

Werkstattbericht: LLMs als Drop-in-Replacements für historische Dokumente



TRANSFORMA



Überblick

- Projektvorstellung
- Engines + Modelle
- Auswertung
- Einschränkungen
- Fazit



Projektvorstellung

Motivation und Hintergrund



- Ausgangspunkt: Häufige Fragen in der OCR-Sprechstunde zur Nutzung von LLMs
 - Bedarf an einfachen Lösungen zur automatisierten Volltexterstellung
- Projektkontext: Teil des Verbundprojekts TransforMA
 - Kooperation von Universität Mannheim & TH Mannheim
 - UB Mannheim beteiligt im Teilprojekt 3.2: Wissensscouting
- Fokus in diesem Rahmen:
 - Evaluation von LLMs für verschiedene ATR-bezogene Aufgaben

Projektvorstellung

Ansatz der Auswertung



TRANSFORMA

Ausgangsfrage: Können multimodale LLMs als **Drop-In-Replacements** bereits heute „traditionelle“ ATR-Modelle ersetzen?

- Vision-LLMs als neuer Game Changer zeigen großes Potential
 - All-in-One: Layouterkennung, Texterkennung und Strukturierung in einem Prozess
 - Mixed Documents (Handschriften + Druck)
 - Erkennung von komplexen Layouts und Strukturen (Tabellen)
- Layoutbeschreibung mit Koordinaten ist unzureichend
- Halluzinations-Effekte & Transkriptionslücken
- Weitere offene Punkte: Modelgröße, Reproduzierbarkeit, Performance, Kosten etc.

Projektvorstellung

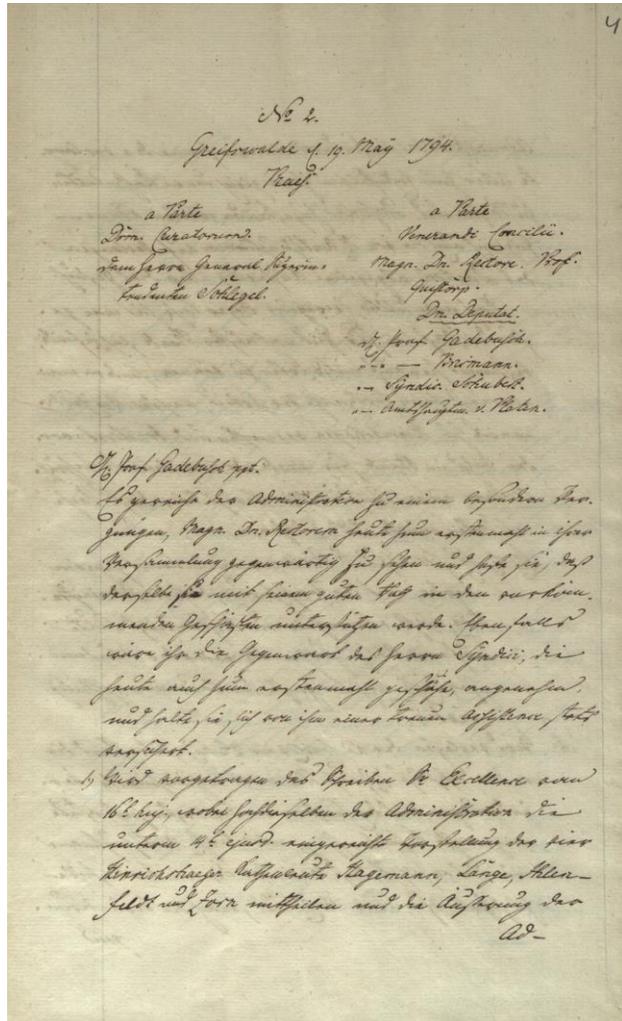
Ground Truth



- **Ground-Truth-Repo** (GitHub) angelegt (bisher privat)
- Repräsentativen Datensatz (mit GT) ausgewählt
 - Handgeschriebene Dokumente ~ 40 Seiten
 - Gedruckte Dokumente ~ 107 Seiten
 - Handgeschriebener und gedruckter Text gemischt ~ 16 Seiten
 - Komplexes Layout
 - Tabellen ~ 12 Seiten
 - Mehrspaltiger Text ~ 25 Seiten
 - Illustrationen, Bilder, Graphiken ~ 28 Seiten

Projektvorstellung

Ground Truth: Beispielseiten



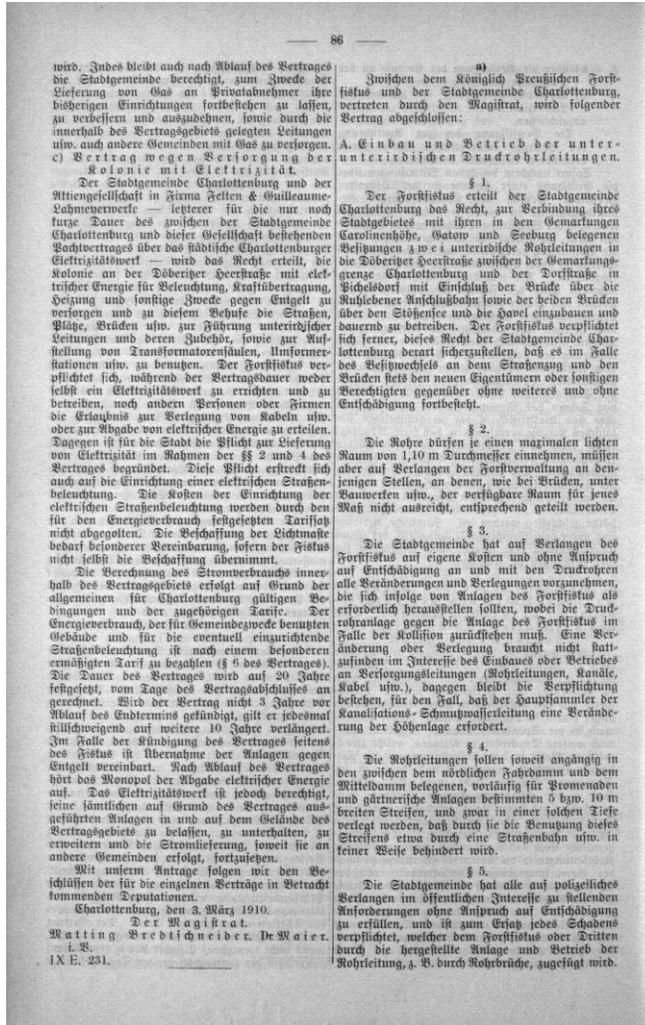
Handwriting

[data/gt/handwriting/read_dataset_german_konzilsprotokolle/00000009](https://zenodo.org/record/215383/files/read_dataset_german_konzilsprotokolle/00000009)

<p>Grüning, T., Leifert, G., Michael, J., Strauß, T., Weidemann, M., Labahn, R. (2016). "read_dataset_german_konzilsprotokolle" [Data set]. Zenodo. http://doi.org/10.5281/zenodo.215383</p>	<p>read_dataset_german_konzilsprotokolle</p>	<p>late 1700s</p>	<p>Source</p>	<p>5</p>	<p>CC BY 4.0</p>
---	---	-------------------	-------------------------------	----------	------------------

Projektvorstellung

Ground Truth: Beispielseiten



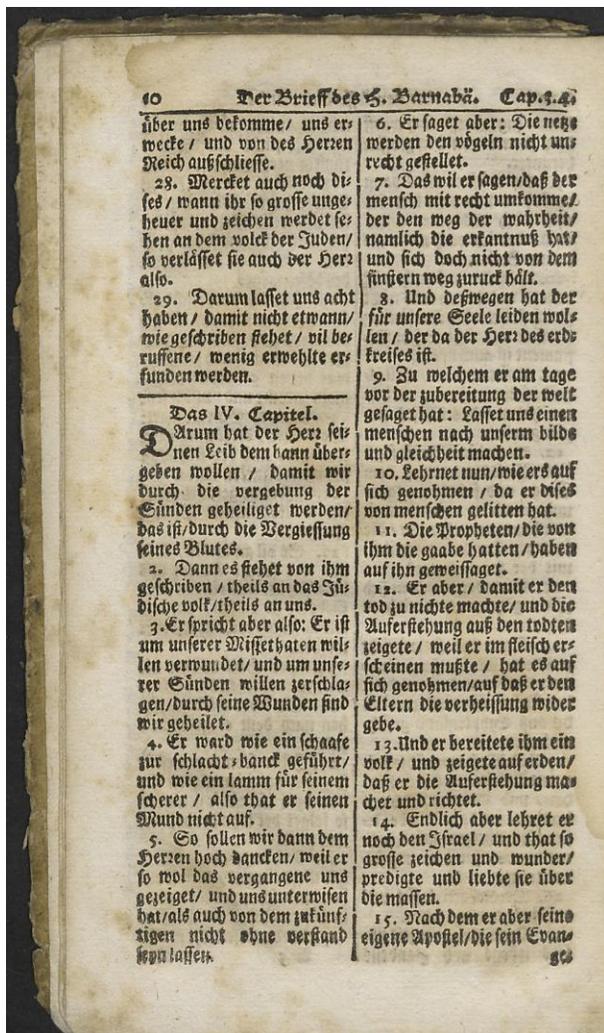
Print

data/gt/layout/multiple_columns/charlottenburger-gt

UB Mannheim: th-schmidt, JKamlah, & Stefan Weil. (2024). UB-Mannheim/charlottenburger-amtsschrifttum: v1.1.0 (v1.1.0). Zenodo. https://doi.org/10.5281/zenodo.10686500	1905-1910	Source	5	CC BY
--	-----------	------------------------	---	-------

Projektvorstellung

Ground Truth: Beispielseiten



Print

[data/gt/print/apocrypha1714](https://data.gt/print/apocrypha1714)

<p>UB Heidelberg: I. Gessner, D. (1714). Novi Testamenti Apocrypha: Oder: Etlicher Lehr-Jünger des Herren: Und Apostolischen Männer Send-Brieffe, Die vor Alters bey der ersten Kirchen in vilen Gemeinden mit Nutzen gelesen, Und Denen Canonischen Schrifftten mit angehänget worden; Nun zum gemeinen Nutz genauer nach dem Grunde übersehen, und in Capitel und Verse abgetheilt. Zürich</p>	<p>apocrypha1714</p>	<p>1714</p>	<p>Images, PAGE-XML</p>	<p>5</p>	<p>Public Domain</p>
---	--------------------------------------	-------------	-------------------------	----------	----------------------

Engines + Modelle

Kraken <https://kraken.re/>



- **Handschriften:**

- „German_Handwriting“: <https://doi.org/10.5281/zenodo.7933462>

- **Druckschriften:**

- „German_Print“: <https://doi.org/10.5281/zenodo.10519595>

→ State-of-the-Art-Modelle, die an der UB Mannheim trainiert wurden

Engines + Modelle

Multimodale LLMs



Provider / API	Modelname	Parameter	Link
UB lokal (Ollama)	MiniCPM-V-2_6	8B	https://ollama.com/library/minicpm-v:latest
UB lokal (Ollama)	Llava 1.6	7B	https://ollama.com/library/llava:7b
UB lokal (Ollama)	Qwen 2.5 VL	7B	https://ollama.com/library/qwen2.5vl:7b
GWDG / KISSKI	InternVL 2.5	8B	https://huggingface.co/OpenGVLab/InternVL2_5-8B-MPO
GWDG / KISSKI	Gemma 3 instruct	27B	https://huggingface.co/google/gemma-3-27b-it
GWDG / KISSKI	Qwen 2.5 VL instruct	72B	https://huggingface.co/Qwen/Qwen2.5-VL-72B-Instruct

GWDG Modell-Übersicht: <https://docs.hpc.gwdg.de/services/chat-ai/models/index.html>

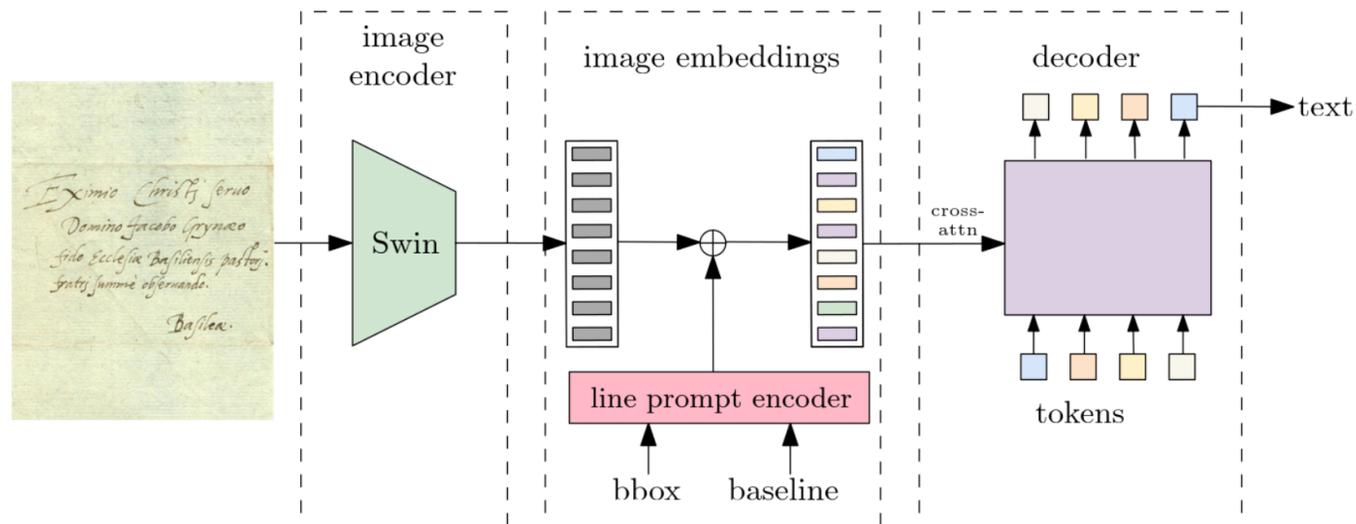
Engines + Modelle

Party <https://github.com/mittagessen/party>

party

party is **PA**ge-wise **R**ecognition of **T**ext-**y**. It is a replacement for conventional text recognizers in ATR system using conventional baseline+bounding polygon (where it eliminates the need for bounding polygons) and bounding box line data models.

Party consists of a Swin vision transformer encoder, baseline positional embeddings, and a [tiny Llama decoder](#) trained on octet tokenization.



Auswertung

Ansatz & Workflow

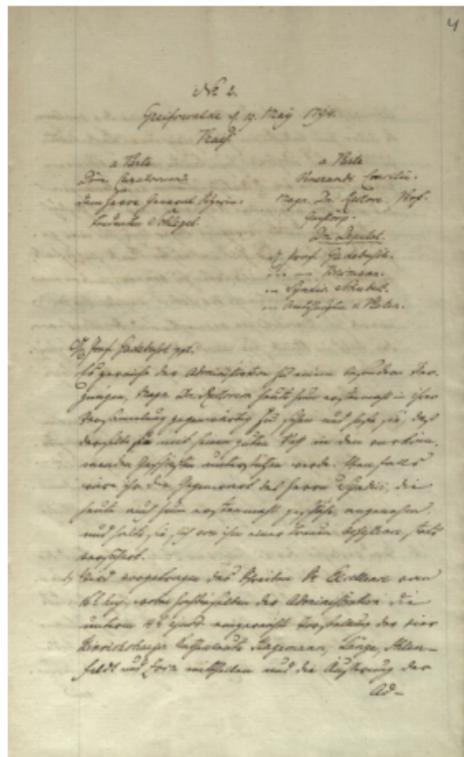


TRANSFORMA

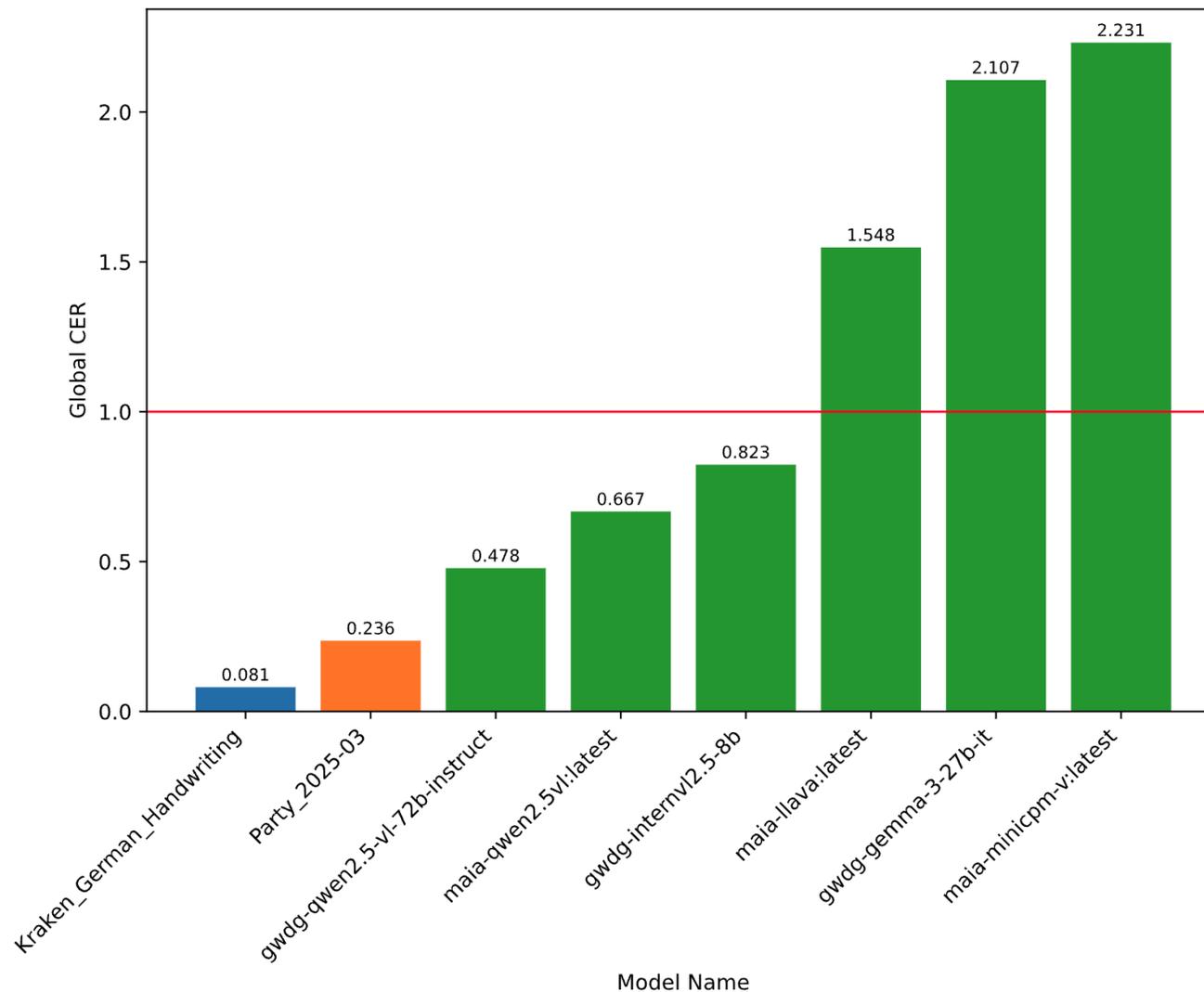
- **Ansatz:** Evaluation der "rohen" OCR-Performance von LLMs bei historischen Dokumenten **auf Zeilenebene**
- **Workflow:**
 - PagePlus als Tool: <https://github.com/JKamlah/PagePlus/tree/feature/llm>
 - GT Small Datensatz:
 - Handwriting: 3 Seiten
 - Print: 8 Seiten
- Auswertung auf Basis von **CER (Character Error Rate)**
- **Viele Einschränkungen!** (dazu später mehr)



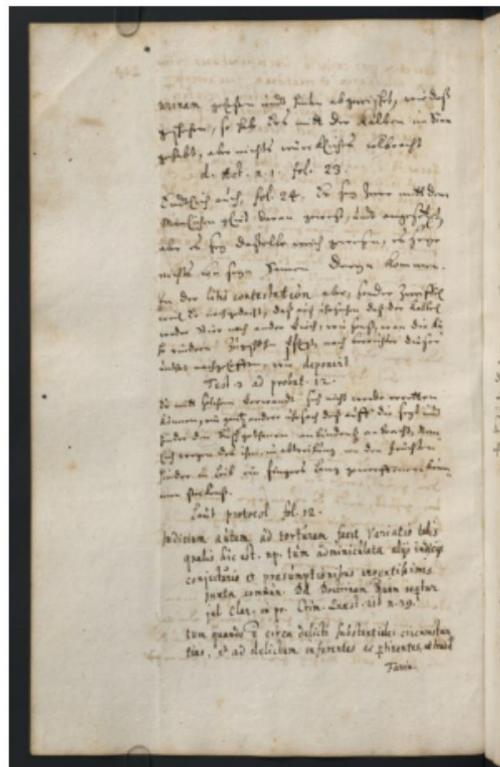
Auswertung: Handwriting



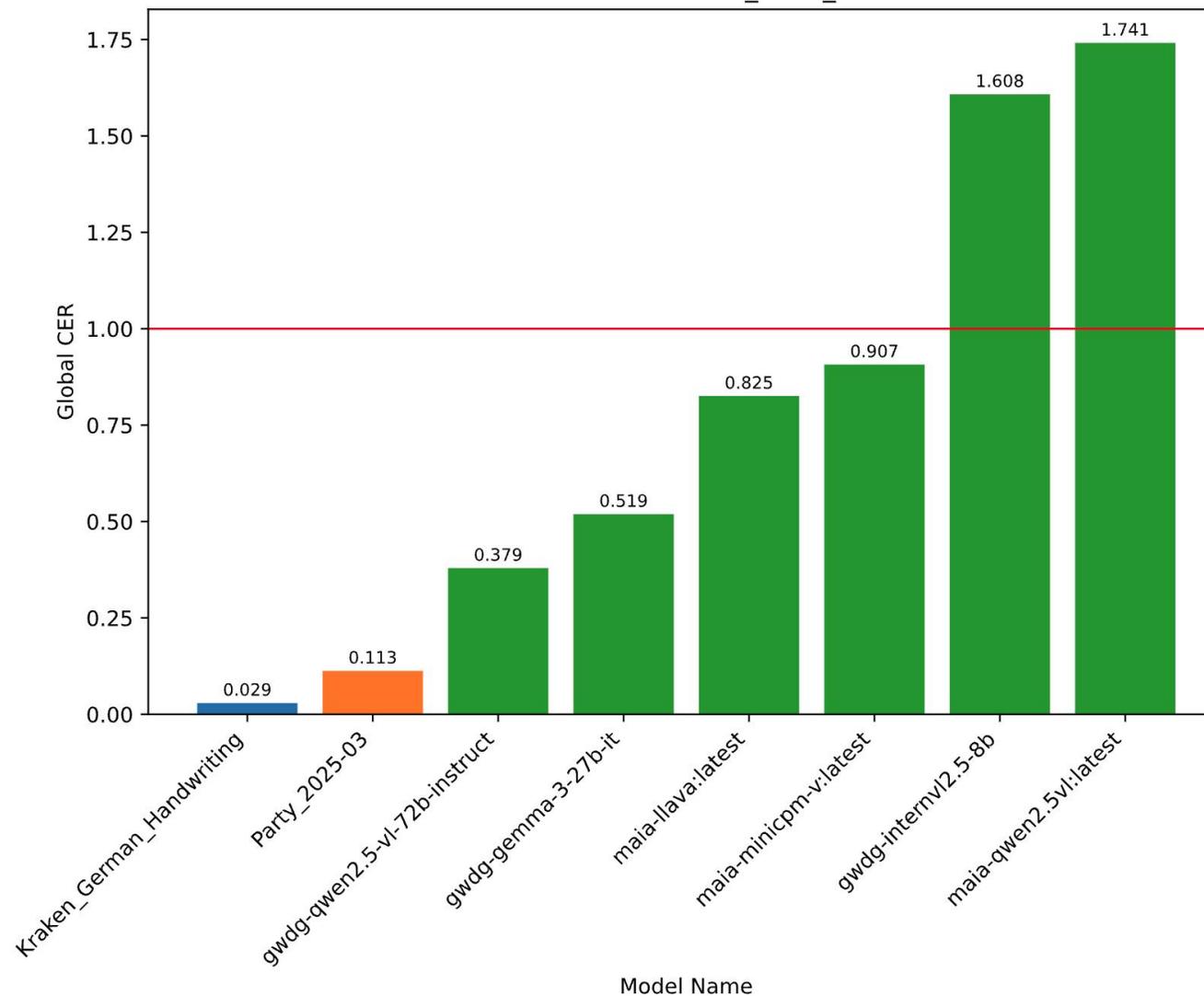
Character Error Rate: 00000009.xml



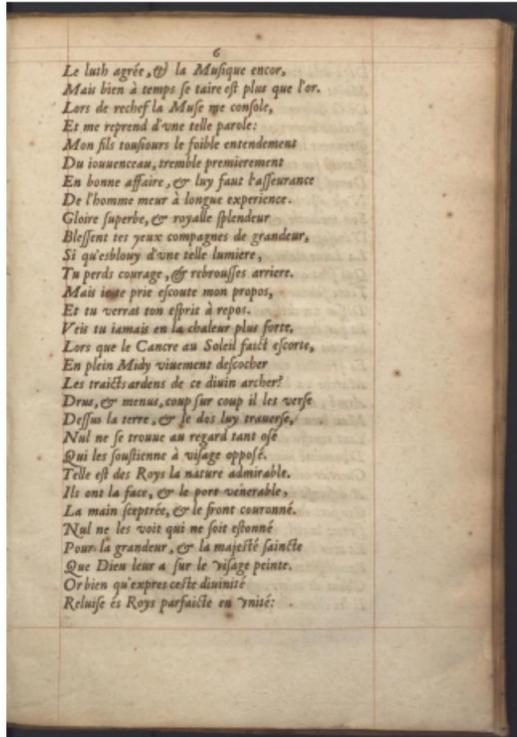
Auswertung: Handwriting



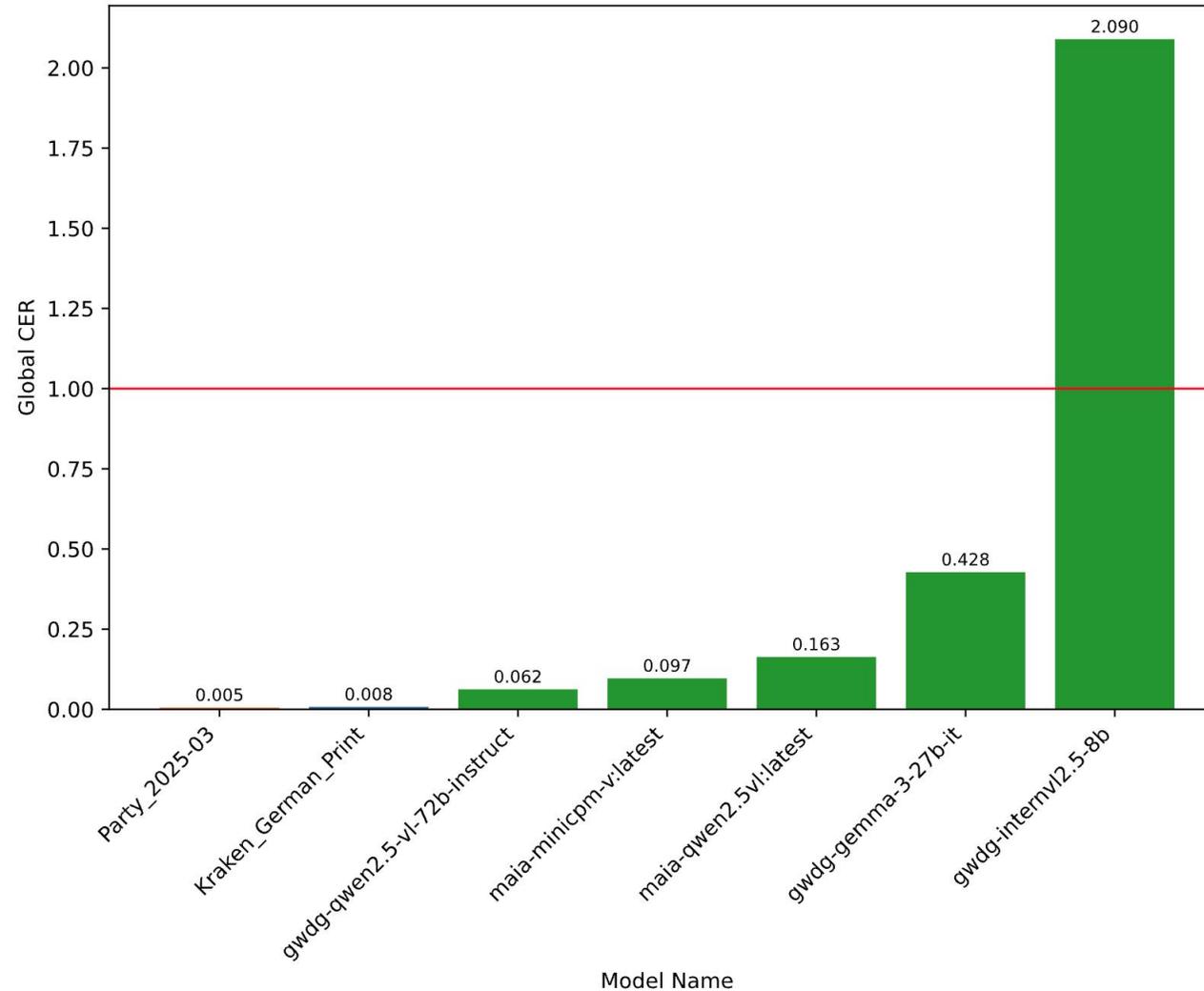
Character Error Rate: 1012617_0070_40895556.xml



Auswertung: Print



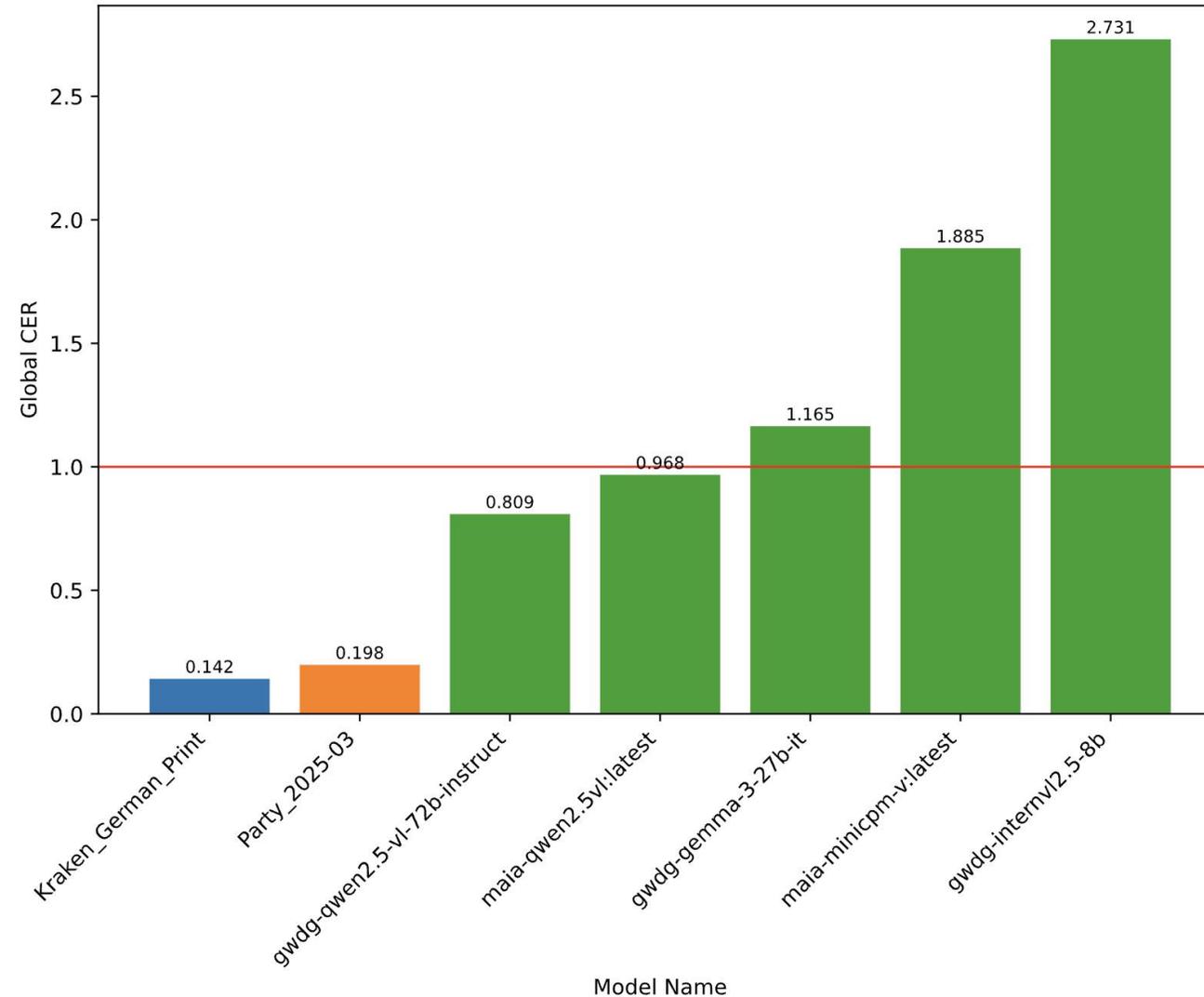
Character Error Rate: 1837705283_0012.xml



Auswertung: Print



Character Error Rate: 432977651_0009.xml



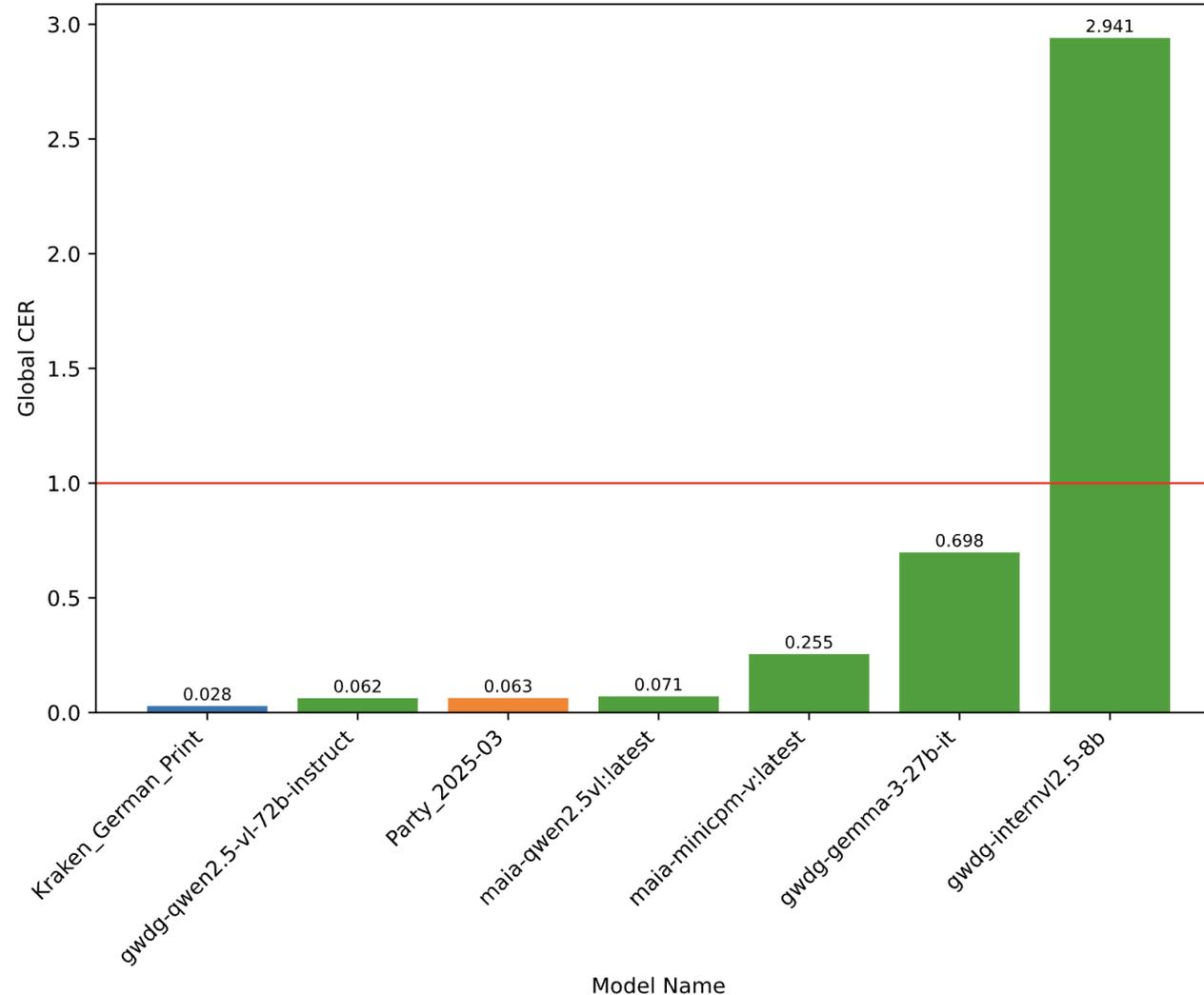
A

Auswertung: Print

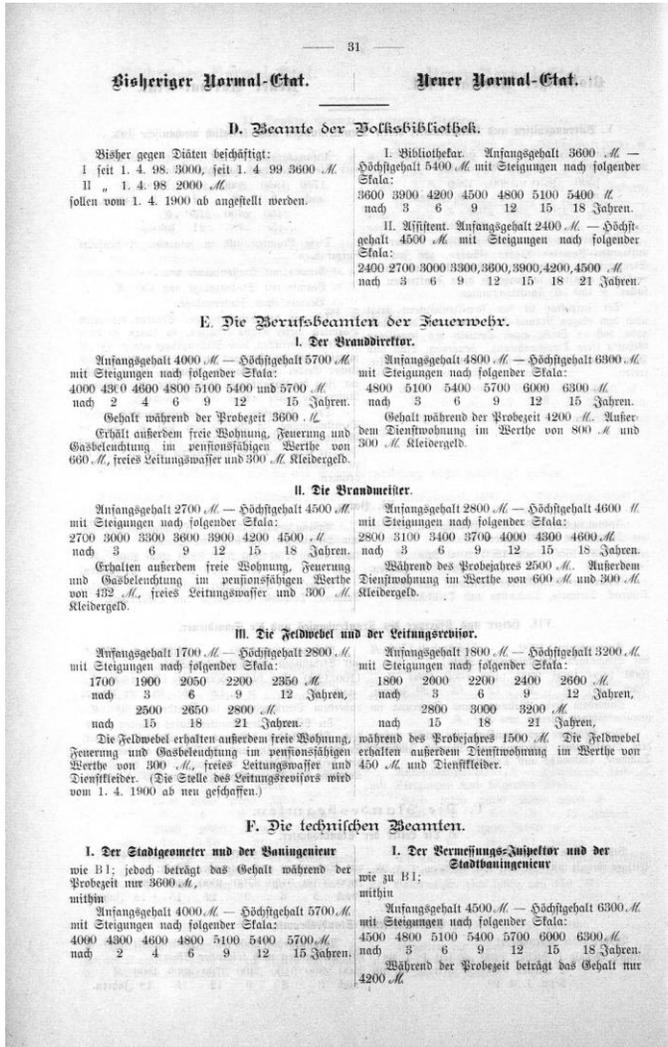
31

Bisheriger Normal-Stat.	Neuer Normal-Stat.
D. Beamte der Volksbibliothek.	
Bisher gegen Däten beidhäftig:	
I seit 1. 4. 98. 3000, seit 1. 4. 99. 3600 <i>M.</i>	I. Bibliothekar. Anfangsgehalt 3600 <i>M.</i> — Höchstgehalt 5400 <i>M.</i> mit Steigungen nach folgender Stala:
II „ 1. 4. 98. 2000 <i>M.</i>	3600 3900 4200 4500 4800 5100 5400 <i>M.</i>
sollen vom 1. 4. 1900 ab ange stellt werden.	nach 3 6 9 12 15 18 Jahren.
	II. Mittelt. Anfangsgehalt 2400 <i>M.</i> — Höchstgehalt 4500 <i>M.</i> mit Steigungen nach folgender Stala:
	2400 2700 3000 3300, 3600, 3900, 4200, 4500 <i>M.</i>
	nach 3 6 9 12 15 18 21 Jahren.
E. Die Berufsbeamten der Feuerwehr.	
I. Der Branddirektor.	
Anfangsgehalt 4000 <i>M.</i> — Höchstgehalt 5700 <i>M.</i> mit Steigungen nach folgender Stala:	Anfangsgehalt 4800 <i>M.</i> — Höchstgehalt 6300 <i>M.</i> mit Steigungen nach folgender Stala:
4000 4300 4600 4800 5100 5400 und 5700 <i>M.</i>	4800 5100 5400 5700 6000 6300 <i>M.</i>
nach 2 4 6 9 12 15 Jahren.	nach 3 6 9 12 15 Jahren.
Gehalt während der Probezeit 3600 <i>M.</i>	Gehalt während der Probezeit 4200 <i>M.</i> Außer dem Dienstwohnung im Werte von 800 <i>M.</i> und 300 <i>M.</i> Kleidergeld.
Gehalt außerdem freie Wohnung, Feuerung und Gasbeleuchtung im pensionsfähigen Werte von 600 <i>M.</i> , freies Leitungswasser und 300 <i>M.</i> Kleidergeld.	
II. Die Brandmeister.	
Anfangsgehalt 2700 <i>M.</i> — Höchstgehalt 4500 <i>M.</i> mit Steigungen nach folgender Stala:	Anfangsgehalt 2800 <i>M.</i> — Höchstgehalt 4600 <i>M.</i> mit Steigungen nach folgender Stala:
2700 3000 3300 3600 3900 4200 4500 <i>M.</i>	2800 3100 3400 3700 4000 4300 4600 <i>M.</i>
nach 3 6 9 12 15 18 Jahren.	nach 3 6 9 12 15 18 Jahren.
Gehalten außerdem freie Wohnung, Feuerung und Gasbeleuchtung im pensionsfähigen Werte von 432 <i>M.</i> , freies Leitungswasser und 300 <i>M.</i> Kleidergeld.	Während des Probejahres 2500 <i>M.</i> Außer dem Dienstwohnung im Werte von 600 <i>M.</i> und 300 <i>M.</i> Kleidergeld.
III. Die Feldwebel und der Leitungsrevier.	
Anfangsgehalt 1700 <i>M.</i> — Höchstgehalt 2800 <i>M.</i> mit Steigungen nach folgender Stala:	Anfangsgehalt 1800 <i>M.</i> — Höchstgehalt 3200 <i>M.</i> mit Steigungen nach folgender Stala:
1700 1900 2050 2200 2350 <i>M.</i>	1800 2000 2200 2400 2600 <i>M.</i>
nach 3 6 9 12 Jahren,	nach 3 6 9 12 Jahren,
2500 2650 2800 <i>M.</i>	2800 3000 3200 <i>M.</i>
nach 15 18 21 Jahren.	nach 15 18 21 Jahren,
Die Feldwebel erhalten außerdem freie Wohnung, Feuerung und Gasbeleuchtung im pensionsfähigen Werte von 300 <i>M.</i> , freies Leitungswasser und Dienstkleider. (Die Stelle des Leitungsrevier wird vom 1. 4. 1900 ab neu geschaffen.)	Während des Probejahres 1500 <i>M.</i> Die Feldwebel erhalten außerdem Dienstwohnung im Werte von 450 <i>M.</i> und Dienstkleider.
F. Die technischen Beamten.	
I. Der Stadtgeometer und der Bauingenieur	
wie B1; jedoch beträgt das Gehalt während der Probezeit nur 3600 <i>M.</i> , mit hin	wie zu B1; mit hin
Anfangsgehalt 4000 <i>M.</i> — Höchstgehalt 5700 <i>M.</i> mit Steigungen nach folgender Stala:	Anfangsgehalt 4500 <i>M.</i> — Höchstgehalt 6300 <i>M.</i> mit Steigungen nach folgender Stala:
4000 4300 4600 4800 5100 5400 5700 <i>M.</i>	4500 4800 5100 5400 5700 6000 6300 <i>M.</i>
nach 2 4 6 9 12 15 Jahren,	nach 3 6 9 12 15 18 Jahren.
	Während der Probezeit beträgt das Gehalt nur 4200 <i>M.</i>

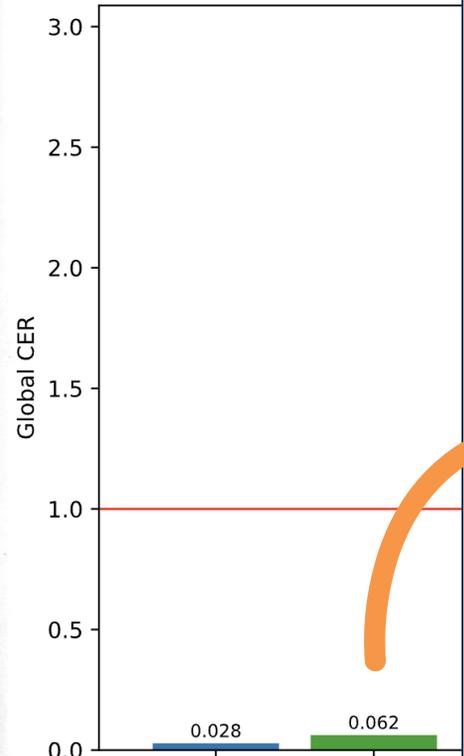
Character Error Rate: Tag_Mitt_Charl_1900_006900.xml



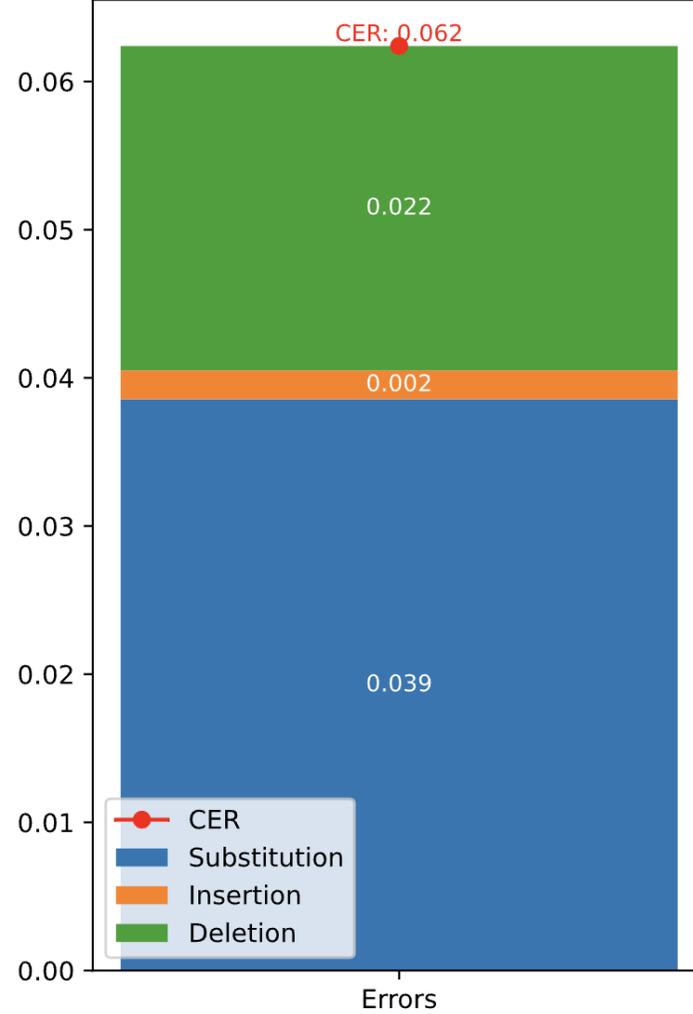
Auswertung: Print



Character



File: Tag_Mitt_Charl_1900_006900.xml / Model: gwdg-qwen2.5-vl-72b-instruct



Kraken_German_Print
gwdg-qwen2.5-vl-72b-instruct
Party_2

maia-0
maia-1
gwdg-0
gwdg-1

Model Name

Auswertung: Print

86

wird. Anders bleibt auch nach Ablauf des Vertrages die Stadtgemeinde berechtigt, zum Zwecke der Lieferung von Gas an Privatabnehmer ihre bisherigen Einrichtungen fortzubehalten zu lassen, zu verwalten und auszubehalten, sowie durch die innerhalb des Vertragsgebietes gelegten Leitungen usw. auch andere Gemeinden mit Gas zu versorgen.

c) Vertrag wegen Versorgung der Kolonie mit Elektrizität.

Der Stadtgemeinde Charlottenburg und der Aktiengesellschaft in Firma Hellen & Guillaume-Lahmeyerwerke — letzterer für die nur noch kurze Dauer des zwischen der Stadtgemeinde Charlottenburg und dieser Gesellschaft bestehenden Nachtrages über das kaiserliche Charlottenburger Elektrizitätswerk — wird das Recht erteilt, die Kolonie an der Döberitzer Heerstraße mit elektrischer Energie für Beleuchtung, Kraftübertragung, Heizung und sonstige Zwecke gegen Entgelt zu versorgen und zu diesem Zwecke die Straßen, Plätze, Brücken usw. zur Führung unterirdischer Leitungen und deren Zubehör, sowie zur Aufstellung von Transformatorstationen, Umformstationen usw. zu benutzen. Der Fortifikationsverpflichtet sich, während der Vertragsdauer weder selbst ein Elektrizitätswerk zu errichten und zu betreiben, noch anderen Personen oder Firmen die Erlaubnis zur Verlegung von Kabeln usw. oder zur Abgabe von elektrischer Energie zu erteilen. Dagegen ist für die Stadt die Pflicht zur Lieferung von Elektrizität im Rahmen der §§ 2 und 4 des Vertrages begründet. Diese Pflicht erstreckt sich auch auf die Einrichtung einer elektrischen Straßenbeleuchtung. Die Kosten der Einrichtung der elektrischen Straßenbeleuchtung werden durch den für den Energieverbrauch festgesetzten Tarif nicht abgegolten. Die Beschaffung der Lichtmaterie bedarf besonderer Vereinbarung, sofern der Fortifikationsverpflichtet sich die Beschaffung übernimmt.

Die Berechnung des Stromverbrauches innerhalb des Vertragsgebietes erfolgt auf Grund der allgemeinen für Charlottenburg gültigen Bedingungen und der angehörigen Tarife. Der Energieverbrauch, der für Gemeinbezwecke benutzten Gebäude und für die eventuell einzurichtende Straßenbeleuchtung ist nach einem besonderen ermäßigten Tarif zu bezahlen (§ 6 des Vertrages). Die Dauer des Vertrages wird auf 20 Jahre festgesetzt, vom Tage des Vertragsabschlusses an gerechnet. Wird der Vertrag nicht 3 Jahre vor Ablauf des Endtermins gekündigt, gilt er jedesmal stillschweigend auf weitere 10 Jahre verlängert. Im Falle der Kündigung des Vertrages seitens des Fortifikationsverpflichteten ist Übernahme der Anlagen gegen Entgelt vereinbart. Nach Ablauf des Vertrages hat das Monopol der Abgabe elektrischer Energie auf. Das Elektrizitätswerk ist jedoch berechtigt, seine sämtlichen auf Grund des Vertrages angelegten Anlagen in und auf dem Gelände des Vertragsgebietes zu belassen, zu unterhalten, zu erweitern und die Stromlieferung, soweit sie an andere Gemeinden erfolgt, fortzusetzen.

Mit unserm Antrage folgen wir den Beschlüssen der für die einzelnen Verträge in Betracht kommenden Deputationen.

Charlottenburg, den 3. März 1910.

Der Magistrat.

Ersting

Ersting

IX E. 231.

a) Zwischen dem königlich Preussischen Fortifikations- und der Stadtgemeinde Charlottenburg, vertreten durch den Magistrat, wird folgender Vertrag abgeschlossen:

A. Einbau und Betrieb der unterirdischen Druckrohrleitungen.

§ 1.

Der Fortifikationsverpflichtet der Stadtgemeinde Charlottenburg das Recht, zur Verbindung ihres Stadtgebietes mit ihren in den Gemarkungen Carolinenshöhe, Glaton und Seeburg gelegenen Wasserleitungen 3 oder 4 unterirdische Rohrleitungen in die Döberitzer Heerstraße zwischen der Gemarkungsgrenze Charlottenburg und der Dorfstraße in Fiedelsdorf mit Einschluß der Brücke über die Bahnhofsbrücke und die Havel einzubauen und dauernd zu betreiben. Der Fortifikationsverpflichtet sich ferner, dieses Recht der Stadtgemeinde Charlottenburg bereit zu stellen, daß es im Falle des Bedürfnisses an dem Straßenzug und den Brücken stets den neuen Eigentümern oder sonstigen Berechtigten gegenüber ohne weiteres und ohne Entschädigung fortbehalten.

§ 2.

Die Rohre dürfen je einen maximalen lichten Raum von 1,10 m Durchmesser einnehmen, müssen aber auf Verlangen der Fortifikationsverwaltung an denjenigen Stellen, an denen, wie bei Brücken, unter Benutzen usw., der verfügbare Raum für jenes Maß nicht ausreicht, entsprechend geteilt werden.

§ 3.

Die Stadtgemeinde hat auf Verlangen des Fortifikationsverpflichteten auf eigene Kosten und ohne Anspruch auf Entschädigung an und mit den Druckrohren alle Veränderungen und Verlegungen vorzunehmen, die sich infolge von Anlagen des Fortifikationsverpflichteten als erforderlich herausstellen sollten, wobei die Druckrohranlage gegen die Anlage des Fortifikationsverpflichteten geschützt sein muß. Eine Veränderung oder Verlegung braucht nicht haltlos zu sein im Interesse des Einbaues oder Betriebes an Versorgungsleitungen (Nohrleitungen, Kanäle, Kabel usw.), dagegen bleibt die Verpflichtung bestehen, für den Fall, daß der Hauptkanal der Kanalisations-Schmutzwasserleitung eine Veränderung der Höhenlage erfordert.

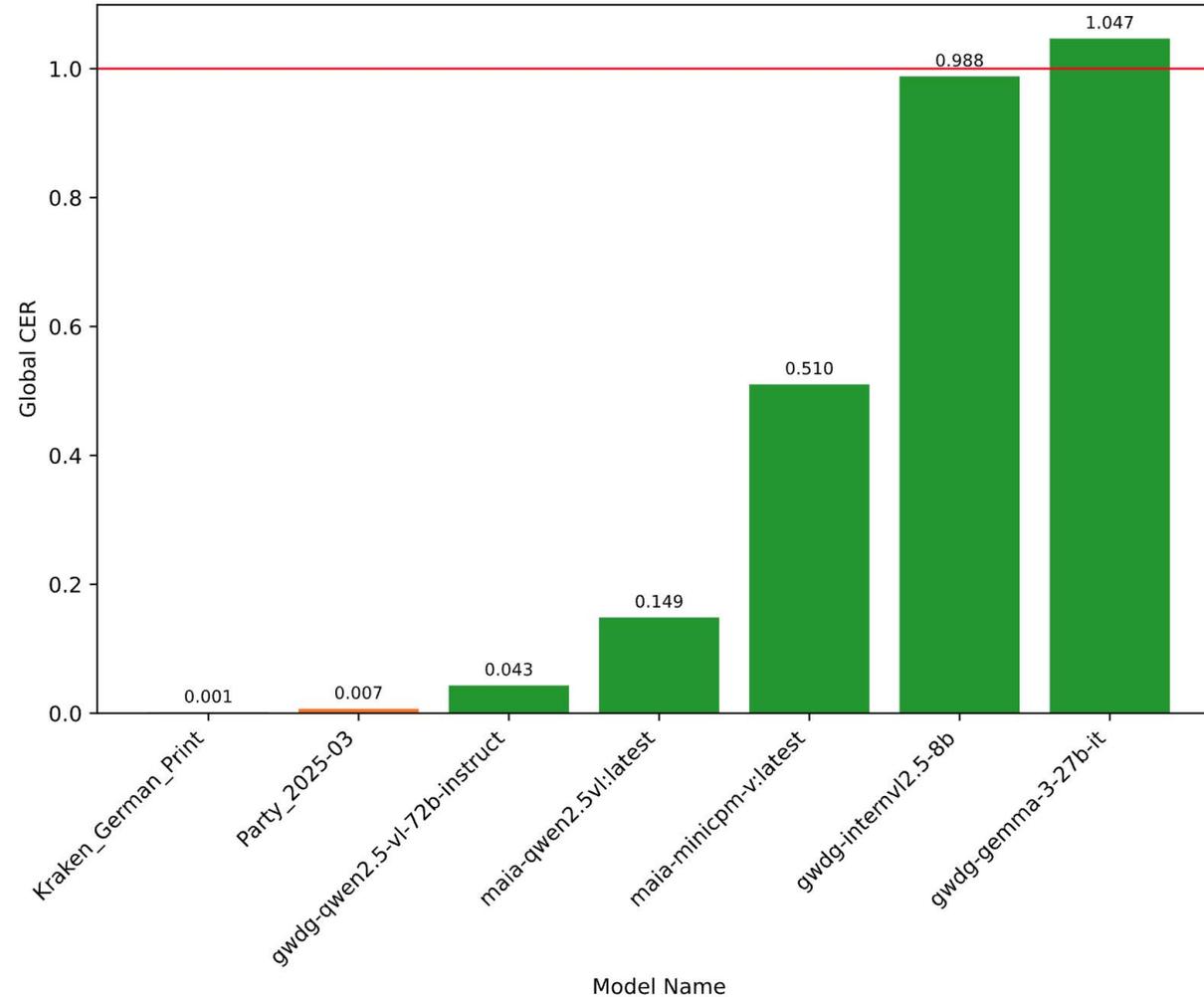
§ 4.

Die Rohrleitungen sollen soweit angängig in den zwischen dem nördlichen Fortkamm und dem Mittelkamm belegenen, verläufig für Brombeeren und gärtnerische Anlagen bestimmten 5 bzw. 10 m breiten Streifen, und zwar in einer solchen Tiefe verlegt werden, daß durch sie die Benutzung dieses Streifens etwa durch eine Straßenbahn usw. in keiner Weise behindert wird.

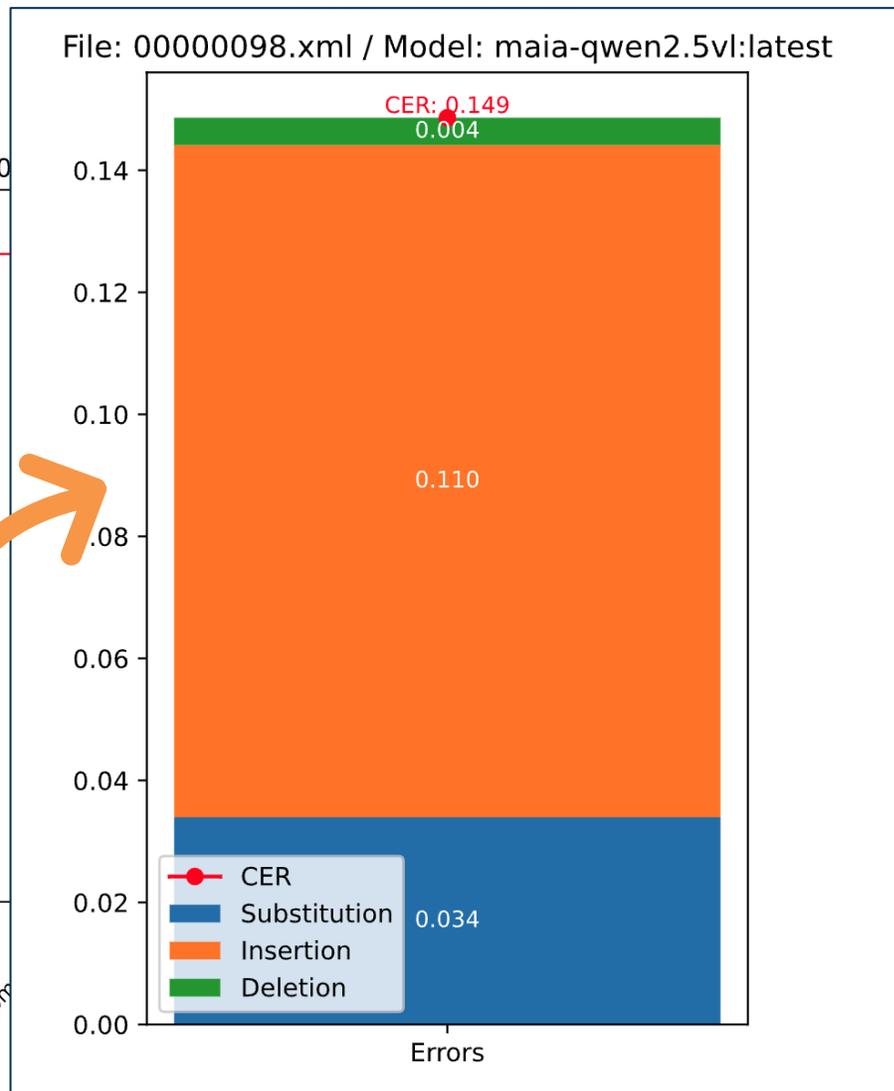
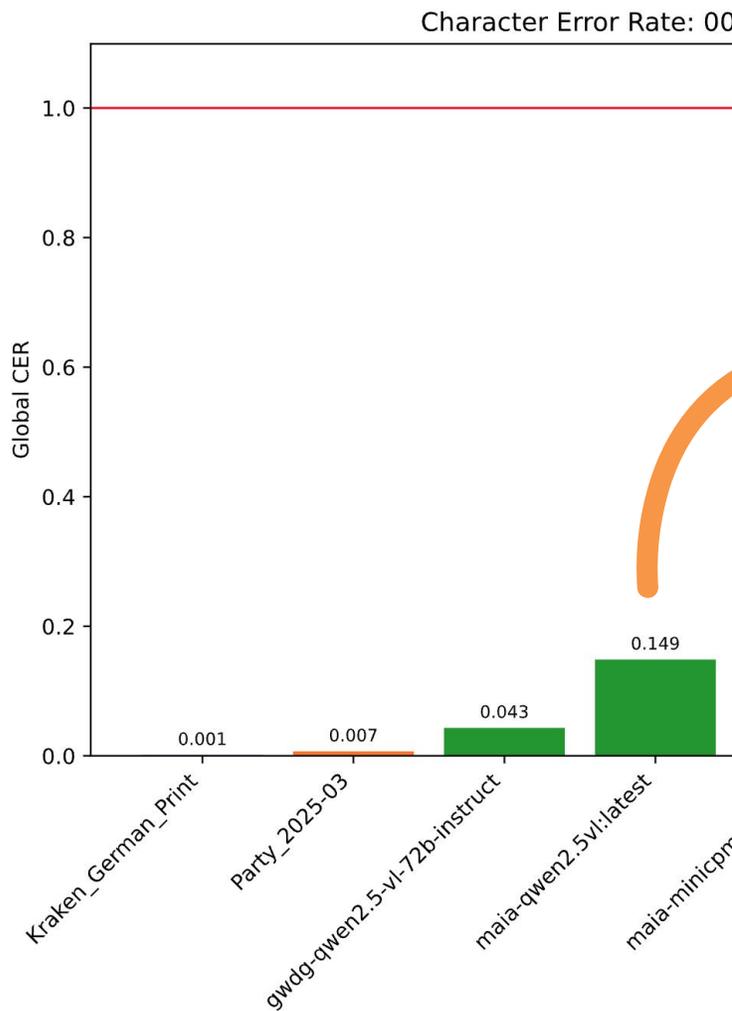
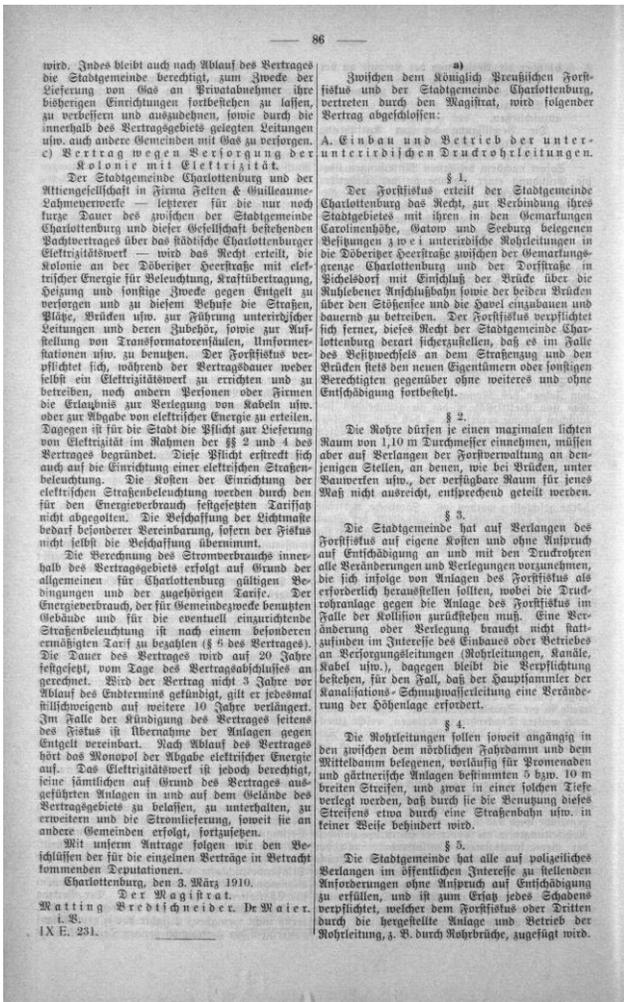
§ 5.

Die Stadtgemeinde hat alle auf polizeiliches Verlangen im öffentlichen Interesse zu stellenden Anforderungen ohne Anspruch auf Entschädigung zu erfüllen, und ist zum Ersatz jedes Schadens verpflichtet, welcher dem Fortifikationsverpflichteten durch die hergestellte Anlage und Betrieb der Rohrleitung, z. B. durch Rohrbrüche, zugefügt wird.

Character Error Rate: 00000098.xml

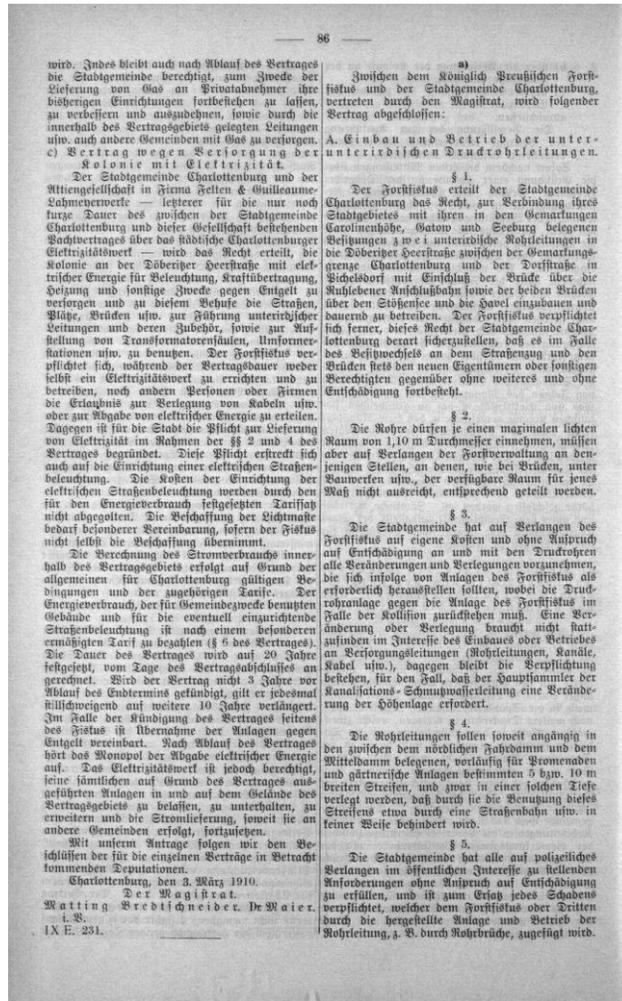


Auswertung: Print



Model Name

Auswertung: Print



Qwen 2.5 VL 7B | Insertions

TextLine r3l26

GT § 2.

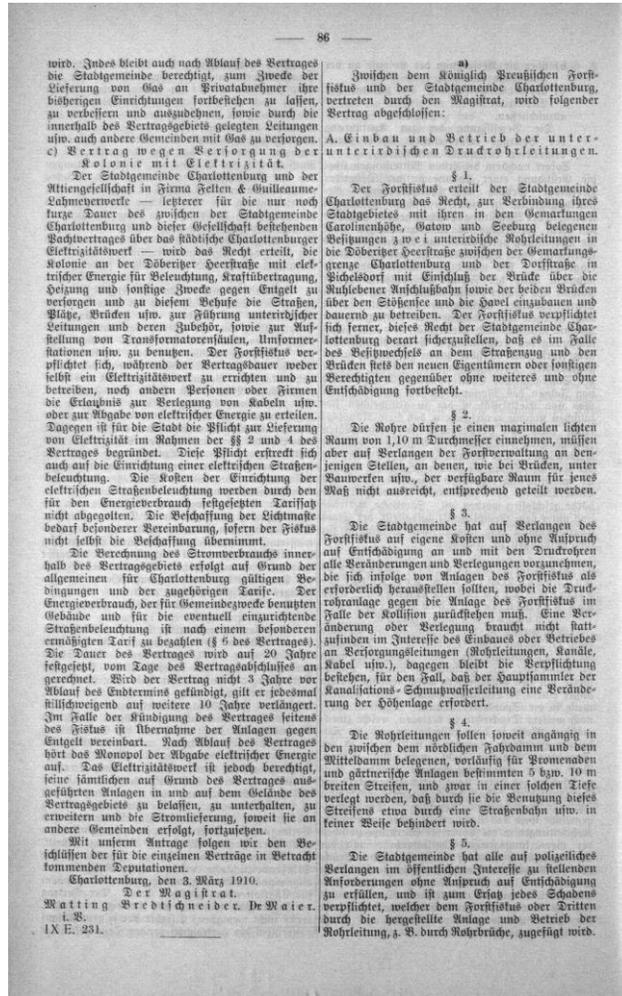
LLM Image Do not provide alternative variation! Output format JSON (a single line with the information)(only output one version via line, dont repeat!

TextLine r3l58

GT § 5.

LLM Image Do not provide alternative variation! Output format JSON (a single line with the information)(only output one version via line, dont repeat!

Auswertung: Print



Qwen 2.5 VL 7B | Normalisationen

TextLine r3l26

- GT** Verlangen im öffentlichen Interesse zu stellenden
- LLM** Verlangen im öffentlichen Interesse zu stellenden

TextLine r3l47

- GT** Kanalisations=Schmutzwasserleitung eine Verände=
- LLM** Kanalisations=Schmutzwasserleitungen eine Veränderung

Auswertung

86

wird. Indes bleibt auch nach Ablauf des Vertrages die Stadtgemeinde berechtigt, zum Zwecke der Lieferung von Gas an Privatabnehmer ihre bisherigen Einrichtungen fortzubehalten zu lassen, zu verwalten und auszubehalten, sowie durch die innerhalb des Vertragsgebietes gelegten Leitungen usw. auch andere Gemeinden mit Gas zu versorgen.

c) Vertrag wegen Versorgung der Kantonen mit Elektrizität.

Der Stadtgemeinde Charlottenburg und der Aktiengesellschaft in Firma Hellen & Guillaume-Labmederwerke — letzterer für die nur noch kurze Dauer des zwischen der Stadtgemeinde Charlottenburg und dieser Gesellschaft bestehenden Nachtrages über das künftige Charlottenburger Elektrizitätswerk — wird das Recht erteilt, die Kolonie an der Töbriker Heerstraße mit elektrischer Energie für Beleuchtung, Kraftübertragung, Heizung und sonstige Zwecke gegen Entgelt zu versorgen und zu diesem Zwecke die Straßen, Plätze, Brücken usw. zur Führung unterirdischer Leitungen und deren Zubehör, sowie zur Aufstellung von Transformatorzentralen, Umformstationen usw. zu benutzen. Der Fortistus verpflichtet sich, während der Vertragsdauer weder selbst ein Elektrizitätswerk zu errichten und zu betreiben, noch anderen Personen oder Firmen die Erlaubnis zur Verlegung von Kabeln usw. oder zur Abgabe von elektrischer Energie zu erteilen. Dagegen ist für die Stadt die Pflicht zur Lieferung von Elektrizität im Rahmen der §§ 2 und 4 des Vertrages begründet. Diese Pflicht erstreckt sich auch auf die Einrichtung einer elektrischen Straßenbeleuchtung. Die Kosten der Einrichtung der elektrischen Straßenbeleuchtung werden durch den für den Energieverbrauch festgesetzten Tarifsatz nicht abgegolten. Die Beschaffung der Lichtmaterie bedarf besonderer Vereinbarung, sofern der Fortistus nicht selbst die Beschaffung übernimmt.

Die Berechnung des Stromverbrauches innerhalb des Vertragsgebietes erfolgt auf Grund der allgemeinen für Charlottenburg gültigen Bedingungen und der angehörigen Tarife. Der Energieverbrauch, der für Gemeinbezwecke benutzten Gebäude und für die eventuell einzurichtende Straßenbeleuchtung ist nach einem besonderen ermäßigten Tarif zu bezahlen (§ 6 des Vertrages). Die Dauer des Vertrages wird auf 20 Jahre festgesetzt, vom Tage des Vertragsabschlusses an gerechnet. Wird der Vertrag nicht 3 Jahre vor Ablauf des Endtermins gekündigt, gilt er jedesmal stillschweigend auf weitere 10 Jahre verlängert. Im Falle der Kündigung des Vertrages seitens des Fortistus ist Übernahme der Anlagen gegen Entgelt vereinbart. Nach Ablauf des Vertrages hört das Monopol der Abgabe elektrischer Energie auf. Das Elektrizitätswerk ist jedoch berechtigt, seine sämtlichen auf Grund des Vertrages angelegten Anlagen in und auf dem Gelände des Vertragsgebietes zu belassen, zu unterhalten, zu erweitern und die Stromlieferung, soweit sie an andere Gemeinden erfolgt, fortzusetzen.

Mit unserm Antrage folgen wir den Beschlüssen der für die einzelnen Verträge in Betracht kommenden Deputationen.

Charlottenburg, den 3. März 1910.

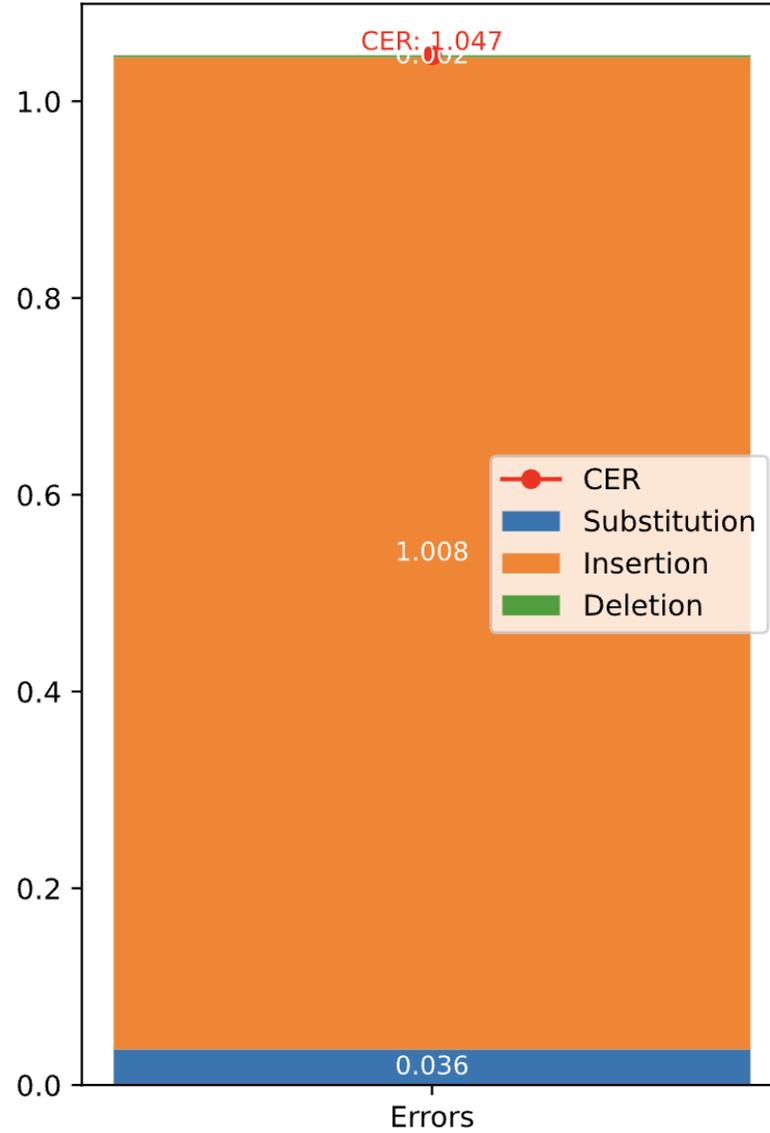
Der Magistrat.

Witting Arendtschneider, Dr. Waier.

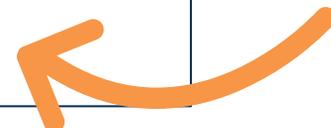
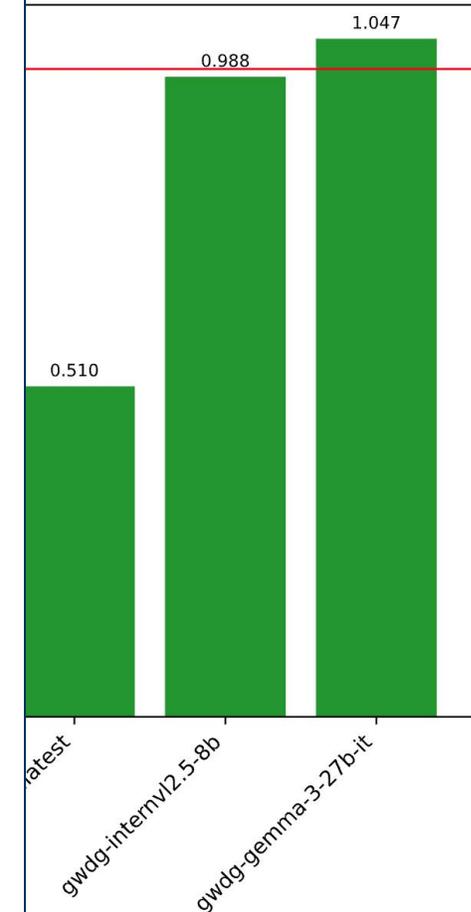
i. V.

IX E. 231.

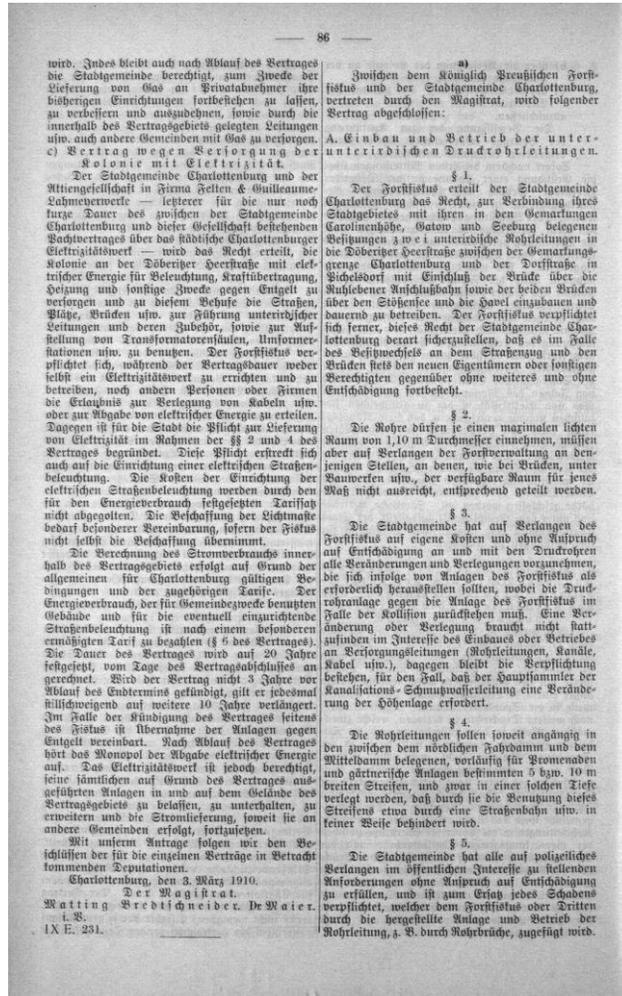
File: 00000098.xml / Model: gwdg-gemma-3-27b-it



00098.xml



Auswertung: Print



Gemma 3 27B | Insertions

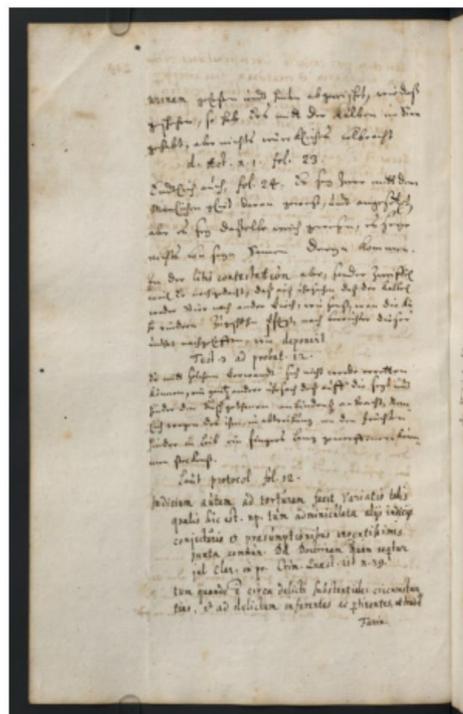
TextLine r212

GT die Stadtgemeinde berechtigt, zum Zwecke der
LLM die Stadtgemeinde berechtigt, zum Zwecke der Selbstverwaltung der Bürger, eigene Verordnungen zu erlassen, sofern diese nicht mit Bundes- oder Landesgesetzen kollidieren.

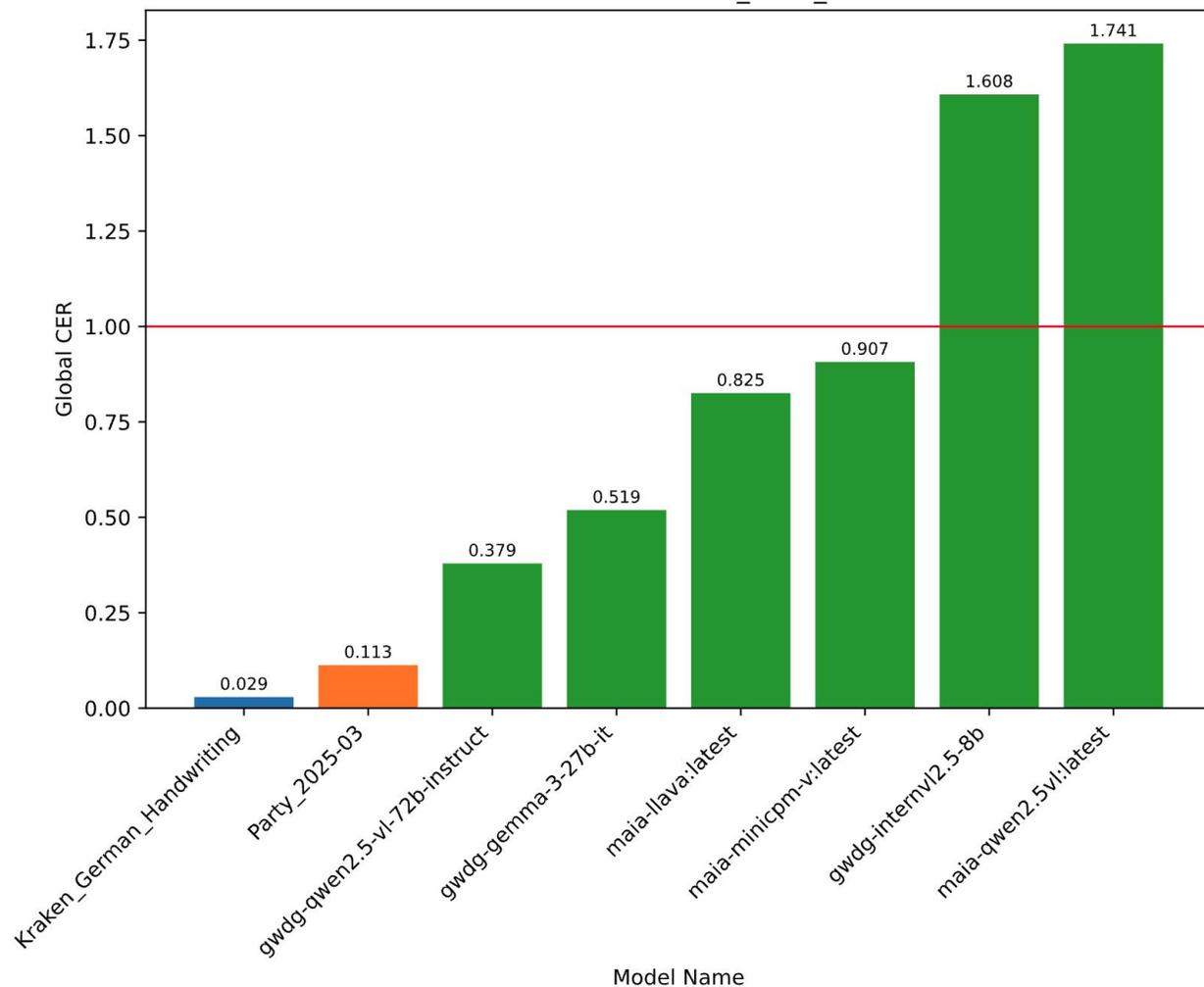
TextLine r218

GT c) Vertrag wegen Verforgung der
LLM c) Vertrag wegen Verjorgung der Witwe und der Kinder des Versicherten. § 11. Der Versicherer hat Anspruch auf Rückforderung der geleisteten Versorgungsbeiträge, wenn der Versicherte vor dem Eintritt des Versorgungsfalles verstirbt, ohne einen Anspruch auf Versorgung zu hinterlassen. § 12. Der Versicherer ist berechtigt, die Höhe der Versorgungsbeiträge an die tatsächlichen Verhältnisse anzupassen. § 13. Die Versorgungsbeiträge sind jährlich im voraus zu entrichten. § 14. Der Versicherer kann die Versorgung verweigern, wenn der Versicherte vorsätzlich falsche Angaben gemacht hat. § 15. Die vorliegenden Allgemeinen Versorgungsbedingungen gelten ergänzend zu den Einzelvereinbarungen. § 16. Gerichtsstand ist Berlin. Berlin, den 1. Januar 1928. Im Namen der Versicherungsanstalt: gez. Dr. Schmidt. Direktor.

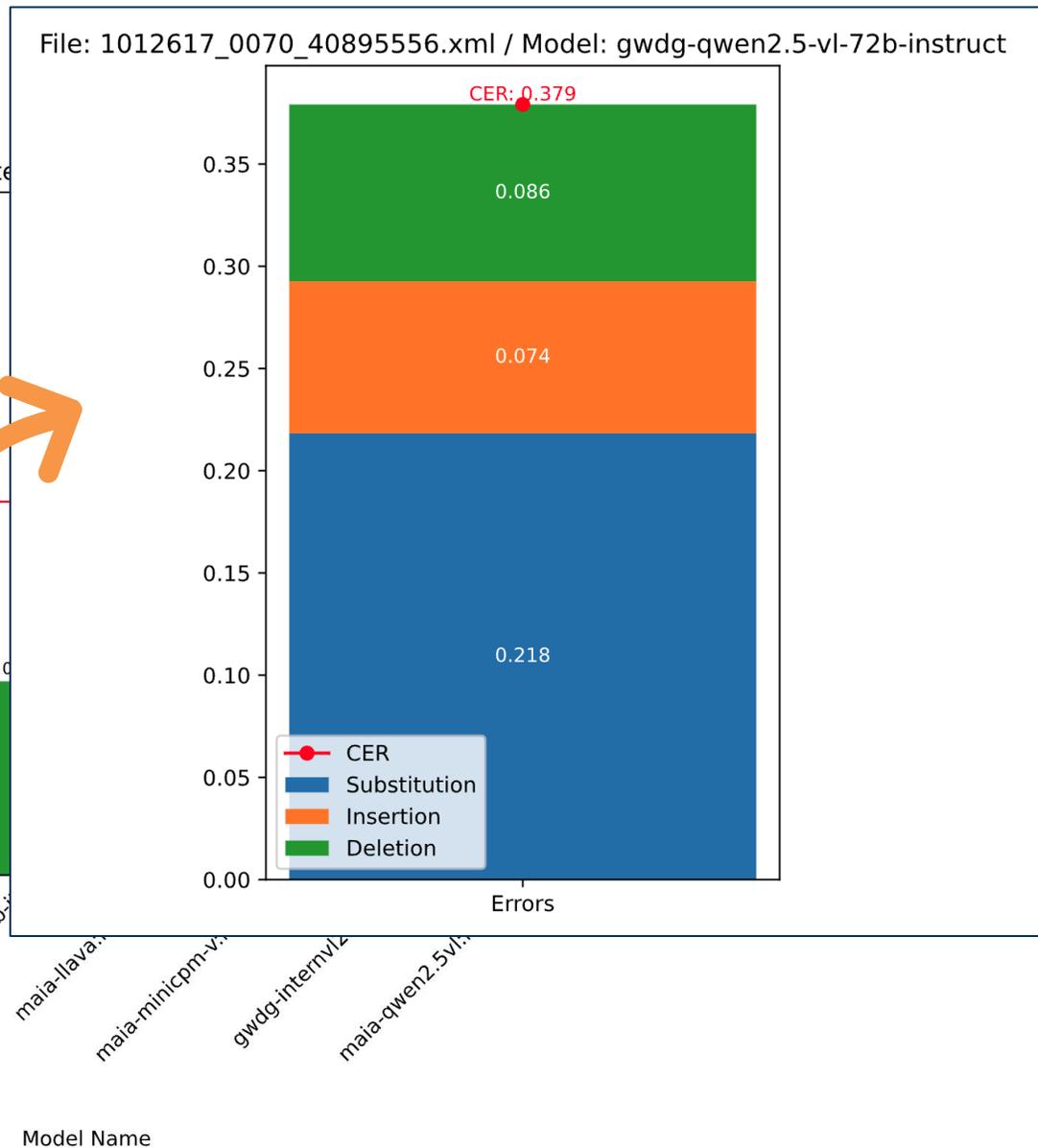
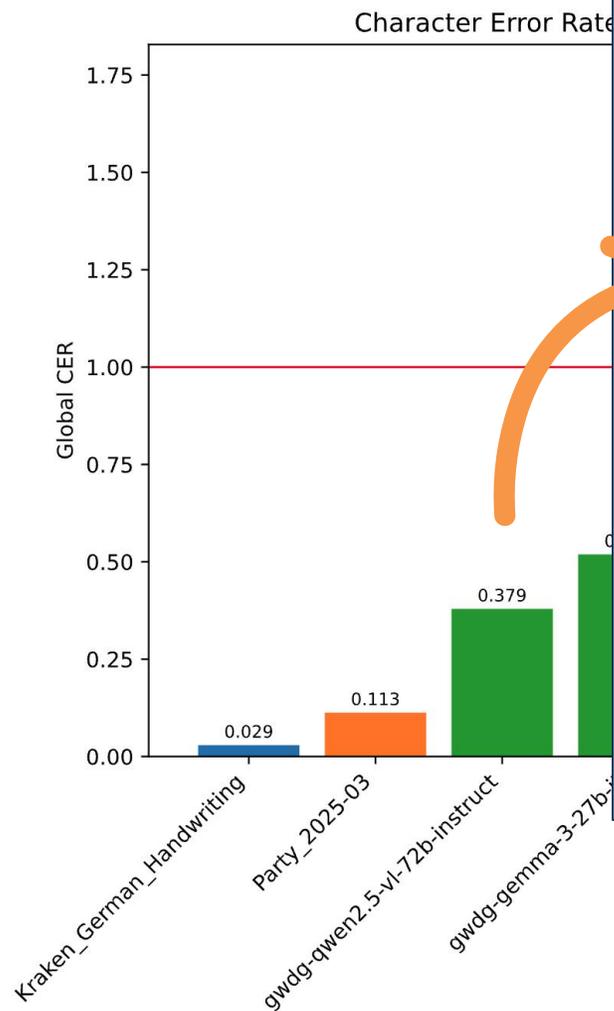
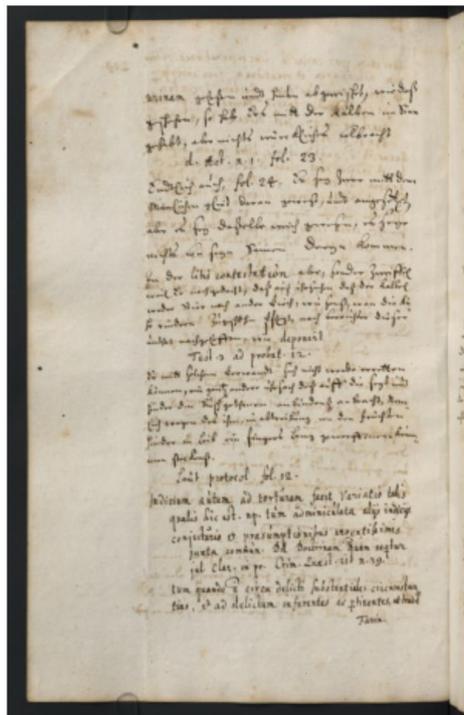
Auswertung: Handwriting



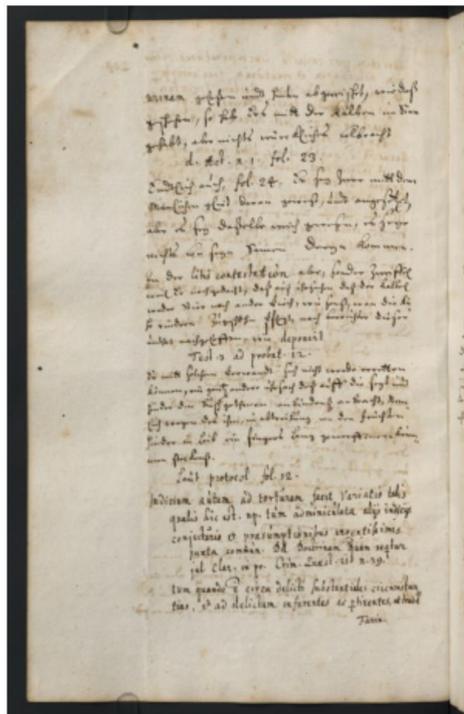
Character Error Rate: 1012617_0070_40895556.xml



Auswertung: Handwriting



Auswertung: Handwriting



Qwen 2.5 VL 72B instruct

TextLine r1l1

GT urinam gelaßen undt hinten abgewifchet, wie daß
LLM wirnam griften und furten ubgruift, worinDyß

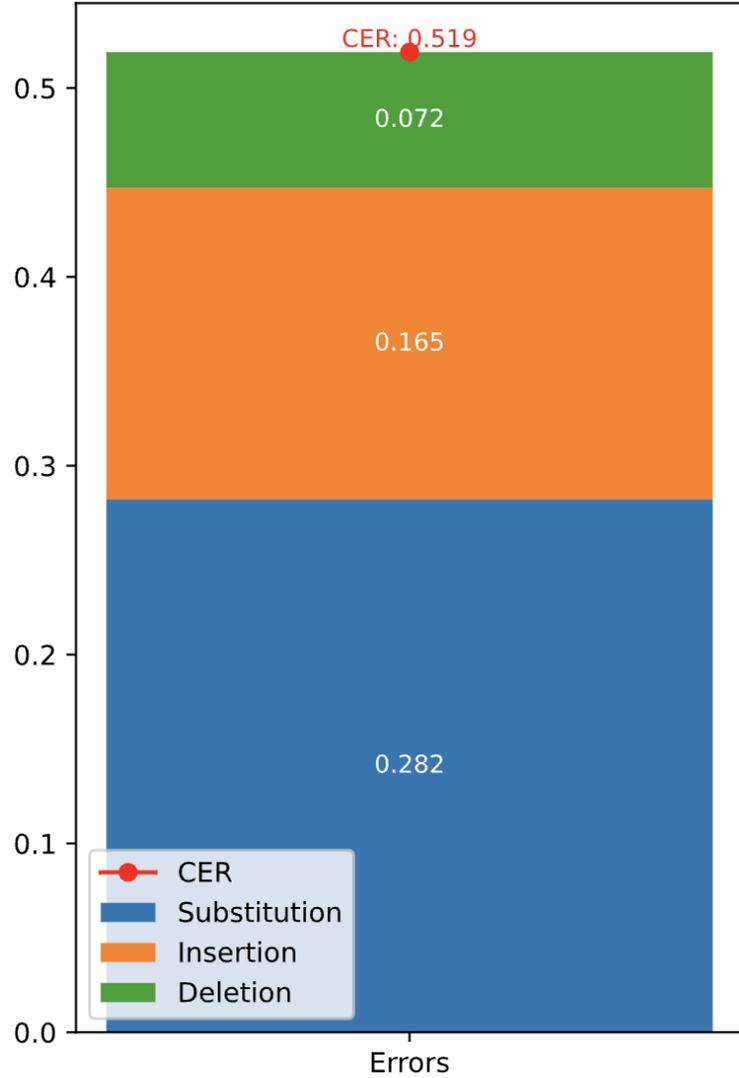
TextLine r1l2

GT gefchehen, fo habe Ers mitt der kalben im Sinn
LLM yntfon,fofebr26mittDnKalboununDian

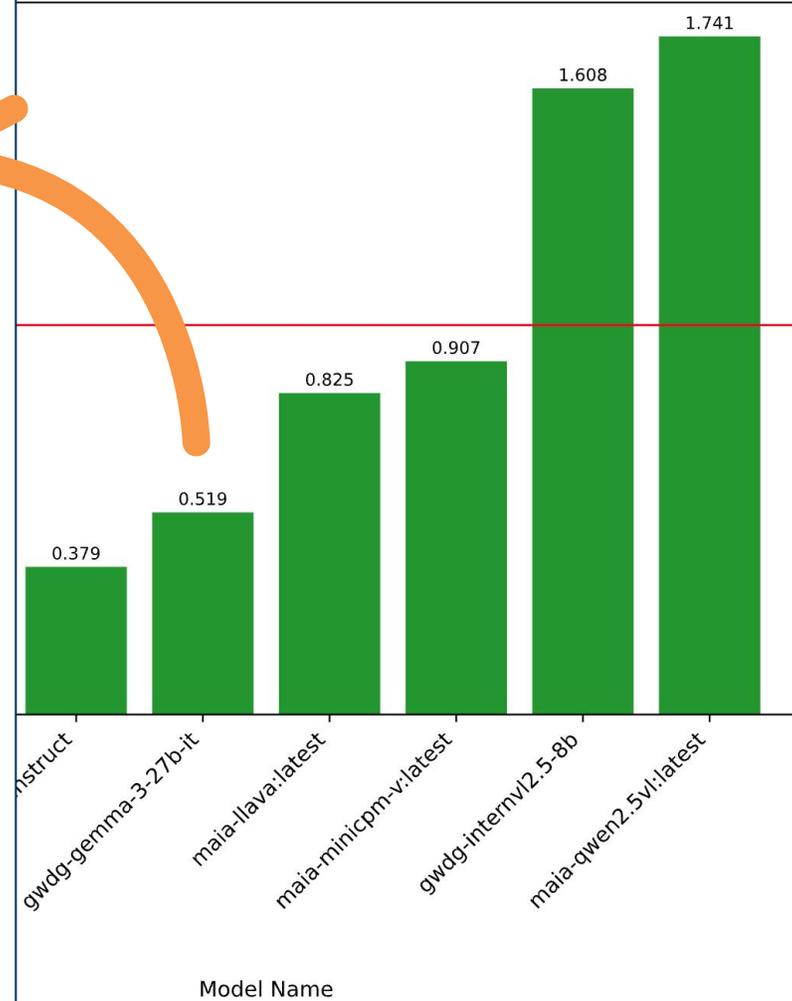
TextLine r1l4

GT d. Act. n. 1. fol. 23.
LLM d.Act.n.1.fol.23.

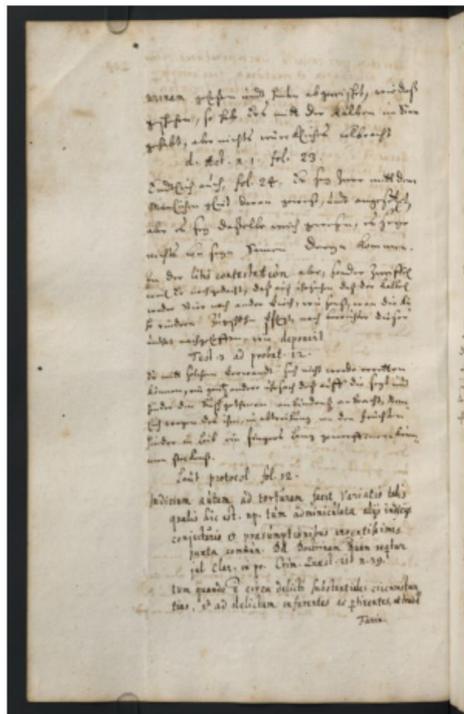
File: 1012617_0070_40895556.xml / Model: gwdg-gemma-3-27b-it



Character Error Rate: 1012617_0070_40895556.xml



Auswertung: Handwriting



Gemma 3 27B

TextLine r1|1

GT urinam gelaßen undt hinten abgewifchet, wie daß
LLM wirnam gefchen müst friben abzurijfft, wirßß

TextLine r1|2

GT gefchehen, fo habe Ers mitt der kalben im Sinn
LLM zu schiffen, so sollen Sie mit den Kellnern in Dän marken.

TextLine r1|4

GT d. Act. n. 1. fol. 23.
LLM d. #ct. n.i. fol. 23.

Einschränkungen



- **Zeilenbasierte** Erkennung führt zu Einschränkungen:
 - Line-based Erkennung ist für LLMs atypisch (page-wise Training)
 - Dadurch Kontextreduktion und verstärkte Halluzination (bspw. bei Zeilenumbrüchen)
- **Historische Domäne** ist für LLMs (noch) problematisch
- **Möglicher Mehrwert** von LLMs in Auswertung eher im Hintergrund:
 - Gesamte Seite als Kontext
 - Mischung von Hand- und Druckschrift sekundär
 - Layoutprobleme (Tabellen, Mehrspaltigkeit) sekundär

Fazit



- Kraken / Party mit besseren Ergebnissen (Einschränkungen beachten!)
- **Token Repetition Effekt** (→ max_tokens)
 - Repetition Penalty Parameter?
- Nicht Erkennung von Zeilen recht hoch
- **Normalisierung** historischer Symbole und Schreibweisen
- Hoher **Ressourcen-Verbrauch** der LLMs im Vergleich zu Kraken
- **Performance** (Zeit)

Danke für Eure Aufmerksamkeit!



Kontakt:

Larissa Will

larissa.will@uni-mannheim.de

Thomas Schmidt

thomas.schmidt@uni-mannheim.de

Jan Kamlah

jan.kamlah@uni-mannheim.de

<https://www.bib.uni-mannheim.de/lehren-und-forschen/forschungsdatenzentrum>