



Generative KI als Unterstützung in einem digitalen bibelwissenschaftlichen Projekt

Nicole Oesterreich^a und Kathrin Väterlein^b

^a  <https://orcid.org/0000-0002-2412-5727>, ^b  <https://orcid.org/0009-0004-0946-6891>

Abstract Wie generative KI sinnvoll für ein digitales bibelwissenschaftliches Projekt eingesetzt werden kann, erläutert dieser Werkstattbericht. Er behandelt vor allem die Themen OCR und deren Nachbearbeitung sowie Übersetzungen verschiedener, darunter vor allem antiker Sprachen unter Einsatz bestimmter LLMs. Dabei werden Vor- und Nachteile beschrieben sowie Ergebnisse von Tasks verschiedener LLMs verglichen.

Keywords Large Language Models, generative AI, translation, Biblical Studies, ancient languages

1. Einleitung

Das Corpus Judaeo-Hellenisticum (CJH) ist eine Sammlung aller erhaltenen Zeugnisse des frühen Judentums (300 v. Chr. bis 200 n. Chr.) unter hellenistisch-römischem Einfluss. Das CJHNTdigital-Projekt¹ zielt darauf ab, diese Quellen digital zu dokumentieren, in Teilen neu zu übersetzen und zu kommentieren sowie ihre Bedeutung für die Interpretation des Neuen Testaments zu analysieren. Es entsteht eine mehrsprachige digitale Forschungsplattform, die sowohl bestehende Datenbanken integriert als auch Schnittstellen für zukünftige Forschung an weiteren antiken Quellen bereitstellt.

Für die Bearbeitung von allen neutestamentlichen Schriften sind die finanzierten 15 Jahre, auch wenn sich ab einer bestimmten Stelle einiges wiederverwenden lassen wird, nicht allzu reichlich bemessen. Der Projektbeginn fiel in eine Zeit, in der generative KI in einer Qualität vorhanden war, die für das Projektteam sinnvoll war. Somit wurde die Nutzung von Anfang an mit in die Überlegungen einbezogen. Generative KI „refers to a subset of artificial intelligence models and techniques that are designed to generate new data samples that are similar in nature to a given set of input data. The goal is to produce content or data that wasn't part of the original

1 Mehr Informationen zum Projekt: <https://cjhnt-info.saw-leipzig.de/de> (abgerufen am 26.06.2025).

training set but is coherent, contextually relevant, and in the same style or structure“ (Campeato 2024, 1).

Large Language Models (LLM) gibt es als offene oder geschlossene Modelle. Offene Modelle beruhen auf open source code, sind meist für geringere Rechenleistungen konzipiert, können daher lokal implementiert werden und sind beliebig anpassbar. Das ist insbesondere bei sensiblen Daten die bessere Wahl. Offene Modelle setzen jedoch umfangreiches technisches Wissen sowie strukturelle Unterstützung durch das lokale Rechenzentrum der eigenen Einrichtung voraus (Alammar/Grootendorst 2024, 34f.). Denn auch diese Modelle benötigen Rechenleistung, die weit über die eines normalen Laptops hinaus geht. Beide Voraussetzungen waren beim CJHNTdigital nicht gegeben, werden aber angestrebt.

Bei der Auswahl eines geschlossenen Modells spielt die Art der Trainingsdaten eine Rolle sowie die darin verwendeten Sprachen, denn die allermeisten Modelle sind hauptsächlich mit englischsprachigen Quelldaten trainiert. Zentral ist außerdem die Größe des Kontextfensters, also die Anzahl von Wörtern oder Tokens aus dem zurückliegenden Chat oder einem Datenset, die ein Modell bei der Generierung seiner Antwort einbeziehen kann. Alle Modelle halluzinieren in unterschiedlichem Ausmaß (Kamath et al. 2024, 219–221), was wiederum vom jeweiligen Finetuning sowie der Implementation von Retrieval Augmented Generation (RAG) abhängt und mittels des Prompts beeinflusst werden kann (Alammar/Grootendorst 2024, 130). RAG ist die Möglichkeit dem LLM eine eigene Datenbasis speziell für einen Task zur Verfügung zu stellen, die über das Kontextfenster hinaus geht (Kamath et al. 2024, 275f.). Nach mehreren Tests mit unterschiedlichen Modellen fiel die Wahl des CJHNTdigital-Teams auf Anthropic Claude. Das Modell Sonnet 3.5 arbeitet mit einem Kontextfenster von 200.000 Tokens² und verarbeitet Griechisch, Hebräisch und Latein korrekt. Mittlerweile gibt Anthropic die Möglichkeit RAG für Claude zu nutzen.³

Die ethischen (Kamath et al. 2024, 229–233.247f.256f.; Alammar/Grootendorst 2024, 32) und ökologischen (Alammar/Grootendorst 2024, 32) Schwierigkeiten von LLMs sind dem Team bewusst, sollen aber in diesem Werkstattbericht keine Rolle spielen.⁴

2 <https://www.anthropic.com/news/claude-3-5-sonnet> (abgerufen am 05.03.2025).

3 <https://support.anthropic.com/de/articles/11473015-retrieval-augmented-generation-rag-fur-projekte> (abgerufen am 12.06.2025).

4 Eine Preprint-Studie kam zu dem Schluss, dass es bei den geschlossenen LLMs große Unterschiede beim ökologischen Fußabdruck gibt, wobei die Modelle von Anthropic vergleichsweise gut abschneiden: Jegham et al. 2025.

2. Anwendungsbereiche

2.1 OCR (optical character recognition)

Eins der Projektziele ist es, frühjüdische Quellentexte in ihrer Überlieferungssprache auf einer digitalen Plattform zu präsentieren. Dabei kann auf die Arbeit von einigen Forschungsgenerationen zurückgegriffen werden. Jedoch sind lange nicht alle Quellentexte bereits digitalisiert.⁵ Das Team kann auf freie Daten von perseus.tufts.org sowie von qumran-digital.org zugreifen. Mit einem weiteren Forschungsteam, das sich mit der Digitalisierung frühjüdischer Quellen befasst, sind Gespräche aufgenommen worden. Jedoch bleiben auch nach der möglichen Übernahme dieser Daten reichlich undigitalisierte Quellen aus dem Corpus übrig. Diese sind nicht nur Griechisch, sondern in vielen Sprachen überliefert.

2.1.1 Altgriechisch

Nachdem OCR-Versuche mit der *tesseract* und *OlmoOCR* mäßig erfolgreich, jedoch mit hohem Nacharbeitsaufwand verbunden waren (abgesehen von der äußerst komplexen Installation von weiteren Sprachen und der nicht gerade anwenderfreundlichen fehlenden GUI bei *tesseract*), startete das Team erste Versuche von griechischsprachigem OCR mit *Claude Sonnet 3.5*. Diese verliefen äußerst zufriedenstellend. *Claude 3.5* erkennt nicht nur nahezu alle Akzente richtig, es verbessert auch je nach Prompt mögliche Fehler des OCR (allerdings auch Fehler in den Editionen, die gegebenenfalls sichtbar bleiben sollen). Mit einer sehr geringen Fehlerquote, die sich üblicherweise im Bereich der Akzente finden (bei entsprechend guter Qualität des vom Team bereit gestellten Scans), überbietet *Claude* die Kombination von *tesseract* und studentischer Hilfskraft nicht nur in der Geschwindigkeit, sondern auch in der Qualität. Hier ein Beispiel aus der *ApkMose 29*, den *Claude* fehlerfrei erkannt hat:

5 Die ursprüngliche Absicht mit einem Hersteller von Konkordanzprogrammen zusammen zu arbeiten, ist als gescheitert anzusehen. Zu OCR und Digitalisierung vgl. Stökl Ben Ezra 2024.

Please extract and review the ancient Greek text.	
<p>³ ἀποκριθεὶς δὲ ὁ πατὴρ ὑμῶν εἶπεν τοῖς ἀγγέλοις· ἰδοὺ, ἐκβάλλετε με. δέομαι ὑμῶν, ἄφετέ με ἄραι εὐωδίας ἐκ τοῦ παραδείσου, ἵνα μετὰ τὸ ἐξελθεῖν με ἀνεπέγκω θυσίαν τῷ θεῷ, ὅπως εἰσακούσεται μου ὁ θεός.</p> <p>⁴ καὶ προσελθόντες εἶπον οἱ ἄγγελοι τῷ κυρίῳ· Ἰαὴλ αἰώνι βασιλεῦ, κέλευσον δοθῆναι τῷ Ἀδὰμ θυμιάματα εὐωδίας ἐκ τοῦ παραδείσου.</p>	<p>3 ἀποκριθεὶς δὲ ὁ πατὴρ ὑμῶν εἶπεν τοῖς ἀγγέλοις· ἰδοὺ, ἐκβάλλετε με. δέομαι ὑμῶν, ἄφετέ με ἄραι εὐωδίας ἐκ τοῦ παραδείσου, ἵνα μετὰ τὸ ἐξελθεῖν με ἀνεπέγκω θυσίαν τῷ θεῷ, ὅπως εἰσακούσεται μου ὁ θεός.</p> <p>4 καὶ προσελθόντες εἶπον οἱ ἄγγελοι τῷ κυρίῳ· Ἰαὴλ αἰώνι βασιλεῦ, κέλευσον δοθῆναι τῷ Ἀδὰμ θυμιάματα εὐωδίας ἐκ τοῦ παραδείσου.</p>

2.1.2 Latein

Gleiche Erfahrungen hat das Team mit lateinischen Texten gemacht. Hier gab es einen direkten Vergleich mit der freien Version von ChatGPT 4-turbo. Nicht nur hat ChatGPT 4 mehr Fehler gemacht, sondern es hatte auch wesentlich mehr Halluzinationen: Ein Problem, das bei Claudes Sonnet 3.5 in der OCR-Anwendung bisher kaum auftrat. Getestet wurde ein Extrakt aus LAB 3. Die Vorlage war eine Ausgabe der Kisch-Edition von 1949 (Kisch 1949), die zahlreiche Annotationen mit Bleistift enthält. Beide LLMs erhielten den gleichen, bereits auf ChatGPTs Ergebnissen angepassten Prompt (s. Appendix), der die Extraktion des Textes mit einer speziellen Formatierung für Kapitel- und Verszahlen verband und verhindern sollte, dass v statt u genutzt wird.

ChatGPT machte in der freien Version ca. 30 Fehler, wobei die ersten zwei Zeilen zunächst komplett fehlten. Diese konnten per Prompt nachgefordert werden. Claudes Text enthielt 8 u-Fehler, die jedoch am Prompt (s. Appendix) lagen, der an ChatGPT angepasst war. Der einzige tatsächliche Fehler war die Auslassung von *desiit*. Beide LLMs haben die Bleistiftannotationen erfolgreich ignoriert und die Formatierungsanweisungen befolgt. Getestet wurde im Vergleich mit freien online Tools,⁶ die kein brauchbares Ergebnis lieferten und die Bleistiftannotationen übernahmen. Tesseract erzielt hier das beste Ergebnis, jedoch bleibt das Ergebnis weit unter dem von ChatGPT. Claude liefert somit mit Abstand die besten Ergebnisse für lateinisches OCR aus modernen Editionen.

⁶ Z. B. <https://www.i2ocr.com/free-online-latin-ocr> (abgerufen am 26.06.2025) oder <https://conversion.co/de/ocr/latin/> (abgerufen am 26.06.2025).

2.1.3 Geʿez (Altäthiopisch)

Geʿez gehört zu den Sprachen, für die nur sehr wenige Trainingsdaten für LLMs existieren. Jedoch sind LLMs auf eine bestimmte Menge an Trainingsdaten angewiesen, um ausreichend gute Übersetzungen zu erstellen oder die Ergebnisse von OCR zu kontrollieren. Grundsätzlich gibt es die Möglichkeit, ein spezialisiertes LLM mit relativ wenigen Daten zu trainieren, insbesondere bei Sprachen, die andere verwandte Sprachen haben (Wassie 2023, 1). Ein solches vortrainiertes Modell wäre EthioLLM von der Gruppe Ethio NLP um Atnafu Lambebo Tonja.⁷ Dies konnte im CJHNTdigital noch nicht getestet werden. Die folgenden Erläuterungen gelten für Claude 3.7 im Vergleich mit einem spezialisierten OCR-Tool.

Bei der OCR von Geʿez zeigt sich Claude schlechter als mit auf Fidäl-spezifizierter OCR-Software (Tigrinya, Amharisch, Geʿez).⁸ Eine der Schwierigkeiten ist die relative Ähnlichkeit der Buchstaben. Die IV. Ordnung beim Verb wird meist durch ein kürzeres „Bein“ dargestellt (z. B. አ anstelle von ለ), was in einigen Editionen auch für menschliche Augen mitunter nur mit Mühe zu unterscheiden ist.

Versuche mit Fidäl in Sonnet 3.7 erzielten schlechte Ergebnisse (s. Appendix): Dem LLM wurde ein Bild der Fleming-Edition von 1Hen 1,1–6 gegeben mit der Bitte den äthiopischen Text zu extrahieren. Nach ein paar Zeilen Informationen über den Text empfahl Claude 3.7 die Hinzuziehung eines Spezialisten, was durchaus positiv zu bewerten ist. Nachdem es gebeten wurde, trotzdem eine Extraktion zu versuchen, gab es zusammen mit einer weiteren Warnung Text aus. Von den 418 Zeichen im Text hat Claude 3.7 79 Zeichen falsch oder nicht gedeutet, was eine Fehlerquote von 19 % ergibt. Im Tool von metaappz.com gab es 12 Fehler, die sich jedoch ausschließlich auf fehlende Satzzeichen und ein lateinisches Zeichen beziehen (macht eine Fehlerquote von ca. 3 %). Rechnet man bei beiden Versuchen ohne Satzzeichen, ergibt sich für Claude 3.7 eine Fehlerquote von 21% und beim Tool von metaappz.com von 0,3 % (bei 343 Zeichen ohne Satzzeichen). Eine Fehleranalyse für die Ausgabe von Claude 3.7 zeigt keine sich wiederholenden Fehler. Es sollte deshalb nicht für Geʿez-Texte verwendet werden.

Ein spezialisiertes LLM, das Fidäl-OCR sowie Übersetzung beherrschen würde, möglicherweise auch von den normalerweise sehr gut lesbaren äthiopischen Manuskripten, wäre eine enorme Bereicherung auch für die neutestamentliche Forschung. Viele frühchristliche Texte sind in altäthiopischen Übersetzungen überliefert (z. B. die Apokalypse des Petrus sowie zahlreiche Apostelakten) und womöglich warten in den mittlerweile dank großartiger Forschungsbemühungen (u. a. vom Akademieprojekt

⁷ <https://ethionlp.github.io/index.html> (abgerufen am 26.06.2025).

⁸ ChatGPT besitzt nach eigener Auskunft keine Möglichkeit des OCR für Fidäl aufgrund der fehlenden Spracherkennung von Geʿez, aber auch der modernen äthiopischen Sprachen mit Fidäl: Tigrinya, Amharisch und Tigre.

Beta Masaheft)⁹ gescannt und damit für die Nachwelt bewahrten Manuskripten unbekannte Texte.¹⁰

Weitere Versuche gab es außerdem mit Syrisch und Altkirchenslawisch. Alle drei Sprachen kann Claude nach eigener Auskunft nicht, sodass die Verbesserung der erkannten Texte durch das LLM wegfällt. Sowohl bei Syrisch als auch bei Altkirchenslawisch macht die OCR-Erkennung sehr typische und wiederkehrende Fehler (Verwechslung der gleichen Buchstaben).

2.2 Übersetzungen

Im ersten dreiviertel Jahr des Projektes wurde zunehmend klarer, dass sehr viel Zeit in die Neuübersetzungen der Texte fließt, wobei hier insbesondere Philo ein besonderes Schwergewicht darstellt.¹¹ Dazu kommt, dass sich die Projektleiter Jens Herzer, Karl-Wilhelm Niebuhr, Roland Deines und Christfried Böttrich entschieden haben, das Projekt in Deutsch und Englisch umzusetzen. Zu den altsprachlichen Übersetzungen¹² kommen also noch solche der deutschen Kommentierungen ins Englische.

2.2.1 Alte Sprachen

Zunächst sollen die grundlegenden Kompetenzen der LLMs bei den Übersetzungen der drei alten Sprachen aufgezeigt werden. Getestet wurde zunächst die Übersetzungskompetenz in Latein von ChatGPT und Claude 3.5 im Vergleich mit dem oben genutzten Beispiel aus dem Liber Antiquitatum Biblicarum 3. In den Übersetzungen bestehen kaum Unterschiede. ChatGPT bekam dafür den korrigierten lateinischen Text (s. Appendix):

et obtulit holocaustomata in ara, et acceptum est Domino, tamquam odor requietionis. {9} Et dixit Deus: Non adiciam iam maledicere terram pro homine, quoniam figura cordis hominis desiit a iuventute sua, et ideo non adiciam simul disperdere omnes viventes sicut feci. Erit autem cum peccaverint habitantes terram, diiudicabo eos in fame sive in gladio, sive in igne, sive in morte, et erit terremotus et dispergentur in inhabitabilia.

9 <https://www.betamasaheft.eu/> (abgerufen am 26.06.2025).

10 Der Bürgerkrieg in Äthiopien, der von Ende 2020 bis zum Friedensabkommen 2022 um die Region Tigray ausgefochten wurde, machte auch vor Klöstern und ihren Bibliotheken nicht halt. Der Frieden ist keineswegs stabil. Touati/(Chandler) de Waal 2024.

11 Die „Philo of Alexandria Commentary Series“ von Greg Sterling und David T. Runia leistet hier hervorragende Unterstützung. Jedoch sind bisher nur acht Bände des umfangreichen Projekts erschienen.

12 Weitere Sprachen kommen eventuell noch hinzu.

Und er brachte Brandopfer auf dem Altar dar, und es wurde vom Herrn angenommen wie ein Duft der Ruhe. {9} Und Gott sprach: Ich werde die Erde nicht mehr um des Menschen willen verfluchen, denn die Neigung des menschlichen Herzens weicht nicht von seiner Jugend an, und deshalb werde ich nicht mehr alle Lebewesen zugleich vernichten, wie ich es getan habe. Es wird aber geschehen, wenn die Bewohner der Erde gesündigt haben, werde ich sie richten – mit Hunger oder mit dem Schwert, mit Feuer oder mit dem Tod. Und es wird ein Erdbeben geben, und sie werden in unbewohnbare Gegenden zerstreut werden.

ChatGPT übersetzte *desiit* (Z. 2) fälschlicherweise präsentisch. Gibt man ChatGPT den Kontext hinzu, klingt die Übersetzung stark lutherdeutsch (s. Appendix).¹³ Gegebenenfalls könnte man über komplexere Prompts mit Hinweis auf grammatikalische Genauigkeit noch bessere Übersetzungen erreichen.

Dies ist die Übersetzung von Claude 3.5:

Und er brachte Brandopfer auf dem Altar dar, und es war dem Herrn wohlgefällig, wie ein Duft der Ruhe. {9} Und Gott sprach: Ich werde die Erde nicht mehr verfluchen um des Menschen willen, weil die Gestalt des menschlichen Herzens von seiner Jugend an aufgehört hat, und deshalb werde ich nicht alle Lebewesen gleichzeitig vernichten, wie ich es getan habe. Wenn aber die Bewohner der Erde sündigen werden, werde ich sie richten mit Hunger oder mit dem Schwert, oder mit Feuer, oder mit Tod, und es wird ein Erdbeben sein, und sie werden in unbewohnbare Gebiete zerstreut werden.

Claude lässt *erit* (Z. 4) unübersetzt, was möglich ist. Claudes Übersetzung ist gegenüber der von ChatGPT also fehlerfrei. Beide Tools bieten bei den Übersetzungen beeindruckende Leistungen. Sie sind nicht nur weitgehend korrekt, sondern auch sprachlich gut. Hingewiesen auf die Konjektur von Cohn (*desipit* statt *desiit*) ergänzen beide den korrekten deutschen Satz: „weil die Gestalt des menschlichen Herzens von seiner Jugend an töricht ist“. Die LLMs wurden außerdem befragt, was Besonderheiten gegenüber dem Standardlatein sind. Die Antworten fielen bis auf die Feststellung der fehlenden Diphthongierung komplett unterschiedlich aus. Sie sind daher ohne eigene Kenntnisse in dem Bereich wenig brauchbar. Hier sind spezialisierte NLP-Tools die sinnvollere Wahl. Die LLMs kamen trotzdem auf eine korrekte Datierung der

13 Hingewiesen auf den Fehler bei *desiit* antwortete ChatGPT, dass es seine Übersetzung jeweils an die Formulierungen des Alten Testaments (scheinbar in der Lutherübersetzung) angeglichen hatte.

lateinischen Übersetzung zwischen dem 3. und 5. Jahrhundert¹⁴ aufgrund der Ähnlichkeit zu den Vetus Latina-Übersetzungen.¹⁵

Der Übersetzungsprozess sollte für solche Quellenabschnitte, die im Rahmen des Projekts kommentiert werden, nicht regelhaft durch die Übersetzung eines LLM ersetzt werden. So bildet die eingehende Beschäftigung mit Grammatik und Semantik des vorliegenden Textes die Grundlage für die spätere Interpretation. Vokabular, das im Frühjudentum eine besondere Bedeutungsnuance hatte, wird u.U. von Claude 3.5 nicht mit dieser übersetzt. Beim Prompten ist daher zu beachten, dass man das Modell auf den Kontext hinweist.¹⁶ Hilfreich ist ein solches LLM insbesondere bei bisher unübersetzten Texten oder solchen, wo eine Übersetzung nicht zur Hand ist, um sich einen Überblick über den Inhalt und den weiteren Kontext der Quelle zu verschaffen (z. B. auch Inschriften und Papyri).

Der Übersetzungsprozess im Projekt sah die Überprüfung der Übersetzungen von Teammitgliedern durch einen Altphilologen vor. Dies kann Claude 3.5 übernehmen:

Prompt: „Bitte korrigiere die deutsche Übersetzung des griechischen Philotextes. Ignoriere die Latex-commands. $\begin{longtable} \{l\} \{p\} \{34em\} \} 114$ & $\text{textgreek}\{\rho\omicron\upsilon\ \gamma\acute{\alpha}\rho\ \mu\omicron\upsilon\ \tau\acute{o}\ \sigma\tilde{\omega}\mu\alpha\ \pi\rho\acute{o}\ \gamma\epsilon\nu\acute{\epsilon}\sigma\epsilon\omega\varsigma\ \eta\tilde{\nu};\ \rho\omicron\tilde{\iota}\ \delta\acute{\epsilon}\ \kappa\alpha\iota\ \chi\omega\rho\acute{\eta}\sigma\epsilon\iota\ \mu\epsilon\tau\alpha\sigma\tau\acute{\alpha}\nu\tau\omicron\varsigma\ \mu\omicron\upsilon;\}$ || [...] Wo ist der Säugling, wo das Kind, wo der kleine Junge, wo der kürzlich zum jungen Mann gewordene, wo der Jugendliche mit dem ersten Bartflaum, der junge Mann, der erwachsene Mann? || & Woher kam die Seele, wohin wird sie sich entfernen, wie lange Zeit wird sie mit uns zusammenleben? || & Was aber ist die Existenz, sind wir imstande [sie] zu beschreiben?|| [...]“

Claude 3.5: „[...] Wo ist der Säugling, wo das Kind, wo der Heranwachsende, wo der gerade Erblühende, wo der Jüngling, der mit erstem Bartflaum, der junge Mann, der erwachsene Mann? Woher kam die Seele, wohin wird sie sich entfernen, wie lange Zeit wird sie mit uns zusammenleben? Was ist sie ihrem Wesen nach, können wir sagen? [...]“

- 14 Kisch zitiert Cohn, der aufgrund der sprachlichen Eigenheiten der lateinischen Übersetzung als späteste Möglichkeit das 4. Jahrhundert nennt (Kisch 1949, 18).
- 15 Hier sollte man freilich mit allergrößter Vorsicht vorgehen und sich keinesfalls auf derlei Angaben verlassen. LLMs sind weiterhin keine Recherche-Tools.
- 16 LLMs werden auf dem Gebiet der Übersetzungen in der Regel durch ein Judge-Evaluator-System (Kamath et al. 2024, 350) mit einem Human-in-the-Loop (Kamath et al. 2024, 359) trainiert und weiterentwickelt. Ein solcher Prozess könnte bei einem selbst trainierten offenen Modell signifikant bessere Ergebnisse liefern, die sowohl grammatikalisch korrekt sind als auch semantisch dem Kontext entsprechen. Aber auch mit geschlossenen LLMs kann man im Laufe eines Chats und Hinweisen der nutzenden Person einige Verbesserungen erreichen.

Claude wurde danach angewiesen, seine Änderungen anzugeben, um diese prüfen zu können. Nach Übernahme bestimmter Korrekturen wurde per Prompt einerseits der LaTeX-Code wieder hinzugefügt, andererseits eine möglichst genaue englische Übersetzung mit Blick auf den griechischen Text angefordert, sodass die deutsche und die englische Übersetzung des Philo-Textes den gleichen Sinn haben. An einer Stelle sollte Claude 3.5 die englische Übersetzung präziser fassen:

Prompt: Geht das noch genauer: What is it in its essence, can we say?

Claude 3.5: „A more precise translation of „τίς δέ ἐστι τὴν οὐσίαν, ἔχομεν εἰπεῖν;“ would be: „And what is it in respect to its substance/essence, are we able to say?“ The Greek phrase uses „τὴν οὐσίαν“ (tēn ousian) as an accusative of respect, indicating the specific aspect about which the question is being asked. The term „οὐσία“ (ousia) in philosophical Greek texts specifically refers to the fundamental nature, substance, or essence of something. The construction with „ἔχομεν εἰπεῖν“ (echomen eipein) is an idiomatic way of asking about capability or possibility („are we able/capable to tell/say“).“

Die korrigierte Übersetzung konnte nach der Auswahl eines der beiden Substantive zur Auswahl übernommen werden.

Eine weitere Spezialaufgabe für Claude 3.5 sind Überlegungen, welche Begriffe ein potentieller, aber nicht erhaltener griechischer Originaltext enthalten haben könnte. Claude 3.5 übernimmt hier Rückübersetzungen aus anderen Sprachen ins Griechische inklusive der Angabe von möglichen alternativen Begriffen, die durch die bearbeitende Person überprüft und weiterentwickelt werden können.

Mit Claude 3.5 wird bei den griechischen, lateinischen und hebräischen Übersetzungen Projektzeit gespart, denn der Altphilologe kann sich syrischen Übersetzungen sowie der Unterstützung des Teams bei komplexeren Übersetzungsaufgaben widmen und die Bearbeitenden können die gesparte Zeit auf die Kommentierung verwenden.

2.2.2 Moderne Sprachen

Bei Übersetzungen ins Englische ist Claude 3.5 eine wertvolle Hilfe, insbesondere, wenn der Prompt die Eingrenzung des Zielstils (akademisches Englisch) sowie einen Arbeitsschritt der Verbesserung des Ausdrucks enthält. Das Ergebnis ist grammatikalisch korrekt, enthält jedoch an einigen Stellen Formulierungen, die für Muttersprachler*innen ungewohnt oder im akademischen Raum unangebracht sind. Hier

ist also ebenfalls eine Überprüfung durch eine*n Muttersprachler*in unerlässlich.¹⁷ Sehr gute Ergebnisse erzielt Claude bei der Verbesserung des englischen Ausdrucks von durch Menschen verfassten Fachtexten in englischer Sprache.

Weiterhin wurden Übersetzungen aus italienischen, spanischen und französischen Aufsätzen vorgenommen, um sie wissenschaftlich rezipieren zu können, was eine große Unterstützung bei der Diversifizierung der Forschung darstellt. Hier könnten in Zukunft auch noch weitere Sprachen dazu kommen, um beispielsweise die orthodoxe Forschung zum 2. Henochbuch abzudecken.

2.3 Umgang mit deutschem Text

2.3.1 Korrekturen nach OCR mit anderen Tools

Von zahlreichen Texten des CJH existieren deutsche Übersetzungen, die auf Grund ihres Alters bereits gemeinfrei sind. Diese sind jedoch selten digitalisiert bzw. liegen lediglich als Scan oder stark fehlerhafte plain texts vor (z. B. bei archive.org, erstellt mit dem dort verwendeten OCR-Tool). Das Team benötigt für die Umwandlung in XML-Dateien eine TXT-Datei, in der Kapitel und Versnummern jeweils mit doppelt oder einfach geschweiften Klammern gefasst sind. Der Text sollte am Ende möglichst fehlerfrei sein. Fehler in der Edition sollen übernommen werden.

Es folgt ein Beispiel aus De Opificio Mundi übersetzt von Leopold Cohn (Cohn 1909). Der Text wurde vorbereitet, indem die am Rand stehenden Versnummern mit dem Original abgeglichen, eventuell weitere auftretende Ziffern am Ende oder Anfang einer Zeile sowie sämtliche Fußnoten und Kopfzeilen entfernt wurden. Das folgende Beispiel lieferte Claude 3,5 nach bereits umfassender Arbeit an diesen Cohn-Digitalisaten. Der verwendete Prompt lautete wie folgt:

„Bitte entferne aus dem folgenden Text Zahlen und überflüssige Zeichen, die nicht in den deutschen Text gehören. Bitte behalte alle Angaben in geschweiften Klammern. Wenn eine Zahl ohne Klammern am Anfang oder Ende einer Zeile steht, dann verschiebe diese Zahl vor den Satzbeginn in dieser Zeile und setze sie in geschwungene Klammern.“

17 Die Problematik, die das bei professionellen Übersetzenden verursacht, ist dem Projektteam sehr bewusst. Allerdings sind auch Zeit und Mittel eines Drittmittel-geförderten Projekts endlich, so dass das LLM die Möglichkeit bietet den straffen Zeitplan zu schaffen.

txt-Datei aus Archive.org	Durch Hilfskraft vorbereiteter Text	Von Claude bearbeitete Version
<p>(31.) Nun müssen wir auch zur anderen Art der Siebenzahl 95 übergehen, die sich innerhalb der Zehnzahl befindet und eine merkwürdige und ebenso bedeutsame Beschaffenheit zeigt wie die zuerst genannte Art. Die 7 besteht nämlich aus 14244, die zwei sehr harmonische Verhältnisse bilden, nämlich das zweifache (2:1) und das vierfache (4:1); das zweifache Verhältnis stellt aber die Konsonanz da rao@v (Oktave) dar, das vierfache die Konsonanz dis 64 rasav (Doppeloktave).</p>	<p>{{XXXI}} Nun müssen wir auch zur anderen Art der Siebenzahl 95 übergehen, die sich innerhalb der Zehnzahl befindet und eine merkwürdige und ebenso bedeutsame Beschaffenheit zeigt wie die zuerst genannte Art. Die 7 besteht nämlich aus 14244, die zwei sehr harmonische Verhältnisse bilden, nämlich das zweifache (2:1) und das vierfache (4:1); das zweifache Verhältnis stellt aber die Konsonanz da rao@v (Oktave) dar, das vierfache die Konsonanz dis 64 rasav (Doppeloktave).</p>	<p>{{XXXI}} {95} Nun müssen wir auch zur anderen Art der Siebenzahl übergehen, die sich innerhalb der Zehnzahl befindet und eine merkwürdige und ebenso bedeutsame Beschaffenheit zeigt wie die zuerst genannte Art. Die 7 besteht nämlich aus 1+2+4, die zwei sehr harmonische Verhältnisse bilden, nämlich das zweifache (2:1) und das vierfache (4:1); das zweifache Verhältnis stellt aber die Konsonanz Oktave dar, das vierfache die Konsonanz Doppeloktave.</p>

Claude 3.5 kann überflüssige Zahlen im Text, Randnotizen (z. B. p- 15M.) und Fehler von gewünschten Inhalten wie Bibelstellen (z. B. 1 Mos. 9, 25) unterscheiden. Es hat sich als notwendig erwiesen, dem Prompt die Bemerkung hinzuzufügen, dass der Text vollständig und ohne Änderungen übernommen werden soll. Claude prüft eigenständig Zahlen bzw. Rechnungen im Text auf ihre Sinnhaftigkeit und ergänzt, wo nötig, korrekt. Lediglich die im Original enthaltene griechische Transkription, die bereits in der Ausgangs-Textdatei nicht lesbar war, entfällt nun ersatzlos (Original: „das Verhältnis 4:3 ergibt die erste Harmonie $1 \frac{1}{3}$, die $\delta\acute{\iota}\alpha\ \tau\epsilon\sigma\acute{\sigma}\acute{\alpha}\rho\omega\nu$ [Quart] heisst.“).

In einem Fall halluzinierte Claude 3.5 eine Fortsetzung des Textes, obwohl die Vorlage bereits zu Ende war und produzierte eine Art pseudepigraphische Fortführung des Philotextes, die so gut zur Vorlage passte, dass es zunächst unbemerkt blieb. Eine Korrektur durch eine Hilfskraft mit entsprechenden inhaltlichen Kenntnissen ist also notwendig.

Jede Bearbeitungssitzung mit Claude 3.5 verläuft etwas anders. Die benutzende Person muss daher stets beobachten und wenn nötig reagieren, um möglichst akkurate Ergebnisse zu erzielen. Insgesamt kann das LLM diese Arbeit sehr viel schneller und gründlicher erledigen als eine Hilfskraft das könnte.

2.4 Coden und Debuggen

Trotz einer IT-Vollzeitkraft wird Claude 3.5 im Projekt zum Coden oder Debuggen eingesetzt. So ist mit Hilfe des LLM ein Visual Basic Makro entstanden, um große Word-dateien automatisch in Einzeldateien aufzuteilen, die außerdem einen bestimmten Dateinamen bekamen. Die Entwicklung erfolgte mit mehreren Debugging-Schritten bis das Makro fehlerfrei lief. Das LLM wurde dabei jeweils auf die Fehlermeldungen des VB-Programms von MS Word hingewiesen oder auf Fehler bei der Endausgabe. Die Entwicklung erfolgte durch eine Person, die VB zwar nicht selbst von Grund auf schreiben, jedoch lesen und verstehen kann, sodass ein potentieller Eintrag von schädlichem Code aufgefallen wäre.

Für den Buchsatz wird im Projekt LaTeX verwendet. Dabei werden die XML-Dateien der Arbeitsplattform in LaTeX-Code umgewandelt und zu einer Satzdatei zusammengesetzt. Claude 3.5 hilft hier bei der Ergänzung von LaTeX-Code und sucht die Fehlerquelle von LaTeX-Fehlermeldungen, die oft nicht leicht zu interpretieren sind.

Beispielhaft war die Suche nach einem Fehler in der Präambel des LaTeX-Dokuments, der dazu führte, dass syrischer Text nicht korrekt dargestellt wurde. Der Prompt beschrieb den Fehler und gab den Präambel-Code vor mit der Bitte nach dem Fehler zu suchen. Claude 3.5 gab folgende Antwort, die zur Lösung des Problems führte:

„I notice the issue in your LaTeX code. There’s a typo in your Syriac font declaration that’s likely causing the problem with the final and middle characters. In your code, you have: `\newfontfamily|syriacfont[Script=Syriac, Scale=1.1]{Estrangelo Edessa}` The issue is in the `Script=Syriac` parameter – it should be `Script=Syriac`. Here’s the corrected line: `\newfontfamily|syriacfont[Script=Syriac, Scale=1.1]{Estrangelo Edessa}`“

Bei komplexeren Fehlern ist Claude jedoch nur mäßig hilfreich bzw. benötigt es einen langen Chat mit umfassender Übermittlung der Logfiles, um den richtigen Fehler zu finden.¹⁸ Hier ist der fortgeschrittene Anwender in der Regel schneller bei der Befragung einschlägiger Foren.

3. Zusammenfassung

Richtig eingesetzt kann generative KI im Bereich der Bibelwissenschaften vor allem an den Stellen viel Zeit sparen bzw. überhaupt Ergebnisse produzieren, an der an Digitalisierung und Textverbesserung sowie Übersetzungen gearbeitet wird (McGrath

¹⁸ Hier wäre die Verwendung von Overleaf mit integriertem LLM eine weitere Möglichkeit.

2025, 132). Zu beachten ist die Auswahl des richtigen Tools und Modells für die anstehende Aufgabe, denn die Fähigkeiten unterscheiden sich bei unterschiedlichen Modellen des gleichen Herstellers. Aufgrund des christlichen und zum Teil antijudaistischen Bias, den man mit der Verwendung von RAG zu einem bestimmten Maße dämpfen kann, sowie häufig fehlenden Kontexten sollten LLMs jedoch keinesfalls ohne Prüfung von Fachpersonen für zu publizierende Übersetzungen verwendet werden. Ein LLM ist für einfache Formatierungs- und Korrekturaufgaben sinnvoll einzusetzen und spart hier Zeit, damit Forscher*innen sowie Hilfskräfte komplexere Tätigkeiten übernehmen können. Die Übersetzungen in modernen Sprachen schaffen Möglichkeiten der Rezeption von Forschung, die mangels Sprachkenntnissen bisher nicht möglich war und ermöglicht Forscher*innen in kleinen Sprachgemeinschaften eine Teilhabe an der internationalen Forschungscommunity.

Ein Langzeitziel der europäischen Bibelwissenschaftler*innen in Verbindung mit Altphilologie und Althistoriker*innen sollte sein, ein europäisches auf alte Sprachen trainiertes LLM für die Forschungsgemeinschaft zu entwickeln, das unabhängig von ideologischer Einflussnahme und Gewinnstreben einzelner Firmen ist. Das Team von CJHNTdigital wird dafür gerne ihre Datensets zur Verfügung stellen.

4. Appendix

4.1 Zu 2.1.2

ChatGPT-4: „Bitte extrahiere den lateinischen Text ohne Fußnoten. Übernimm die Kapitel und Verszahlen. Gib die Kapitelnummern in arabischen Zahlen in doppelten geschweiften Klammern an, die Verszahlen in einfachen geschweiften Klammern. Nutze v anstelle von u in solchen Worten, wo beides möglich wäre. Korrigiere den lateinischen Text.“

Claude 3.5: „Bitte extrahiere den lateinischen Text ohne Fußnoten. Übernimm die Kapitel und Verszahlen. Gib die Kapitelnummern in arabischen Zahlen in doppelten geschweiften Klammern an, die Verszahlen in einfachen geschweiften Klammern.“

4.2 Zu 2.1.3

<p>Flemming-Edition</p>	<p style="text-align: center;">መጽሐፈ ሄኖክ።</p> <p>CAP. 1. ቃለ ፡ በረከት ፡ ዘሄኖክ ፡ በከመ ፡ ባረከ ፡ ጎሩያነ ፡ ወጻድቃነ ፡ እለ ፡ ሀለዉ ፡ ይኩኑ ፡ በዕለተ ፡ ምንዳቤ ፡ ለአሰሰሎ ፡ ኩሎ ፡ እኩያነ ፡ ወረሲዓን ። 2 ወአውሥአ ፡ እንከ ፡ ሄኖክ ፡ ወይቤ ፡ ብእሲ ፡ ጻድቅ ፡ ዘእምኅበ ፡ እግዚአብሔር ፡ [እንዘ ፡] አዕይንቲሁ ፡ ከሡታ 5 ት ፡ ወይሬኢ ፡ ራእየ ፡ ቅዱስ ፡ ዘበሰማያት ፡ ዘአርአዩኒ ፡ መላእክት፤ ወሰማዕኩ ፡ እምኅቤሆሙ ፡ ኩሎ ፡ ወአእመርኩ ፡ አነ ፡ ዘእሬኢ ፡ ወአኩ ፡ ለዛ ፡ ትውልድ ፡ አላ ፡ ለዘይመጽእ ፡ ርሕቅት። 3 በእንተ ፡ ጎሩያን ፡ እቤ ፡ ወአውሣእኩ ፡ በእንቲአሆሙ ፡ ምስለ ፡ ይወፅእ ፡ ቅዱስ ፡ ወዐቢይ ፡ እማኅደሩ፤ 4 ወአምላከ ፡ ዓለም ፡ [ወ]እምሆየ ፡ ይ 10 ከይድ ፡ ዲበ ፡ ሲና ፡ ደብር ፡ ወያስተርኢ ፡ በትዕይንቱ ፡ ወያስተርኢ ፡ በኅይለ ፡ ጽንዑ ፡ እምሰማይ። 5 ወይፈርሀ ፡ ኩሎ ፡ ወያድለቀልቁ ፡ ትጉሃን ፡ ወይነሥአሙ ፡ ፍርሀት ፡ ወረዓድ ፡ ዐቢይ ፡ እስከ ፡ አጽናፈ ፡ ምድር። 6 ወይደነግፁ ፡ አድባር ፡ ነዋኝን ፡ ወይቴሐቱ ፡ አ</p>
<p>Claude 3.7</p>	<p>መጽሐፈ ሄኖክ (The Book of Henok/Enoch)</p> <p>CAP. 1. ቃለ ፡ በረከት ፡ ዘሄኖክ ፡ በከመ ፡ ባረከ ፡ ጎሩያነ ፡ ወጻድቃነ ፡ እለ ፡ ሀለዉ ፡ ይኩኑ ፡ በዕለተ ፡ ምንዳቤ ፡ ለአሰሰሎ ፡ ኩሎ ፡ እኩያነ ፡ ወረሲዓን። 2 ወአውሥአ ፡ ሄኖክ ፡ ወይቤ ፡ ብእሲ ፡ ጻድቅ ፡ ዘእምጽባሕ ፡ አዕይንቲሁ ፡ [ሄኖክ] እዩድንተሁ ፡ ከሡተ 5 ተ ፡ ወይሬኢ ፡ ራእየ ፡ ቅዱስ ፡ ዘበሰማያት ፡ ዘአርአይዎ ፡ መላእክት ፡ ወሰማዕኩ ፡ እምኅሆሙ ፡ ኩሎ ፡ ወአእመርኩ ፡ አነ ፡ ዘእሬኢ ፡ ወአኩ ፡ ለዛ ፡ ትውልድ ፡ አላ ፡ ለዘይመጽእ ፡ ርሕቅት። 3 በእንተ ፡ ጎሩያን ፡ እቤ ፡ ወአውሣእኩ ፡ በእንቲአሆሙ ፡ ምስለ ፡ ይወፅእ ፡ ቅዱስ ፡ ወዐቢይ ፡ እማኅደሩ። 4 ወአምላዕለ ፡ ዐለም ፡ [ወ]እምሆየ ፡ ይ 10 ከይድ ፡ ጽባ ፡ ሲና ፡ ደብር ፡ ወያስተርኢይ ፡ በትዕይንቱ ፡ ወያስተርኢ ፡ በኅይለ ፡ ንዑ ፡ እምሰማይ። 5 ወይፈርሀ ፡ ኩሎ ፡ ወያድለቀልቁ ፡ ተጋቢያን ፡ ወይነጽሐሙ ፡ ኒዙናት ፡ ወሬዓዩ ፡ ወዐይ ፡ እስከ ፡ አጽናፈ ፡ ምድር። 6 ወይደነግፁ ፡ አድባር ፡ ነዋኅ ፡ ወይትሐተቱ ፡ አ</p>
<p>https://www.me-taappz.com/Tools/Amharic_OCR</p>	<p>መጽሐፈ ሄኖክ።</p> <p>፤</p> <p>GAP. 1. ቃለ ፡ በረከት ፡ ዘሄኖክ ፡ በከመ ፡ ባረከ ፡ ጎሩያነ ፡ ወጻድቃነ ፡ እለ ፡ ሀለዉ ፡ ይኩኑ ፡ በዕለተ ፡ ምንዳቤ ፡ ለአሰሰሎ ፡ ኩሎ ፡ እኩያነ ፡ ወረሲዓን ። 2 ወአውሥአ ፡ እንከ ፡ ሄኖክ ፡ ወይቤ ፡ ብእሲ ፡ ጻድቅ ፡ ዘእምኅበ ፡ እግዚአብሔር ፡ እንዘ ፡ አዕይንቲሁ ፡ ከሡታ ት ፡ ወይሬኢ ፡ ራእየ ፡ ቅዱስ ፡ ዘበሰማያት ፡ ዘአርአዩኒ ፡ መላእክት፤ ወሰማዕኩ ፡ እምኅቤሆሙ ፡ ኩሎ ፡ ወአእመርኩ ፡ አነ ፡ ዘእሬኢ ፡ ወአኩ ፡ ለዛ ፡ ትውልድ ፡ አላ ፡ ለዘይመጽእ ፡ ርሕቅት። 8 በእንተ ፡ ጎሩያን ፡ እቤ ፡ ወአውሣእኩ ፡ በእንቲአሆሙ ፡ ምስለ ፡ ይወፅእ ፡ ቅዱስ ፡ ወዐቢይ ፡ እማኅደሩ፤ 4 ወአምላከ ፡ ዓለም ፡ [ወ]እምሆየ ፡ ይ ከይድ ፡ ዲበ ፡ ሲና ፡ ደብር ፡ ወያስተርኢ ፡ በትዕይንቱ ፡ ወያስተርኢ ፡ በኅይለ ፡ ጽንዑ ፡ እምሰማይ። 5 ወይፈርሀ ፡ ኩሎ ፡ ወያድለቀልቁ ፡ ትጉሃን ፡ ወይነሥአሙ ፡ ፍርሀት ፡ ወረዓድ ፡ ዐቢይ ፡ እስከ ፡ አጽናፈ ፡ ምድር። 6 ወይደነግፁ ፡ አድባር ፡ ነዋኝን ፡ ወይቴሐቱ ፡ አ</p>

4.3 Zu 2.2.1.

Prompt an ChatGPT:

„Dies ist der korrekte Text: et obtulit holocaustomata in ara, et acceptum est Domino, tamquam odor requietionis. {9} Et dixit Deus: Non adiciam iam maledicere terram pro homine, quoniam figura cordis hominis desiit a iuventute sua, et ideo non adiciam simul disperdere omnes viventes sicut feci. Erit autem cum peccaverint habitantes terram, diiudicabo eos in fame sive in gladio, sive in igne, sive in morte, et erit terremotus et dispergentur in inhabitabilia. Terram autem non adiciam corrumpere aqua diluvii. Et in omnibus diebus terre, semen, et messis, frigus et calor, estas et autumnus, die ac nocte, non requiescent, quousque rememorabor qui inhabitant terram, donec compleantur tempora. Bitte übersetze“

Zweiter Versuch, Prompt: „Der Kontext ist antikes Frühjudentum. Bitte versuche es noch einmal“

„Und er brachte Brandopfer dar auf dem Altar, und es wurde dem Herrn wohlgefällig – wie der Duft der Beruhigung. {9} Und Gott sprach: „Nie mehr werde ich die Erde um des Menschen willen verfluchen, denn die Neigung des menschlichen Herzens ist böse von seiner Jugend an. Darum werde ich auch nicht mehr zugleich alle Lebewesen vernichten, wie ich es getan habe. Wenn aber künftig die Bewohner der Erde sündigen, werde ich über sie Gericht bringen – durch Hunger oder Schwert, durch Feuer oder Tod. Dann wird es ein Beben geben, und sie werden zerstreut werden in öde, unbewohnbare Gegenden. Doch die Erde selbst werde ich nicht noch einmal verderben durch das Wasser einer Flut. Solange die Tage der Erde währen, wird Saat und Ernte, Kälte und Hitze, Sommer und Herbst, Tag und Nacht nicht aufhören – bis ich mich wieder erinnere an die, die auf der Erde wohnen, und die Zeiten zur Erfüllung kommen.“

Literaturverzeichnis

- Alammar, Jay/Grootendorst, Maarten 2024: Hands-on large language models. Language understanding and generation. Sebastopol/CA, O'Reilly.
- Campesato, Oswald 2024: Large Language Models. An Introduction. Boston/MA, Mercury.
- Cohn, Leopold (Hg.) 1909: Die Werke Philos von Alexandria in deutscher Übersetzung 4. SJHL 4. Breslau, Marcus.

- Jegham, Nidhal/Abdelatti, Marwen/Elmoubarki, Lassad/Hendawi, Abdeltawab 2025: How Hungry is AI? Benchmarking Energy, Water, and Carbon Footprint of LLM Inference. In: arXiv. DOI: 10.48550/arXiv.2505.09598.
- Kamath, Uday/Keenan, Kevin/Somers, Garrett/Sorenson, Sarah 2024: Large Language Models: A Deep Dive. Bridging Theory and Practice. Cham, Springer. DOI: 10.1007/978-3-031-65647-7.
- Kisch, Guido 1949: Pseudo-Philo's Liber Antiquitatum Biblicarum. Notre Dame/IN, University of Notre Dame.
- McGrath, James F. 2025: ChatGPT and Biblical Studies. In: Interp. 79 (2): 126–135.
- Stökl Ben Ezra, Daniel 2024: Textdigitalisierung. In: Nunn, Christopher A./van Oorschot, Frederike (Hg.): Kompendium Computational Theology 1. Forschungspraktiken in den Digital Humanities. Heidelberg, heiBooks: 107–125. DOI: 10.11588/heibooks.1459.c21906.
- Touati, Charlotte/Chandler de Waal, Matthew 2024: The Destruction and Looting of Heritage in the Tigray War, In: African Arguments vom 25.09.2024. <https://africanarguments.org/2024/09/the-destruction-and-looting-of-heritage-in-the-tigray-war/> (zuletzt abgerufen am 09.04.2025).
- Wassie, Aman K. 2023: Machine Translation for Ge'ez Language. In: arXiv. DOI: 10.48550/arXiv.2311.14530.

Verzeichnis erwähnter Internetauftritte (alle abgerufen am 26.06.2025):

- Amharic OCR: https://www.metaappz.com/Tools/Amharic_OCR
- Anthropic Claude, Modell Sonnet 3.5: <https://www.anthropic.com/news/claude-3-5-sonnet>
- Anthropic, RAG: <https://support.anthropic.com/de/articles/11473015-retrieval-augmented-generation-rag-fur-projekte>
- Beta Masaheft: <https://www.betamasafeft.eu/>
- CJHNTdigital-Projekt: <https://cjhnt-info.saw-leipzig.de/de>
- Convertio: <https://convertio.co/de/ocr/latin/>
- EthioLLM: <https://ethionlp.github.io/index.html>
- I2OCR: <https://www.i2ocr.com/free-online-latin-ocr>
- Perseus: perseus.tufts.org
- Qumran digital: qumran-digital.org