-Approve-Schwellen. Der Beleg landet im Posteingang der Buchhaltung mit Status `ocr_ready` und vorausgewählter Kostenstelle „Liegenschaft Hauptstraße 14, Sanitär". Die Buchhalterin öffnet ihn um 14:30 Uhr, scrollt durch die Felder, korrigiert den Verwendungszweck auf „WC-Reparatur Wohnung 3.OG", drückt „Bestätigen". Bearbeitungszeit: zwölf Sekunden statt drei Minuten.

Grenzen: was die Pipeline heute nicht kann

Eine ehrliche Beschreibung gehört dazu. Die aktuelle Pipeline hat klare Grenzen, und die werden so kommuniziert:

- **Keine Auto-Buchung ohne menschliche Bestätigung.** Auch ein vollständig grüner Beleg muss bestätigt werden. Die KI schlägt vor, der Mensch entscheidet. Das ist Absicht, kein Zwischenstand.
- **Handgeschriebene Belege sind unzuverlässig.** Vision-LLMs sind bei Druck deutlich besser als bei Handschrift. Wer eine handnotierte Spesenquittung fotografiert, bekommt mit hoher Wahrscheinlichkeit ein Review mit niedrigen Konfidenzen.
- **Fremdwährung ist v1 nicht abgedeckt.** Belege in CHF, USD oder GBP werden zwar extrahiert, aber Wechselkurs-Logik und Mehrwährungs-Buchungen sind nicht Teil des Auto-Approve-Pfads.
- **PDFs mit eingescannten Bildern statt Text-Layer durchlaufen die Vision-Pipeline**, das ist langsamer als bei nativen PDFs mit eingebettetem Text. Eine Hybrid-Pipeline, die zuerst den Text-Layer prüft und nur bei Fehlen das Vision-Modell aufruft, ist auf der Roadmap.

Diese Grenzen sind sichtbar in der UI und nicht in einem versteckten FAQ. Wer einen handschriftlichen Beleg hochlädt, bekommt einen entsprechenden Hinweis und nicht still niedrige Konfidenzen.

Wo wir stehen — IDP-Orchestrator und Few-Shot-Pflege

Die Belegerfassung läuft produktiv über einen Intelligent-Document-Processing-Orchestrator (Epic #102), der die Routen `POST /api/belege` und `POST /idp/extraktionen` bündelt. Der Orchestrator wählt pro Dokumentenklasse das passende Modell und Schema,

ruft die Vision-Inferenz auf, validiert das JSON gegen das Schema, normalisiert die Werte (IBAN-Formatierung, Datums-Parsing, Betragsformat) und schreibt das Ergebnis in die Staging-Tabelle.

Eine Eigenheit des Designs: Es gibt **kein Fine-Tuning** des Modells pro Tenant. Stattdessen pflegt jeder Tenant eine Few-Shot-Bibliothek mit fünf bis zwanzig kuratierten Beispielen pro Dokumentenklasse — eine Hausgeld-Abrechnung sieht bei Verwaltung A anders aus als bei Verwaltung B. Diese Beispiele werden bei jedem Aufruf in den Prompt eingebettet. Das ist günstiger als Fine-Tuning, sofort wirksam ohne Modell-Re-Training und ändert sich pro Tenant in Minuten statt Tagen.

Die offene API hinter dem Orchestrator ist Teil der breiteren Schnittstellen-Strategie zu DATEV, Messdiensten und Smartlocks — eine Belegerfassung, die ihre Ergebnisse nur in die eigene Buchhaltung schreibt und nicht in den DATEV-Export Format 7 oder den SEPA-Sammellauf übergibt, hat ihren Wert nur halb gehoben.

EU-AI-Act-Einordnung: Belegerfassung mit menschlicher Bestätigung fällt nicht unter die Hochrisiko-Kategorie nach Anhang III. Sie ist ein begrenztes Risiko mit Transparenzpflicht — die Bearbeiter sehen ausdrücklich, welche Felder von einer KI vorgeschlagen wurden, und können das in der Versionshistorie jedes Belegs nachvollziehen.

Abschluss

Drei Minuten manuelle Erfassung pro Beleg auf zwölf Sekunden Bestätigung zu reduzieren ist keine Magie. Es ist ein Vision-Modell mit ehrlicher Konfidenzangabe pro Feld, eine harte Kreuzvalidierung gegen die eigenen Stammdaten, ein klarer Staging-Workflow mit menschlicher Bestätigung, eine Mobile-First-Erfassung mit `capture="environment"` und EXIF-Strip — und eine konservative Haltung, die lieber gelb markiert als falsch automatisiert. Wer hier sauber arbeitet, hat einen Wettbewerbsvorteil, der Quartal für Quartal wächst, weil die Few-Shot-Bibliothek sich pro Tenant verfeinert und die Auto-Approve-Quote organisch steigt.

Wer KI in der Hausverwaltung breiter denkt, findet das gleiche Muster im KI-Telefonassistenten für die Hausverwaltung wieder: Modelle übernehmen die mechanische Arbeit, Menschen bleiben in der Entscheidung, Datenschutz wird nicht nachgelagert, sondern eingebaut.

Fragen, Rückmeldungen oder Wünsche zur Belegerfassung — gern an kontakt@immogenio.de.