

Kassenbon-OCR zwischen lokalen Modellen und Cloud-APIs

Eine explorative Studie zu elf Konfigurationen auf 100 identischen synthetischen deutschen Bons

SCFold

Messstand: 16. Oktober 2026 · Sprache der Stichprobe: Deutsch

Zusammenfassung

Wir vergleichen elf lokale und über Azure bereitgestellte OCR-Konfigurationen auf derselben Stichprobe von 100 synthetischen deutschen Kassenbons. Der Korpus enthält 482 Referenzartikel und 34.624 normalisierte Referenzzeichen. Bewertet werden die Zeichengenauigkeit des Volltexts, exakt extrahierte Artikelfelder nach einem gemeinsamen Regelparser sowie beobachtete Latenzen und geschätzte API-Kosten. GLM-OCR erreicht mit vier Zeicheneditierfehlern die höchste Zeichengenauigkeit (99,9884 %). Unter den acht Cloud-Konfigurationen führt GPT-6.1 Sol mit 99,9336 %; die geschätzten Kosten betragen 0,339554 USD für 100 Bons. GPT-5 nano benötigt für dieselbe Stichprobe geschätzt 0,012068 USD bei 97,6404 % Zeichengenauigkeit. Azure Document Intelligence liefert zusätzlich native Artikelfelder: 98,34 % korrekte Produktnamen und 99,79 % korrekte Zeilensummen. Diese direkte Extraktion wird separat ausgewertet. Die Studie zeigt, dass nahezu fehlerfreier OCR-Text noch keine gleichwertige strukturierte Artikelerkennung garantiert. Die Aussagekraft bleibt durch die kleine synthetische Stichprobe, unterschiedliche Reasoning-Einstellungen und unkontrollierte Leistungszustände der lokalen Hardware begrenzt.

Schlüsselwörter: Optical Character Recognition · Kassenbons · Document AI · lokale Inferenz · Azure · Belegextraktion

1 Einleitung und Forschungsfragen

Bei der automatischen Verarbeitung von Kassenbons werden sowohl lesbarer Text als auch verlässliche Artikelpositionen benötigt. Produktbezeichnungen, Mengen und Zeilensummen sind für nachgelagerte Anwendungen unmittelbar relevant. Eine hohe Volltextgenauigkeit allein beschreibt diesen Nutzen nur teilweise. Zugleich unterscheiden sich lokale OCR-Systeme und Cloud-Dienste hinsichtlich Betriebsaufwand, Latenz und Abrechnungsmodell.

Die Untersuchung beantwortet drei Fragen: **(F1)** Welche der getesteten Konfigurationen transkribiert denselben Korpus am genauesten? **(F2)** Wie gut lassen sich aus den Antworten mit identischen Regeln Artikelfelder gewinnen? **(F3)** Welche API-Kosten und Antwortzeiten wurden dabei beobachtet? Direkte strukturierte Felder eines spezialisierten Belegmodells werden als zusätzliche Auswertung behandelt.

Die Ergebnisse sind eine Momentaufnahme konkreter Modell- und Pipelinekonfigurationen. Sie bilden weder einen repräsentativen Test echter Supermarktbelege noch einen kontrollierten Leistungsvergleich der Hardware ab. Der Bericht enthält ausschließlich die elf im Folgenden aufgeführten Konfigurationen.

2 Material und Methoden

2.1 Datensatz und Auswahl

Verwendet wird der öffentliche Datensatz `albertobarnabo/synthetic-receipts-ocr` [1]. Aus dem Split **eval** werden die ersten 100 Beispiele mit **locale=DE** in der ursprünglichen Reihenfolge ausgewählt; Eingabe ist jeweils **image_photo**. Es erfolgt keine Auswahl nach Bildqualität und kein Training innerhalb dieser Studie. Referenzen sind `full_text` und `fields.lines`. Die Dataset-Karte nennt Apache 2.0 als Lizenz. Die Stichprobe enthält keine echten REWE-, Aldi- oder Lidl-Bons.

2.2 Ausführung und Konfigurationen

Alle Modelle erhalten dieselben vorbereiteten PNG-Dateien: EXIF-Drehung, RGB-Konvertierung und proportionale Begrenzung auf 1.003.520 Pixel; kein Deskew oder Unwarping. Lokale Modelle laufen nacheinander auf einer RTX 3070 Ti Laptop GPU mit 8 GiB VRAM. Ein lokaler Warm-up je Modell wird nicht gewertet. Cloud-Aufrufe übertragen Bilder und den festen Transkriptionsprompt ohne Referenzlösungen. Je Cloud-Teilttest sind höchstens zwei Aufrufe gleichzeitig konfiguriert.

Konfiguration	Ausführung	Aufgabe / API	Reasoning
GLM-OCR	Lokal	OCR	-
gpt-6.1-sol	Azure	Chat-Transkription	low
gpt-5.6-sol	Azure	Chat-Transkription	none
gpt-6-sol	Azure	Chat-Transkription	none
gpt-5.6-terra	Azure	Chat-Transkription	none
mistral-ocr-4-0	Azure	Native OCR	-
PP-OCrv6 Medium	Lokal	OCR	-
Azure DI · Receipt	Azure	prebuilt-receipt	-
gpt-5-nano	Azure	Chat-Transkription	minimal
PaddleOCR-VL-1.6	Lokal	Text-Spotting	-
Phi-4-multimodal-instruct	Azure	Chat-Transkription	Standard

Tabelle 1. Untersuchte Konfigurationen, geordnet nach Zeichengenauigkeit. „Azure DI · Receipt“ steht für Azure AI Document Intelligence, Modell `prebuilt-receipt`. Lokale Modellquellen: [2–4]; Belegmodell: [5].

Chatmodelle verwenden `detail=high` und maximal 3.072 Ausgabetoken einschließlich gegebenenfalls anfallender Reasoning-Token. GPT-6.1 Sol benötigt mindestens low; GPT-5 nano verwendet minimal. Mistral OCR nutzt die native OCR-API, Document Intelligence die API-Version 2024-11-30 ohne kostenpflichtige Add-ons. Diese beiden Dienste erhalten keinen Chat-Prompt.

2.3 Bewertungsregeln

Der Textvergleich vereinheitlicht Groß-/Kleinschreibung (Unicode casefold) und Leerraum und entfernt Folgen von mindestens drei Trennzeichen aus `-`, `_` oder `=`. Mit Levenshtein-Distanz d und normalisierter Referenz r gilt über alle Bons:

$$\text{OCR-Accuracy} = \max(0; 100 - 100 \times \sum d(r_i, o_i) / \sum |r_i|).$$

„Bons exakt“ verlangt vollständige Gleichheit der normalisierten Texte. Artikelausgaben werden per globaler eindeutiger Namenszuordnung (Hungarian-Verfahren, `SequenceMatcher` $\geq 0,45$) mit Referenzartikeln verbunden. Danach müssen Namen exakt normalisiert stimmen; bei Zahlen gilt eine Abweichung $< 0,005$. Feldquoten teilen die Summe korrekter Felder durch $\sum \max(\text{Referenzartikel}, \text{Ausgabeartikel})$. Fehlende Werte zählen nicht als korrekt; unbekannte Mengen bleiben null.

3 Ergebnisse: OCR und gemeinsame Artikelpipeline

Jede der elf Konfigurationen liefert 100 erfolgreiche gewertete Antworten. Alle Volltexte durchlaufen denselben Regelparser. Die Feldquoten in Tabelle 2 messen daher OCR plus Nachverarbeitung; sie sind keine Bewertung modellindividuell optimierter JSON-Prompts.

Konfiguration	OCR (%)	Namen (%)	Zeilen- summen (%)	Mengen (%)	Bons exakt
GLM-OCR	99,9884	94,81	95,41	12,97	96/100
gpt-6.1-sol	99,9336	92,43	95,02	12,95	81/100
gpt-5.6-sol	99,8469	88,42	95,21	12,97	65/100
gpt-6-sol	99,7661	87,03	94,61	12,97	63/100
gpt-5.6-terra	99,5754	88,84	94,42	12,95	65/100
mistral-ocr-4-0	98,7610	93,98	94,18	12,45	81/100
PP-OCRv6 Medium	98,5068	93,19	93,99	12,63	57/100
Azure DI · Receipt	97,9610	92,76	93,36	12,27	38/100
gpt-5-nano	97,6404	69,96	83,40	11,07	22/100
PaddleOCR-VL-1.6	97,1609	80,11	81,78	10,41	24/100
Phi-4-multimodal-instruct	78,8066	9,84	12,70	1,02	7/100

Tabelle 2. Micro-gewichtete Volltext- und Feldgenauigkeit; absteigend nach OCR. Zeilensummen sind Artikelbeträge, nicht die Bon-Gesamtsumme. Unterschiedliche Feldnamen berücksichtigen zusätzlich erzeugte Artikel (488-538).

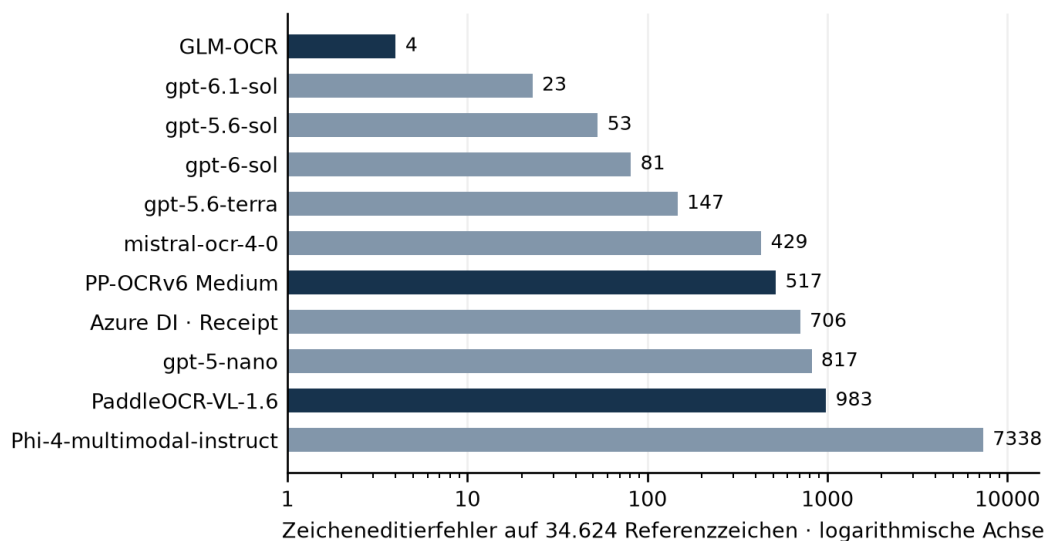


Abbildung 1. Absolute Zeicheneditierfehler. Dunkel: lokale Modelle; hell: Azure. Die logarithmische Achse macht die Spannweite von 4 bis 7.338 Fehlern sichtbar. Alle Balken beruhen auf demselben Referenzumfang.

GLM-OCR erreicht 96 vollständig exakte Bons und vier Zeicheneditierfehler. GPT-6.1 Sol liefert unter den Cloud-Konfigurationen die höchste Zeichengenauigkeit (23 Fehler). GPT-5.6 Sol und GPT-6 Sol folgen mit 53 beziehungsweise 81 Fehlern. Die Reihenfolge bei Produktnamen unterscheidet sich: Mistral OCR und PP-OCRv6 erzielen mit dem gemeinsamen Parser höhere Namensquoten als die hier getesteten Sol-Deployments.

4 API-Kosten und beobachtete Latenzen

Konfiguration	Ø s/Bon	USD/Bon	USD/100 Bons	USD/1.000 Bons
GLM-OCR	5,96	0,000000	0,000000	0,0000
gpt-6.1-sol	6,58	0,003396	0,339554	3,3955
gpt-5.6-sol	3,90	0,006643	0,664284	6,6428
gpt-6-sol	3,32	0,003312	0,331184	3,3118
gpt-5.6-terra	2,72	0,003705	0,370498	3,7050
mistral-ocr-4-0	2,46	0,004000	0,400000	4,0000
PP-OCRv6 Medium	1,54	0,000000	0,000000	0,0000
Azure DI · Receipt	4,48	0,010000	1,000000	10,0000
gpt-5-nano	3,46	0,000121	0,012068	0,1207
PaddleOCR-VL-1.6	14,91	0,000000	0,000000	0,0000
Phi-4-multimodal-instruct	13,90	0,000169	0,016864	0,1686

Tabelle 3. Mittlere Latenz pro erfolgreichem Bon und geschätzte API-Kosten. Lokale Modelle: 0 USD API-Kosten; Strom und Hardware sind nicht Bestandteil dieser Rechnung. Werte für 1.000 Bons sind eine lineare Projektion, keine zusätzliche Messung.

4.1 Verbrauch und Referenztarife

Für Tokenmodelle werden nicht gecachte Eingabe-, Cache- und Ausgabetoken getrennt mit den dokumentierten Preisen multipliziert. Reasoning-Token sind bereits Teil der Ausgabetoken und werden nicht erneut addiert. Native OCR-Dienste werden pro Seite berechnet. Jeder Testbon entspricht einer Seite; bei Document Intelligence stammt die Seitenzahl aus `analyzeResult.pages`, nicht aus einem separaten Abrechnungszähler.

Azure-Konfiguration	Input ¹	Cache ¹	Output ¹	Abrechnung
gpt-6.1-sol	2,00	0,100	10,00	Token
gpt-5.6-sol	4,00	0,400	20,00	Token
gpt-6-sol	2,00	0,200	10,00	Token
gpt-5.6-terra	2,00	0,200	12,00	Token
mistral-ocr-4-0	-	-	-	4,00 / 1.000 Seiten
Azure DI · Receipt	-	-	-	10,00 / 1.000 Seiten
gpt-5-nano	0,05	0,005	0,40	Token
Phi-4-multimodal-instruct	0,08	-	0,32	Token

Tabelle 4. Referenztarife in USD, geprüft am 16.10.2026. ¹ Je 1 Mio. Token; Cache = Cache-Lesen. Cache-Schreiben: 2,50 USD bei GPT-6.1 Sol, GPT-6 Sol und Terra; 5,00 USD bei GPT-5.6 Sol; im Korpus keine entsprechenden Cache-Schreibtoken. Quellen: [6-9].

Die Rechnung nutzt Standard-Global-Referenzpreise; für GPT-6 und GPT-6.1 ist US East die dokumentierte Referenzregion. Regionale Vertragspreise, Rabatte, Guthaben, F0-Kontingente, Steuern, feste Bereitstellungen und Vorab-/Warm-up-Proben sind nicht enthalten. Die Summen sind Verbrauchsschätzungen und keine Azure-Abrechnung.

4.2 Grenzen der Zeitmessung

PP-OCRv6 sowie die ersten drei PaddleOCR-VL-Bons liefen vor Anschluss des Laptop-Netzteils; der vierte Paddle-Bon lag im Übergang. GLM-OCR lief danach. Lokale Zeiten schließen Modellladen aus. API-Zeiten enthalten Netzwerk, Warteschlange und mögliche Wiederholungen; Document Intelligence außerdem Statusabfragen. Die unterschiedlichen Leistungszustände und Messzeitpunkte erlauben keinen kontrollierten Geschwindigkeitsrang. Der gemessene Durchlauf enthält keine Wiederholungsstudie.

5 Direkte Extraktion und Diskussion

5.1 Native Artikelfelder von Document Intelligence

Das Receipt-Modell [5] liefert im selben Aufruf zusätzliche strukturierte Artikelfelder. Description, Quantity, Price und TotalPrice werden in das gemeinsame Artikelschema übertragen. Nicht gelieferte Mengen oder Einzelpreise bleiben null; weder eine Menge 1 noch ein aus anderen Feldern abgeleiteter Preis wird ergänzt. Diese Auswertung ist unabhängig vom gemeinsamen OCR-Regelparser.

Natives Feld	Korrekte Felder / Nenner	Genauigkeit (%)
Produktname	474/482	98,34
Zeilensumme	481/482	99,79
Menge	64/482	13,28
Einzelpreis	64/482	13,28

Tabelle 5. Direkte strukturierte Extraktion auf denselben 100 Bons; der aggregierte Feldnenner beträgt 482. Ein Nenner von 482 allein garantiert nicht, dass in jedem Bon die richtige Artikelstruktur vorliegt.

Die direkte Extraktion erreicht 474 korrekte Namen und 481 korrekte Zeilensummen. Mit OCR plus Regelparser liegen dieselben Quoten des Dienstes bei 92,76 % und 93,36 %. Dieser Unterschied zeigt den Einfluss der nachgelagerten Strukturierung. Er belegt keine allgemeine Überlegenheit gegenüber anderen Modellen, deren direkte JSON-Extraktion hier nicht optimiert oder vergleichbar getestet wurde. Nur 64 Mengen und Einzelpreise stimmen exakt; nicht gelieferte Werte werden konservativ als nicht korrekt gezählt, auch wenn eine Referenzmenge implizit 1 ist.

5.2 Interpretation für die Modellwahl

GLM-OCR liefert in dieser Stichprobe den stärksten lokalen Transkriptionswert. GPT-6.1 Sol erreicht den stärksten Cloud-OCR-Wert und kostet mit 0,339554 USD pro 100 Bons rund halb so viel wie GPT-5.6 Sol (0,664284 USD). Es arbeitet mit low statt none; seine höhere beobachtete Latenz ist nur innerhalb dieser Konfiguration interpretierbar.

GPT-5 nano ist hier die günstigste Cloud-Konfiguration, erreicht mit unserem Parser aber nur 69,96 % exakte Produktnamen. Mistral OCR und PP-OCRv6 erzielen trotz geringerer Volltextgenauigkeit hohe Artikelquoten. Für strukturierte Belegdaten sollten die tatsächlich benötigten Felder daher getrennt geprüft werden.

5.3 Bedrohungen der Validität

Externe Validität. Die Auswahl ist klein, synthetisch und auf deutsche Beispiele begrenzt. Händler, Schriftarten, Layouts und Bildfehler bilden reale Bons nur teilweise ab. Aussagen zu konkreten Supermarktketten oder seltenen Rabattformaten sind nicht ableitbar. Öffentliches Trainingsvorwissen lässt sich nicht ausschließen.

Interne Validität. Reasoning, proprietäre Bildverarbeitung und Nachverarbeitung sind nicht vollständig angeglichen. Mehrzeilige Mistral-BBox-Blöcke werden näherungsweise auf Zeilen verteilt; bei unvollständiger Abdeckung bleibt vollständiger bereinigter Text erhalten. Stromversorgung, Netzwerk und Serverzustand beeinflussen die Zeiten.

Konstruktvalidität und Statistik. Normalisierung ignoriert Groß-/Kleinschreibung und Trennlinien. Die Namenszuordnung kann Feldquoten bei Duplikaten oder starken Textfehlern beeinflussen. Es fehlen Wiederholungsläufe, Konfidenzintervalle und Signifikanztests; Differenzen sind rein deskriptiv. Einheitliche prozessbezogene VRAM-Spitzenwerte fehlen, Cloud-VRAM ist nicht beobachtbar.

6 Schlussfolgerung

Auf dieser kleinen synthetischen Stichprobe führt GLM-OCR bei der Volltextgenauigkeit. GPT-6.1 Sol erzielt den höchsten Cloud-Wert, während GPT-5 nano die geringsten geschätzten Cloud-Kosten aufweist. Document Intelligence liefert besonders genaue native Namens- und Zeilensummenfelder, weist aber Lücken bei Mengen auf. Eine belastbare Produktionsentscheidung erfordert zusätzliche reale Bons, feldbezogene Abnahmekriterien und wiederholte Messungen unter stabilen Leistungsbedingungen.

Anhang A Datenbasis und Reproduzierbarkeit

Dieser PDF sind die ausgewerteten Kennzahlen als JSON, die Vergleichstabelle als CSV und die protokollierten lokalen Modellumgebungen als JSON-Anlagen beigefügt. Die Begleitdaten enthalten 1.100 Einzelwertungen und SHA-256-Hashes der 100 Eingabebilder. API-Zugänge, private Bons und Ressourcenadressen sind nicht enthalten. Vollständige API-Rohantworten und Bilder sind keine PDF-Anlagen; ihre Rekonstruktion beziehungsweise eine neue Inferenz erfordert die angegebenen Quellen und den Benchmark-Code.

Dataset-Revision: ed46e02b9b1136f6b54847c1d4bce9e94d11e55c

Lauf: azure-d1a40be654e0b2d5

SHA-256 der Kennzahlendatei: 972f7d517d7eeca5fe3592038e17b5233cfb6a6430ac853ce042555ad77c857

Lokale Revisionen: GLM-OCR 2e85a62840ccac27daa451df36c736c4636b8628; PaddleOCR-VL c5630abae1d940eafe0697512a0325494b02ab42. PP-OCRv6 nutzt medium_det und medium_rec; Artefakthashes und Paketversionen stehen in modellumgebungen.json. GPT-Antworten melden 5.6 Sol/Terra vom 09.07.2026, 6 Sol vom 22.09.2026, 6.1 Sol vom 29.09.2026 und Nano vom 07.08.2025. Phi meldet lediglich „vision“; Deploymentnamen sind daher keine vollständige unabhängige Versionsverifikation.

Fester Prompt der Chatmodelle (unverändert):

Transcribe all printed text in this receipt image exactly as written, in reading order. Keep each receipt row together, including product name and its price on the same line. Preserve spelling, numbers, punctuation, currency and line breaks. Do not correct, translate, summarize or infer missing information. Return only the plain receipt text, without Markdown, JSON or commentary.

Literatur und Datenquellen

- [1] albertobarnabo. synthetic-receipts-ocr. Dataset-Karte und Datenrevision. Zugriff/Preisprüfung: 16.10.2026.
<https://huggingface.co/datasets/albertobarnabo/synthetic-receipts-ocr>
- [2] PaddlePaddle. PP-OCRv6_medium_rec. Offizielle Modellkarte. Zugriff/Preisprüfung: 16.10.2026.
https://huggingface.co/PaddlePaddle/PP-OCRv6_medium_rec
- [3] PaddlePaddle. PaddleOCR-VL-1.6. Offizielle Modellkarte. Zugriff/Preisprüfung: 16.10.2026.
<https://huggingface.co/PaddlePaddle/PaddleOCR-VL-1.6>
- [4] Z.ai. GLM-OCR. Offizielle Modellkarte. Zugriff/Preisprüfung: 16.10.2026.
<https://huggingface.co/zai-org/GLM-OCR>
- [5] Microsoft. Receipt data extraction: Document Intelligence, API 2024-11-30. Zugriff/Preisprüfung: 16.10.2026.
<https://learn.microsoft.com/en-us/azure/ai-services/document-intelligence/prebuilt/receipt?view=doc-intel-4.0.0>
- [6] Microsoft. Azure OpenAI Service: öffentliche Referenzpreise. Zugriff/Preisprüfung: 16.10.2026.
<https://azure.microsoft.com/en-us/pricing/details/azure-openai/>
- [7] Microsoft. Mistral AI in Foundry: öffentliche Referenzpreise. Zugriff/Preisprüfung: 16.10.2026.
<https://azure.microsoft.com/en-us/pricing/details/ai-foundry-models/mistral-ai/>
- [8] Microsoft. Azure Document Intelligence: öffentliche Referenzpreise. Zugriff/Preisprüfung: 16.10.2026.
<https://azure.microsoft.com/en-us/pricing/details/document-intelligence/>
- [9] Microsoft. Microsoft models in Foundry: öffentliche Referenzpreise. Zugriff/Preisprüfung: 16.10.2026.
<https://azure.microsoft.com/en-us/pricing/details/ai-foundry-models/microsoft/>
- [10] OpenAI. GPT-6.1 Sol: Modell- und Reasoning-Dokumentation. Zugriff/Preisprüfung: 16.10.2026.
<https://developers.openai.com/api/docs/models/gpt-6.1-sol>