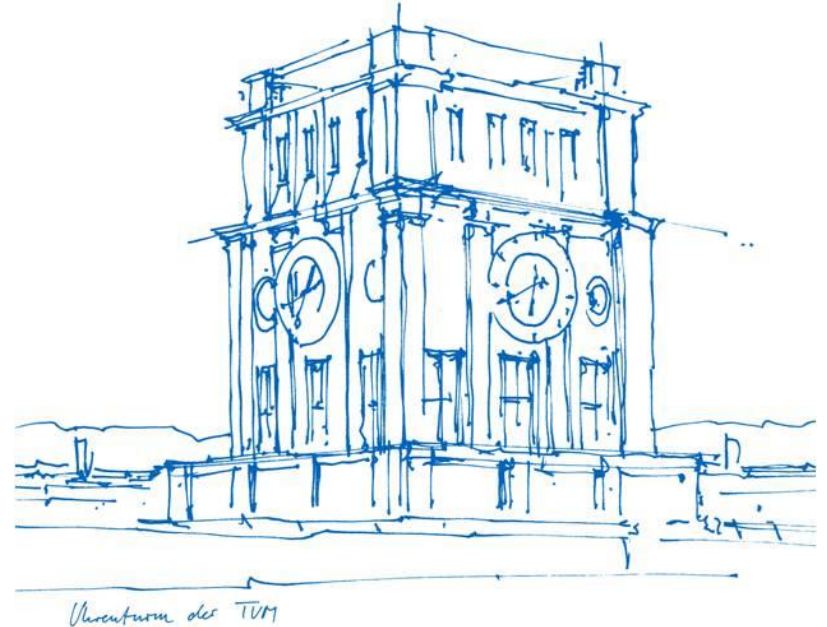


Utilizing Large Language Models for Causal Discovery and Legal Text Interpretation: A Case Study on the German GoZ

Jonas Zausinger

Technische Universität München

München, 13. März 2024



Motivation



Reasoning ability, particularly in extracting causal relationships, is critical for a plethora of applications, including scientific discoveries and legal interpretation



With the evolution of LLMs, there arises an opportunity to harness these models for those advanced tasks

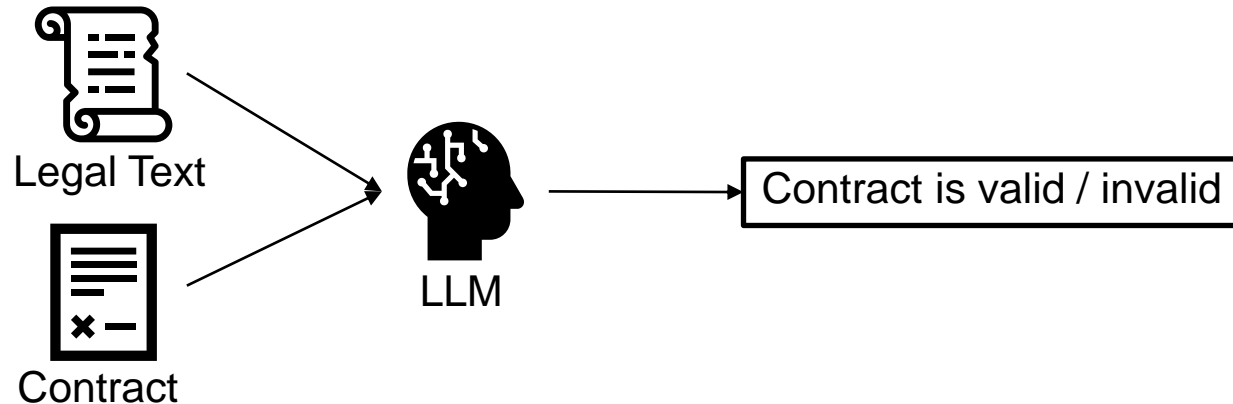
1. How strong are LLMs in causal discovery?

2. Can we use the causal and logical discovery and reasoning capabilities of LLMs to solve a real-world use case in the domain of legal texts?



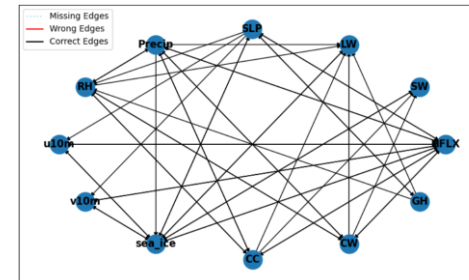
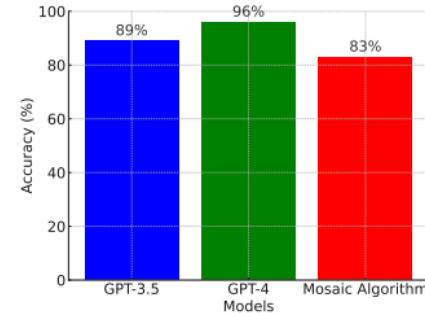
Motivation

Can LLMs be used to extract causal and logic relationships and knowledge from legal texts with such efficiency that it is possible to automatically reason with it and automatically determine the legal correctness of a contract according to the rules of the text?



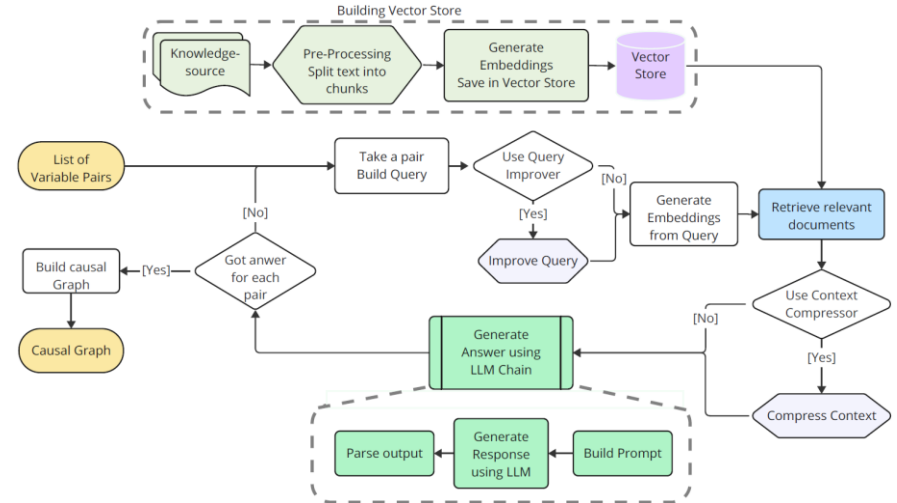
Related Work – Causal Discovery

- Kiciman et al. [15] found that LLMs are able to outperform previous state-of-the-art algorithms
- Pairwise Causal Discover on the Tübingen cause-effect pairs dataset
 - Altitude → Temperature
 - Age → Diastolic blood pressure
 - Number of cars ← Type of day
- Full graph discovery on the arctic sea dataset
 - This scientific dataset is about the drivers of arctic sea ice coverage
 - 12 variables with 48 edges



Methodology – Causal Discovery

- Retrieval-augmented Language Models
 - Optimal Knowledge Source
 - Web-scraped knowledge source
- Query Improver
 - MultiQueryRetriever
 - Forward-Looking Active Retrieval augmented generation (FLARE)
- Context Compression



Results – Causal Discovery

- Pairwise Causal Discover on the Tübingen cause-effect pairs dataset

Prompt Set.	ZS	OS	ZS	OS	OS + CC	GPT-4 ZS	GPT-4 ZS
Knowledge Base	-	-	DP	DP	Automatic	-	DP
Accuracy	0.83	0.86	0.87	0.94	0.7	0.96	1

Baseline
Baseline

- Full graph discovery on the arctic sea dataset

Prompt Set.	ZS	ZS
KB	-	DP
# Edges	38	24
NHD	0.35	0.28
NHD Ratio	0.58	0.56
Acc.	0.62	0.7
F1	0.42	0.44

Baseline

→ Additional contextual information improves causal discovery performance of LLM

Related Work – Automatic Reasoning with Legal Text



Understand legal texts by means of neural networks

- Low performance of traditional machine reading models
- Law-based single declarations vs. statistical language learning.

Try to manually or automatically convert legal text into a form of logic

- Disadvantage: High manual effort
- Advantage: Direct logic queries over network-weights encoded legal knowledge.
- Manual translation of laws into expert systems, primarily using Prolog.

Related Work – Automatic Reasoning with Legal Text



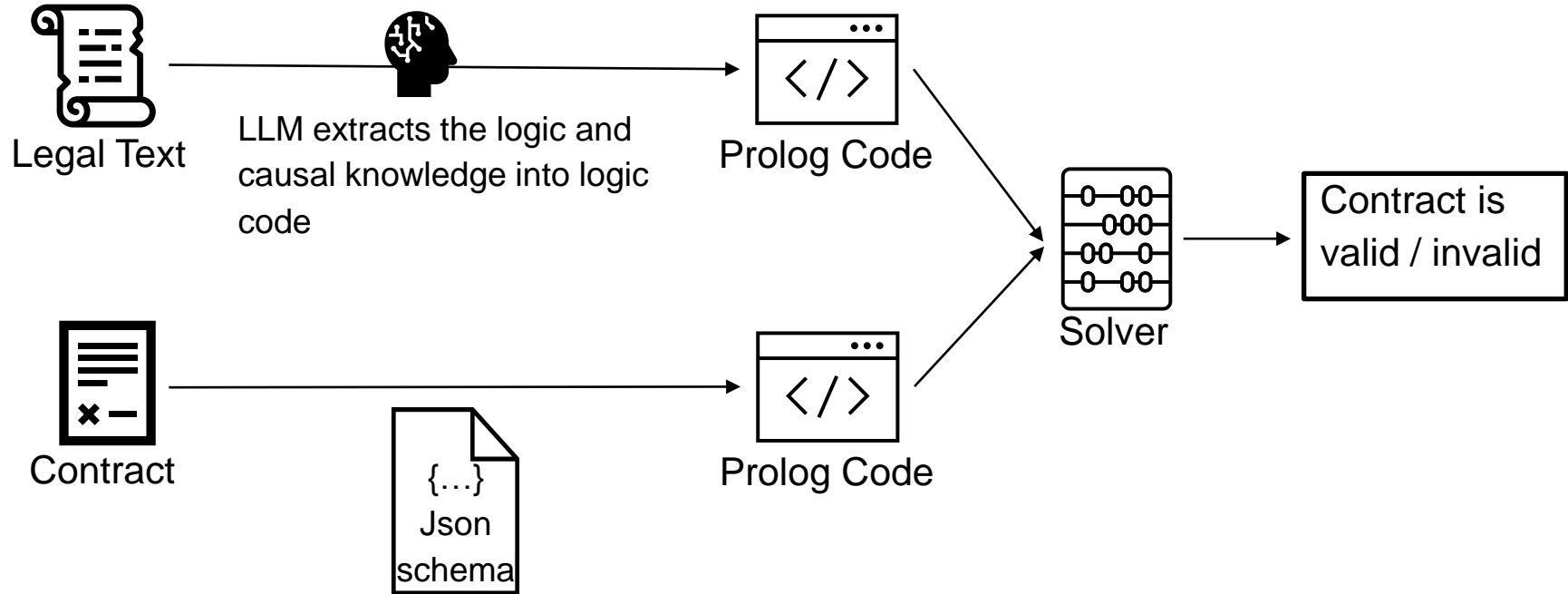
Why Prolog?

- Rules are represented as Horn clauses: $A : - B_1, \dots, B_n$.
- Facts: B_1 .
- Declarative Nature
- Predicate Calculus is inherently suited for representing legal arguments
- Allows to query both the explicitly stated information and the implications of this information

Reasoning with Prolog and LLMs

- Performance of LLMs for logical reasoning tasks can be improved by using Prolog
- LLMs translate the problem into Prolog
- A symbolic solver performs inference on the Prolog code

Methodology



Gebührenordnung für Zahnärzte (GoZ)

§ 5

Bemessung der Gebühren für Leistungen des Gebührenverzeichnisses

(1) Die Höhe der einzelnen Gebühr bemisst sich nach dem Einfachen bis Dreieinhalbfachen des Gebührensatzes. Gebührensatz ist der Betrag, der sich ergibt, wenn die Punktzahl der einzelnen Leistung des Gebührenverzeichnisses mit dem Punktwert vervielfacht wird. Der Punktwert beträgt 5,62421 Cent. Bei der Bemessung von Gebühren sind sich ergebende Bruchteile eines Cents unter 0,5 abzurunden und Bruchteile von 0,5 und mehr aufzurunden; die Rundung ist erst nach der Multiplikation mit dem Steigerungsfaktor nach Satz 1 vorzunehmen.

B Prophylaktische Leistungen

Allgemeine Bestimmungen

Prophylaktische Leistungen nach Abschnitt B sind nur bei Einzelunterweisung (Individualprophylaxe) berechnungsfähig; bei Gruppenunterweisung (Gruppenprophylaxe) sind sie nicht berechnungsfähig.

Nummer	Leistung	Punktzahl	1,0-fach €	2,3-fach €	3,5-fach €
0040	Aufstellung eines schriftlichen Heil- und Kostenplans bei kieferorthopädischer Behandlung oder bei funktionsanalytischen und funktionstherapeutischen Maßnahmen nach Befundaufnahme und Ausarbeitung einer Behandlungsplanung <i>Die Leistungen nach den Nummern 0030 und 0040 sind nicht nebeneinander berechnungsfähig.</i>	250	14,06	32,34	49,21
0050	Abformung oder Teilabformung eines Kiefers für ein Situationsmodell einschließlich Auswertung zur Diagnose oder Planung	120	6,75	15,52	23,62
0060	Abformung beider Kiefer für Situationsmodelle und einfache Bissfestigung einschließlich Auswertung zur Diagnose oder Planung <i>Die Nebeneinanderberechnung der Leistungen nach den Nummern 0050 und 0060 ist in der Rechnung zu begründen.</i>	260	14,62	33,63	51,18
0065	Optisch-elektronische Abformung einschließlich vorbereitender Maßnahmen, einfache digitale Bissregistrierung und Archivierung, je Kieferhälfte oder Frontzahnbereich <i>Neben der Leistung nach der Nummer 0065 sind konventionelle Abformungen nach diesem Gebührenverzeichnis für dieselbe Kieferhälfte oder denselben Frontzahnbereich nicht berechnungsfähig.</i>	80	4,50	10,35	15,75
0070	Vitalitätsprüfung eines Zahnes oder mehrerer Zähne einschließlich Vergleichstest, je Sitzung	50	2,81	6,47	9,84
0080	Intraorale Oberflächenanästhesie, je Kieferhälfte oder Frontzahnbereich	30	1,69	3,88	5,91

Gebührenordnung für Zahnärzte (GOZ) 2012

RECHNUNG

Rechnungs-Nr.: 92545-16
(bitte bei Zahlungen angeben)
Patient/in.: Meike Schnitt
Geburtsdatum: 16.04.1955

Datum: 11.11.2016

Sehr geehrte Frau Muster,

für die erbrachten Leistungen in der Zeit vom 02.09. bis 06.09.16 erlaube ich mir folgende
Beträge in Rechnung zu stellen: **EUR 257,51**

Datum	Zahn	Anz.	Nr.	Leistung	Faktor	EUR
02.09.16		1	1000	Erstellung eines Mundhygienestatus und eingehende Unterweisung zur Vorbeugung gegen Karies und parodontale Erkrankungen, Dauer mindestens 25 Minuten	1,0	11,25
02.09.16	17,16,15, 14,13,12, 11,21,22, 23,24,25, 26,27,37, 36,35,34, 33,32,31, 41,42,43, 44,45,46, 47	28	1040	Professionelle Zahnreinigung	2,0	88,20
06.09.16	27	1	2120	Präparieren einer Kavität und Restauration mit Kompositmaterialien, in Adhäsivtechnik (Konditionieren), mehr als dreiflächig, ggf. einschließlich Mehrschichttechnik, einschließlich Polieren, ggf. einschließlich Verwendung von Inserts Begründung: Überdurchschnittlicher Schwierigkeitsgrad und Zeitaufwand wegen erschwertem Anlegen von Matrizen/Bändern durch Zahnengstand und tiefer, sehr schwer einsehbarer Approximalkavität.	3,5	151,57
	27	1		abzgl. Bema-Sachleistung Nr. 13a bis 13d		- 54,47
Zwischensumme Zahnarzt-Honorar:						251,02
abzgl. Kassenanteil nach Bema:						- 54,47
Eigenlabor:						60,96
(s. beigefügte Materialrechnung)						
Rechnungsbetrag: € 257,51						
umsatzsteuerfrei nach § 4 Nr. 14a UStG						

Der Berechnung der Honorarleistung liegt die **Gebührenordnung für Zahnärzte (GOZ)** zugrunde. Bitte bezahlen Sie den Rechnungsbetrag bis spätestens zum 03.12.2016 auf das unten aufgeführte Konto. Bei der Überweisung des Rechnungsbetrages geben Sie bitte die Rechnungsnummer an.

Methodology – Legal Text Reasoning Pipeline

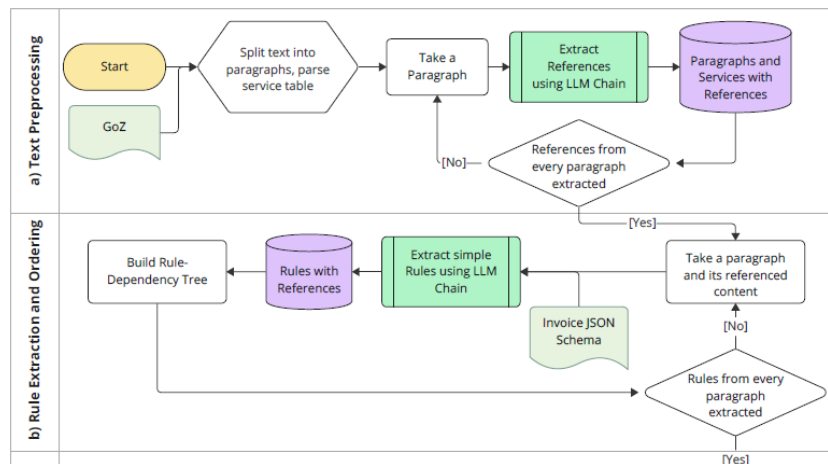
- Many References to other paragraphs
 - Rules are sometimes imprecise and vague or can not be checked programmatically
- ➔ They must be extracted and filtered

§ 10

Fälligkeit und Abrechnung der Vergütung;
Rechnung

(1) Die Vergütung wird fällig, wenn dem Zahlungspflichtigen eine dieser Verordnung entsprechende Rechnung nach der Anlage 2 erteilt worden ist. Künftige Änderungen der Anlage 2 werden durch das Bundesministerium für Gesundheit durch Bekanntmachung veröffentlicht.

(2) Innerhalb des Gebührenrahmens sind die Gebühren unter Berücksichtigung der Schwierigkeit und des Zeitaufwandes der einzelnen Leistung sowie der Umstände bei der Ausführung nach billigem Ermessen zu bestimmen. Die Schwierigkeit der einzelnen Leistung kann auch



Methodology – Legal Text Reasoning Pipeline

Translation of Rules into Prolog code

§ 5

Bemessung der Gebühren für Leistungen des Gebührenverzeichnisses

.....
(1) Die Höhe der einzelnen Gebühr bemisst sich nach dem Einfachen bis Dreieinhalbfachen des Gebührensatzes. Gebührensatz ist der Betrag, der sich ergibt, wenn die Punktzahl der einzelnen Leistung des Gebührenverzeichnisses mit dem Punktwert vervielfacht wird. Der Punktwert beträgt 5,62421 Cent. Bei der Bemessung von Gebühren sind sich ergebende Bruchteile eines Cents unter 0,5 abzurunden und Bruchteile von 0,5 und mehr aufzurunden; die Rundung ist erst nach der Multiplikation mit dem Steigerungsfaktor nach Satz 1 vorzunehmen.

(2) Innerhalb des Gebührenrahmens sind die Gebühren unter Berücksichtigung der Schwierigkeit und des Zeitaufwandes der einzelnen Leistung sowie der Umstände bei der Ausführung nach billigem Ermessen zu bestimmen. Die Schwierigkeit der einzelnen Leistung kann auch durch die Schwierigkeit des Krankheitsfalles begründet sein. Bemessungskriterien, die bereits in der Leistungsbeschreibung berücksichtigt worden sind, haben hierbei außer Betracht zu bleiben. Der 2,3fache Gebührensatz bildet die nach Schwierigkeit und Zeitaufwand durchschnittliche Leistung ab; ein Überschreiten dieses Gebührensatzes ist nur zulässig, wenn Besonderheiten der in Satz 1 genannten Bemessungskriterien dies rechtfertigen; Leistungen mit unterdurchschnittlichem Schwierigkeitsgrad oder Zeitaufwand sind mit einem niedrigeren Gebührensatz zu berechnen.



The amount of the individual fee is based on one to three and a half times (multiplier) the fee rate. The fee rate is the rounded amount that results when the number of points for the individual service in the fee schedule is multiplied by the point value. The point value is 5.62421 cents.



The 2.3-fold fee rate reflects the average service in terms of difficulty and time required; exceeding this fee rate is only permitted if this is justified by special features of the assessment criteria mentioned in sentence 1; services with a below average degree of difficulty or time required are to be charged at a lower fee rate

```
point_value(0.0562421).  
% Predicate to calculate the fee for a service  
fee(Service, Multiplier, Fee) :-  
    point_value(PointValue),  
    service_points(Service, Points),  
    Fee is round((Points * PointValue *  
        Multiplier)*100)/100,  
    Fee >= round((Points * PointValue)*100)/100,  
    Fee <= round((Points * PointValue * 3.5)*100)  
        /100.  
% Predicate to check if an invoice is invalid  
    based on Rule 1  
is_invoice_invalid(Invoice) :-  
    member(Service-Multiplier, Invoice),  
    \+ fee(Service, Multiplier, _),  
    print('Invoice is invalid: Fee for service is  
        not within allowed range.').
```

This rule states that the 2.3-fold fee represents the average service. If the fee exceeds this, it must be justified by the difficulty or time required for the service. Services with below-average difficulty or time requirement should be calculated with a lower fee. This rule is subjective and depends on the individual service, so it cannot be translated into Prolog without making assumptions.

Methodology – Legal Text Reasoning Pipeline

Translation must be grounded on invoice structure

§ 5

Bemessung der Gebühren für Leistungen des Gebührenverzeichnisses

(1) Die Höhe der einzelnen Gebühr bemisst sich nach dem Einfachen bis Dreieinhalbfachen des Gebührensatzes. Gebührensatz ist der Betrag, der sich ergibt, wenn die Punktzahl der einzelnen Leistung des Gebührenverzeichnisses mit dem Punktwert vervielfacht wird. Der Punktwert beträgt 5,62421 Cent. Bei der Bemessung von Gebühren sind sich ergebende Bruchteile eines Cents unter 0,5 abzurunden und Bruchteile von 0,5 und mehr aufzurunden; die Rundung ist erst nach der Multiplikation mit dem Steigerungsfaktor nach Satz 1 vorzunehmen.

(2) Innerhalb des Gebührenrahmens sind die Gebühren unter Berücksichtigung der Schwierigkeit und des Zeitaufwandes der einzelnen Leistung sowie der Umstände bei der Ausführung nach billigem Ermessen zu bestimmen. Die Schwierigkeit der einzelnen Leistung kann auch durch die Schwierigkeit des Krankheitsfalles begründet sein. Bemessungskriterien, die bereits in der Leistungsbeschreibung berücksichtigt worden sind, haben hierbei außer Betracht zu bleiben. Der 2,3fache Gebührensatz bildet die nach Schwierigkeit und Zeitaufwand durchschnittliche Leistung ab; ein Überschreiten dieses Gebührensatzes ist nur zulässig, wenn Besonderheiten der in Satz 1 genannten Bemessungskriterien dies rechtfertigen; Leistungen mit unterdurchschnittlichem Schwierigkeitsgrad oder Zeitaufwand sind mit einem niedrigeren Gebührensatz zu berechnen.



The amount of the individual fee is based on one to three and a half times (multiplier) the fee rate. The fee rate is the rounded amount that results when the number of points for the individual service in the fee schedule is multiplied by the point value. The point value is 5.62421 cents.



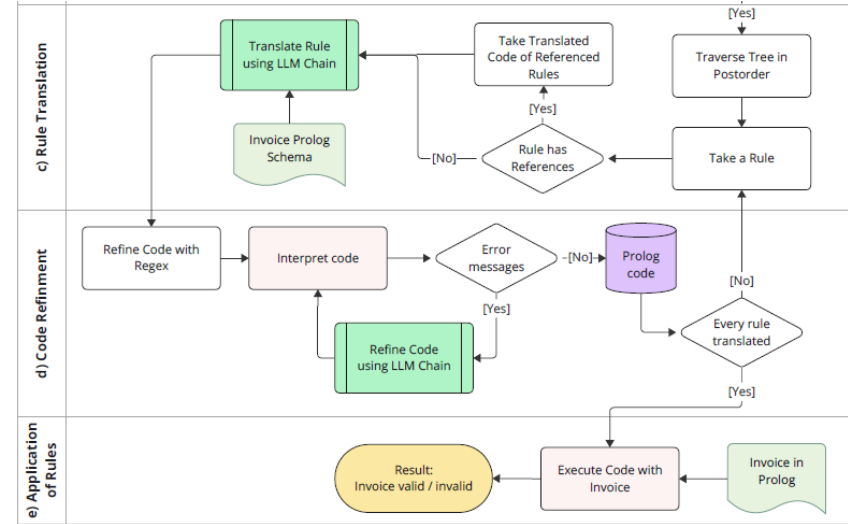
The 2.3-fold fee rate reflects the average service in terms of difficulty and time required; exceeding this fee rate is only permitted if this is justified by special features of the assessment criteria mentioned in sentence 1; services with a below average degree of difficulty or time required are to be charged at a lower fee rate

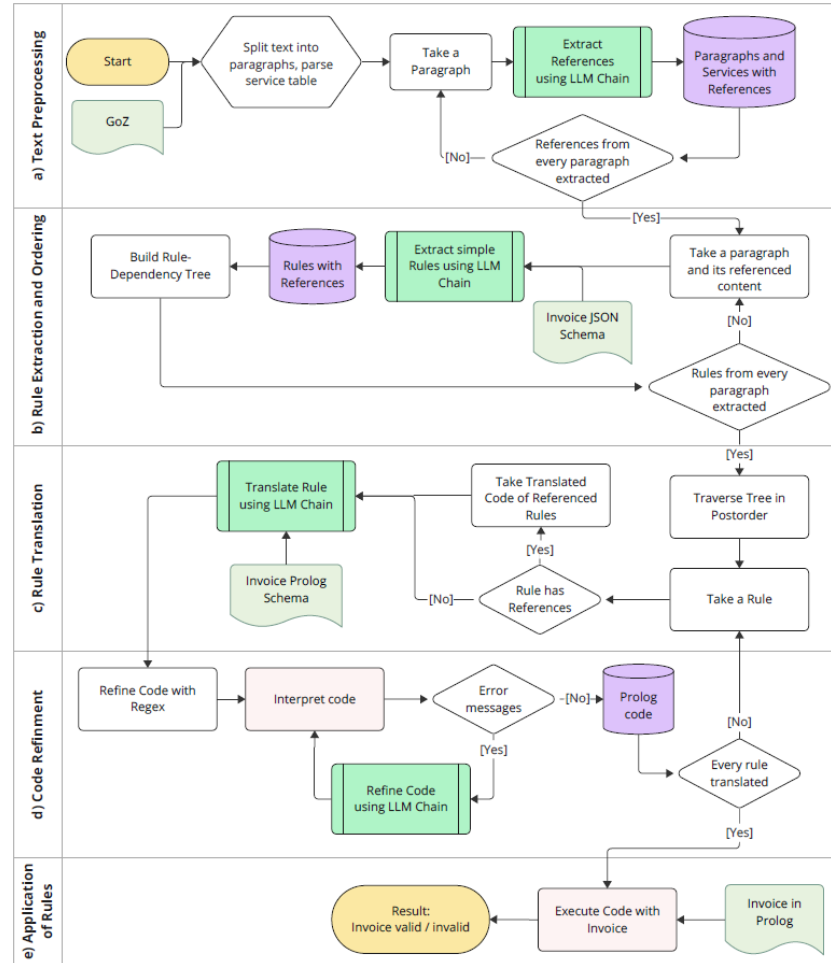
```
calculate_charge(PointScore, Multiplier, Charge)
:-
    PointValue = 0.0562421,
    CalculatedCharge is round(PointValue *
        Multiplier * PointScore * 100) / 100,
    Charge = CalculatedCharge.
is_invoice_invalid(Invoice) :-
    Invoice = invoice(Services, _, _, _),
    member(service(_, _, Multiplier, PointScore,
        Charge, _, _, _), Services),
    \+ calculate_charge(PointScore, Multiplier,
        Charge),
    print('Invalid invoice: The charge for a
        service does not match the calculated
        charge. '),
    true.

is_multiplier_valid(Multiplier, Justification) :-
    Multiplier <= 2.3;
    (Multiplier > 2.3, Justification \= '').
is_invoice_invalid(Invoice) :-
    Invoice = invoice(Services, _, _, _),
    member(service(_, _, Multiplier, _, _, _,
        Justification, _), Services),
    \+ is_multiplier_valid(Multiplier,
        Justification),
    print('Invalid invoice: The multiplier for a
        service exceeds 2.3 without
        justification. '),
    true.
```

Methodology – Legal Text Reasoning Pipeline

- Translation must be grounded on invoice structure
- Refinement step to correct wrong code
- Translated Rules can be applied to invoices





Methodology – Legal Text Reasoning

GoZ Commentary

GOZ Nr. 5190

Funktionelle Abformung des Unterkiefers mit individuellem Löffel

Punktzahl	540 Punkte		
Faktor	1,0-fach	2,3-fach	3,5-fach
Gebühr in €	30,37 €	69,85 €	106,30 €

Kommentar zur Leistungsbeschreibung

Die Durchführung einer funktionellen Abformung ist im Rahmen der Herstellung einer totalen Unterkieferprothese oder einer Unterkiefer-Deckprothese (Cover-Denture) erforderlich. Durch sie wird u. a. der Übergang von beweglicher zu unbeweglicher Schleimhaut dargestellt, um einen Ventirand herstellen und damit die Saugwirkung einer Prothese erzielen zu können. Als individueller Löffel kann auch eine entsprechend vorbereitete vorhandene Prothese dienen. Auch bei der Herstellung von Teilprothesen kann eine Funktionsabformung zur Darstellung der Weichgewebe notwendig werden. Ebenso kann bei der Rebasierung von abnehmbarem Zahnersatz sowie bei Erweiterungen von Prothesen mit funktioneller Randgestaltung die funktionelle Abformung nach dieser Gebührenummer erforderlich werden. Die Abdruckdesinfektion ist als zahntechnische Leistung nach § 9 berechnungsfähig. Das hat auch dann zu gelten, wenn die Abformung Leistungsbestandteil einer anderen Leistung und nicht gesondert berechnungsfähig ist.

Zusätzlich berechnungsfähige Leistungen

- Anatomische Abformungen GOZ 5170
- Funktionsanalytische und funktionstherapeutische Maßnahmen GOZ 8000 ff.
- Erneuerung des Sekundärteils einer Teleskopkrone GOZ 5100
- u. v. m.

GOZ 1988

GOZ Nr.: 519

Funktionelle Abformung des Unterkiefers mit individuellem Löffel

Punktzahl	540 Punkte		
Faktor	1,0-fach	2,3-fach	3,5-fach
Gebühr in €	30,37 €	69,85 €	106,29 €

Results – Legal Text Reasoning

- Self prepared set of sample invoices for each rule
 - Some of them are valid according to the rule
 - Some of them violate the rule
- The set of sample invoices for a rule is structured in such a way that they can fully test the rule
- A rule is only counted as correctly translated when it
 - Determines all valid sample invoices of this rule as correct
 - Determines all invalid sample invoices of this rule as not correct
- Rule translation accuracy (RT Acc.): Ratio of correctly translated rules.
- Simple retrieval augmented LLM (RA LLM) as baseline

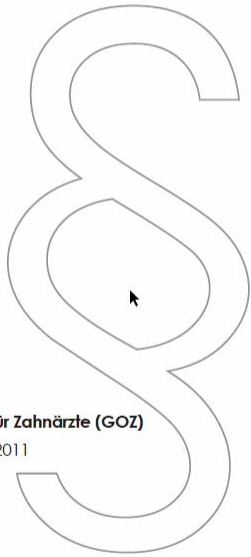
	R.A. LLM	Our Method	Our Method + Commentary	R.A. LLM + Commentary
RT Acc.	0.6 ± 0.09	0.42 ± 0.02	0.67 ± 0.05	0.34 ± 0.07

Alle Tools x

- PDF exportieren
 - PDF bearbeiten
 - PDF erstellen
 - Dateien zus. führen
 - Seiten verwalten
 - Kommentare hinzufügen
 - E-Signaturen anfordern
 - Scan & OCR
 - PDF-Datei schützen
 - PDF-Datei schwärzen
 - PDF komprimieren
 - Formular vorbereiten
 - Ausfüllen und Signieren
- Mehr anzeigen

Konvertiere, bearbeite und e-signiere PDF-Formulare.

Gratis 7-Tage-Testversion



Gebührenordnung für Zahnärzte (GOZ)
Stand 5. Dezember 2011

Conclusion

Limitations

- The performance heavily relies on the capabilities of the LLMs used
- Small translation errors lead to incorrect application of rules
- Requires numerous calls to the LLM.

Future Work

- Reduce erroneous rule translations
- More efficient token management
- Fully automated contract review and due diligence application.

Conclusion

- The ability of LLMs to perform causal discovery can be improved by retrieval augmented generation with a meaningful knowledge base
- LLMs can be used as a very useful assistant to convert a legal text into a formal representation.