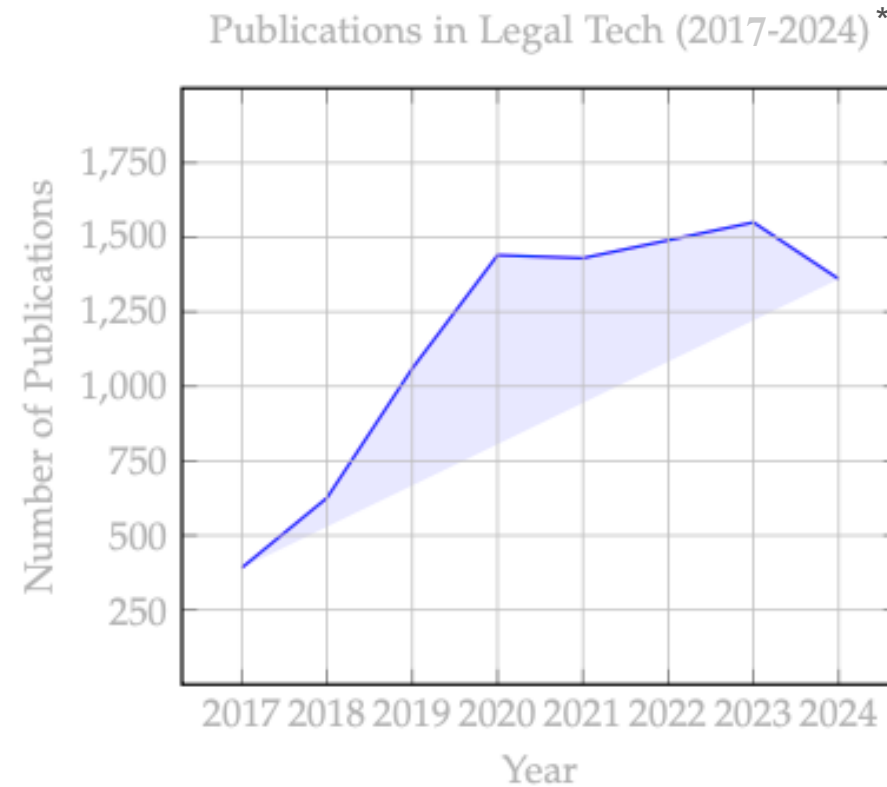


Controlling the Evolution of a Legal Knowledge Base for an In-Context Learning AI System - Design and Implementation

John Graf von Hardenberg

13 January 2025, Bachelor's Thesis Final Presentation

Chair of Software Engineering for Business Information Systems (sebis)
Department of Computer Science
School of Computation, Information and Technology (CIT)
Technical University of Munich (TUM)
www.matthes.in.tum.de



The global legal technology market has grown significantly in recent years and GenAI will accelerate this growth, meaning the market will reach \$50 billion in value by 2027[1].

* until 10.12.2024

TUM - Oliver Wardas

Erstellt von: Oliver Wardas Erstellt am: 08.03.24

1 Fairness-Bewertung

2 Alternative Formulierungen

BACK

NEXT

Filter:

IN ORDNUNG 12

UNFAIR 1

UNZULÄSSIG 1

ZURÜCKSETZEN

NEU LADEN

Technische Universität München

1. Ausfertigung

ARBEITSVERTRAG

Zwischen

dem Freistaat Bayern,

vertreten durch den Präsidenten der Technischen Universität München

und

Herrn Oliver Wardas, M.Sc.RWTH

wird folgender Arbeitsvertrag geschlossen:

§ 1

Herr Oliver Wardas, M.Sc.RWTH, geb. am 16.06.1999, wird ab 21.11.2022 befristet bis zum 20.11.2023 gemäß § 2 Abs. 1 Satz 1 Wissenschaftszeitvertragsgesetz (WissZeitVG) als Vollbeschäftigter eingestellt.

§ 2

Das Arbeitsverhältnis bestimmt sich nach dem Gesetz über befristete Arbeitsverträge in der Wissenschaft (Wissenschaftszeitvertragsgesetz – WissZeitVG) und nach Art. 21 und 22 des Bayerischen Hochschulpersonalgesetzes (BayHSPG) sowie dem Tarifvertrag für den öffentlichen Dienst der Länder (TV-L) und zur Regelung des Übergangsrechts (TVÜ-Länder) und die diese ergänzenden, ändernden oder ersetzenden Tarifverträgen in der für den Bereich der Tarifgemeinschaft deutscher Länder (TdL) jeweils geltenden Fassung, solange der Freistaat Bayern hieran gebunden ist. Außerdem finden die im Bereich des Freistaates Bayern jeweils geltenden sonstigen einschlägigen Tarifverträge sowie die

Seite 1

1. Ausfertigung

ARBEITSVERTRAG

Zwischen dem Freistaat Bayern, vertreten durch den Präsidenten der Technischen Universität München und Herrn Oliver Wardas, M.Sc.RWTH wird folgender Arbeitsvertrag geschlossen:

Ihre Bewertung

FERTIG

§1

Herr Oliver Wardas, M.Sc.RWTH, geb. am 16.06.1999, wird ab 21.11.2022 befristet bis zum 20.11.2023 gemäß § 2 Abs. 1 Satz 1 Wissenschaftszeitvertragsgesetz (WissZeitVG) als Vollbeschäftigter eingestellt.

§ 2

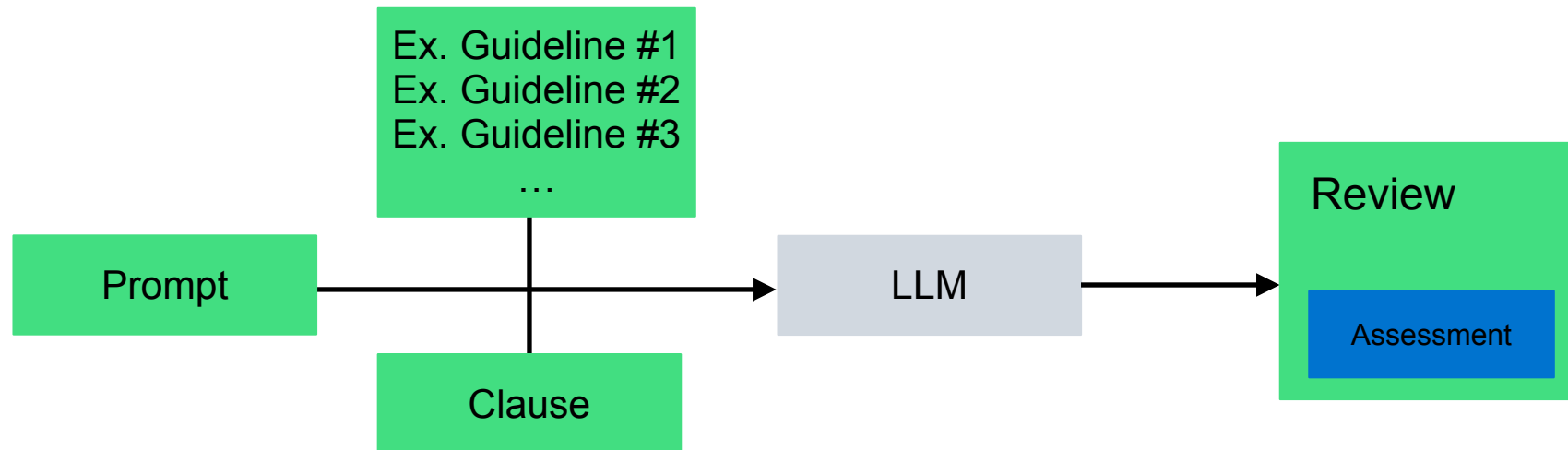
Das Arbeitsverhältnis bestimmt sich nach dem Gesetz über befristete Arbeitsverträge in der Wissenschaft (Wissenschaftszeitvertragsgesetz - WissZeitVG) und nach Art. 21 und 22 des Bayerischen Hochschulpersonalgesetzes (BayHSPG) sowie dem Tarifvertrag für den öffentlichen Dienst der Länder (TV-L), insbesondere nach § 40 TV-L, dem Tarifvertrag zur Überleitung der Beschäftigten der Länder in den TV-L und zur Regelung des Übergangsrechts (TVÜ-Länder) und die diese ergänzenden, ändernden oder ersetzenden Tarifverträgen in der für den Bereich der Tarifgemeinschaft deutscher Länder (TdL) jeweils geltenden Fassung, solange der Freistaat Bayern hieran gebunden ist. Außerdem finden die im Bereich des Freistaates Bayern jeweils geltenden sonstigen einschlägigen Tarifverträge sowie die

Background & Motivation

What is an Examination Guideline?

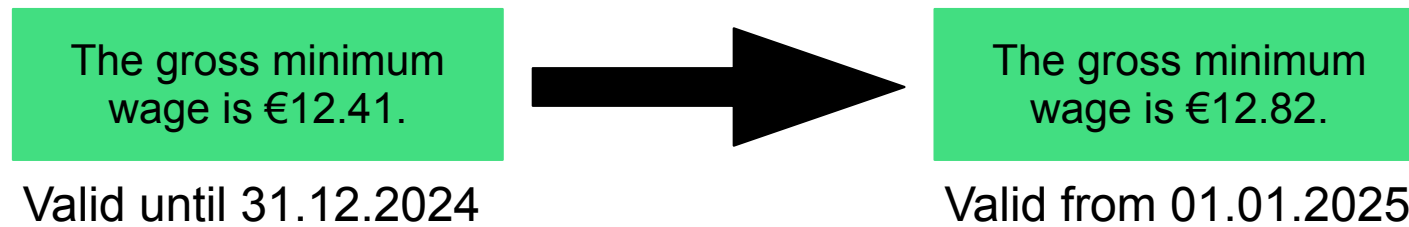
The gross minimum wage is €12.41.

Company training courses must generally be paid for.

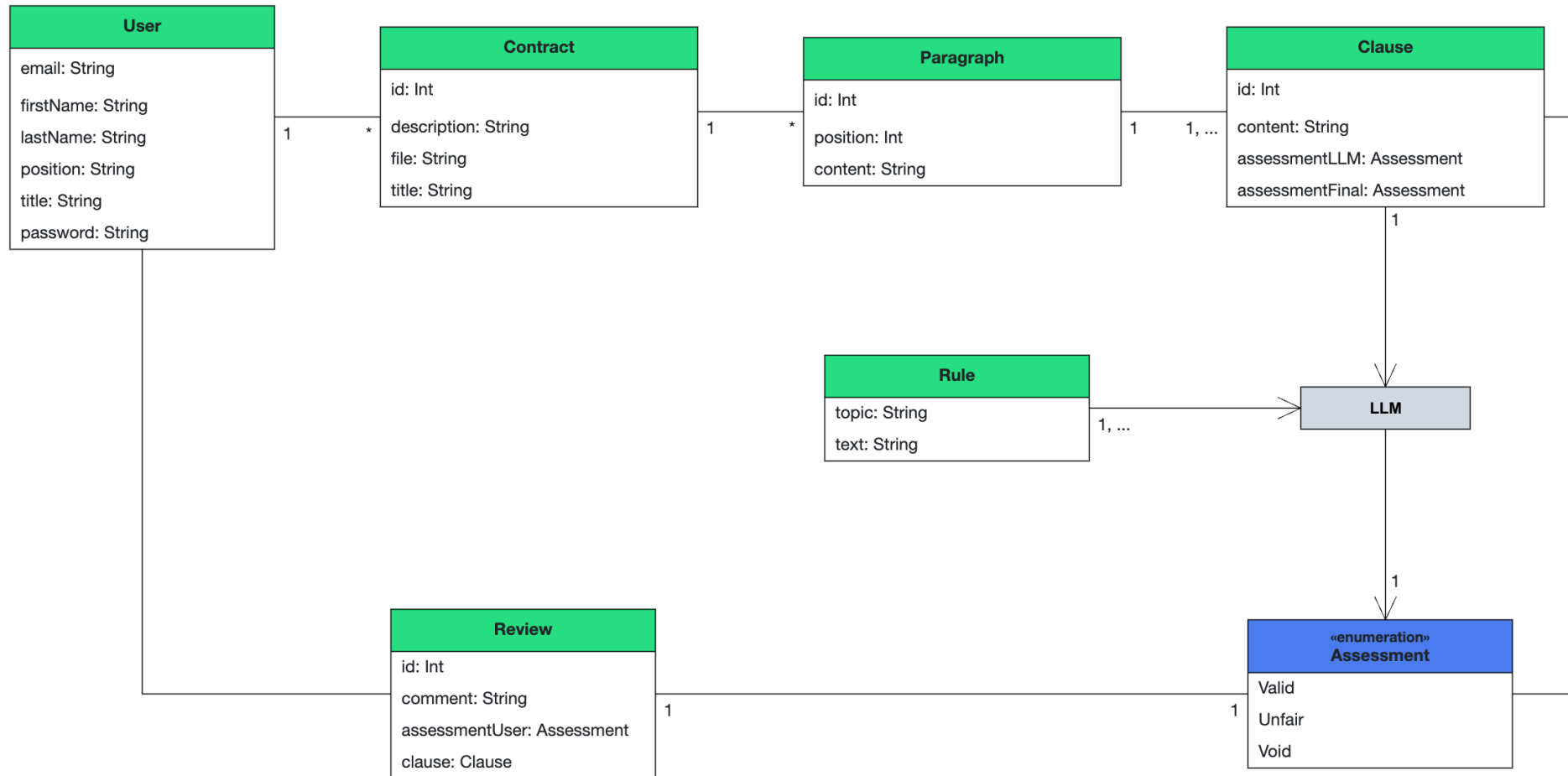


Background & Motivation

What is an Examination Guideline?



Background & Motivation



RQ1	How can changes in examination guidelines and their impact on the evaluation of contract clauses in an in-context learning AI system be tracked and assessed?
RQ2	What is needed to create a user-friendly interface which allows to monitor this evolution without technical knowledge?
RQ3	How can technical aspects and the usability of the system be evaluated?

RQ1

Slowly Changing Dimensions (SCD), Structured Output **[3,4]**

RQ2

Common Software Development Metrics

RQ3

Standard Usability Scale (SUS), Unmoderated Software Testing **[7,8]**

Rule
topic: string
text: string

Regelhistorie: Ausschlussfristen bei der Geltendmachung von Ansprüchen

Version vom 25.11.2024

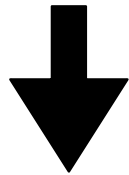
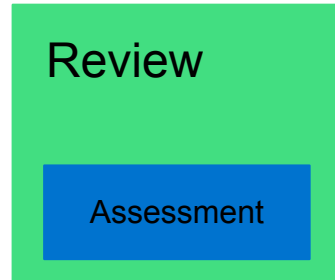
Es darf keine strengere Form als die Textform verlangt werden. Textform bedeutet, dass eine lesbare Erklärung ohne Unterschrift, z. B. per E-Mail oder Fax, abgegeben werden kann, sofern die Person des Erklärenden erkennbar ist.

Version vom 20.11.2024

Es darf keine strengere Form als die Textform verlangt werden

SCHLIESSEN

Pattern Matching



OpenAI's structured output [3]

```
class ClauseValidationExpert(BaseModel):  
    clauseld: str  
    assessment: Literal["void", "unfair", "valid"]  
    rulelds: List[str]  
    confidence: float  
    comment: str
```

«enumeration» Assessment
Valid
Unfair
Void



OpenAI's structured output [3]

```
class ClauseValidationExpert(BaseModel):  
    clauseId: str  
    assessment: Literal["void", "unfair", "valid"]  
    ruleIds: List[str]  
    confidence: float  
    comment: str
```

Review

```
assessment: Assessment  
  
comment: string  
  
clause: Clause  
  
author: Author  
  
rules: string[]  
  
isLLM: boolean  
  
confidence: number
```

2.2 Arbeitszeit

Die regelmäßige wöchentliche Arbeitszeit beträgt 40 Stunden. Die Arbeitszeiten finden Montag bis Freitag von 8.00 Uhr bis 16.30 Uhr statt, wobei eine Pause von 30 Minuten durchgeführt wird. Die Arbeitszeit kann je nach Ermessen variieren. Weitere Details werden mit dem Vorgesetzten abgestimmt.

AI Assessment



Ihre Bewertung

FERTIG

2.2 Arbeitszeit

Die regelmäßige wöchentliche Arbeitszeit beträgt 40 Stunden. Die Arbeitszeiten finden Montag bis Freitag von 8.00 Uhr bis 16.30 Uhr statt, wobei eine Pause von 30 Minuten durchgeführt wird. Die Arbeitszeit kann je nach Ermessen variieren. Weitere Details werden mit dem Vorgesetzten abgestimmt.

AI Assessment

Die Formulierung, dass die Arbeitszeit nach 'Ermessen variieren' kann, benachteiligt den Arbeitnehmer unangemessen, indem sie die notwendige Planbarkeit beschränkt.

Angewandte Regeln:

- Betriebliche Schulungen sind grundsätzlich zu entlohnen.



Ihre Bewertung

FERTIG

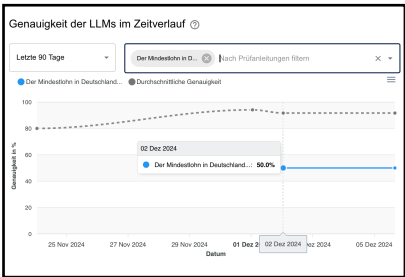
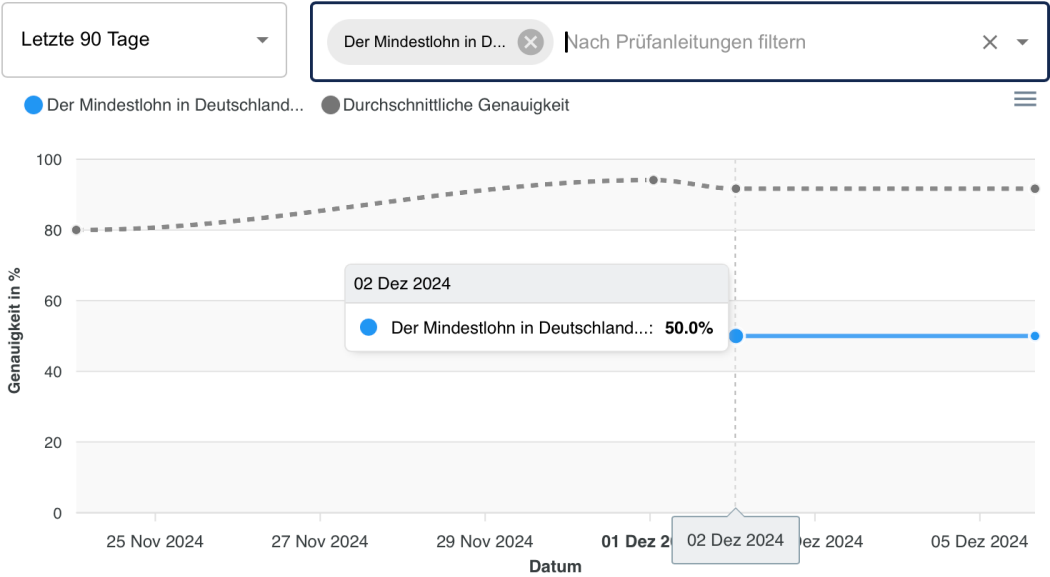
Meta Information

- Current state of the system
- Amount Contracts, Users
- Amount of Examination Guidelines
- Examination Guidelines per Topic
- Avg. Time per Contract (LLM)

AI Information

- Comparable data to prior versions of the system
- Accuracy of LLMs
- Confusion Matrix
- Confidence Score

Genauigkeit der LLMs im Zeitverlauf ?



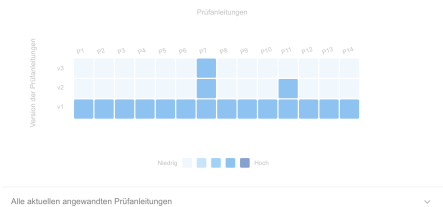
Confusion Matrix ?

Dieses Jahr

1.1.2024 - 10.12.2024
Genauigkeit: 69.0% (153/223)

LLM \ Anwalt	valid	unfair	void
valid	95.45 %	66.20 %	40.91 %
unfair	3.79 %	29.58 %	31.82 %
void	0.76 %	4.23 %	37.27 %

Confidence Score der Prüfanleitungen ?



Confusion Matrix ?

Dieses Jahr

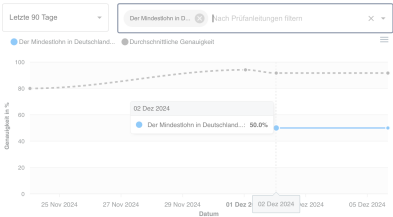
+

1.1.2024 - 10.12.2024

Genauigkeit: 68.0% (153/225)

LLM \ Anwalt	valid	unfair	void
valid	95.45 %	66.20 %	40.91 %
unfair	3.79 %	29.58 %	31.82 %
void	0.76 %	4.23 %	27.27 %

Genauigkeit der LLMs im Zeitverlauf



Confusion Matrix

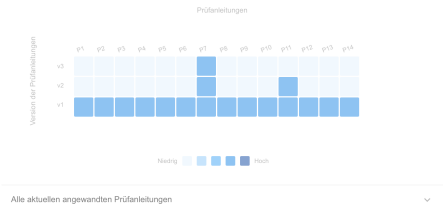
Dieses Jahr

+

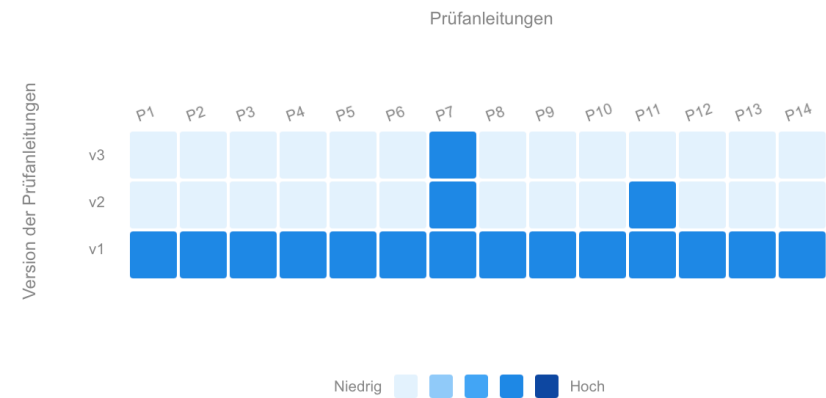
1.1.2024 - 10.12.2024
Genauigkeit: 68.0% (153/225)

LLM \ Anwalt	valid	unfair	void
valid	95.45 %	66.20 %	40.91 %
unfair	3.79 %	29.58 %	31.82 %
void	0.76 %	4.23 %	27.27 %

Confidence Score der Prüfanleitungen

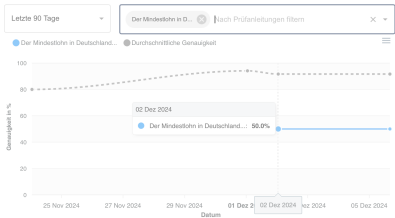


Confidence Score der Prüfanleitungen ?



Alle aktuellen angewandten Prüfanleitungen

Genauigkeit der LLMs im Zeitverlauf ?



Confusion Matrix ?

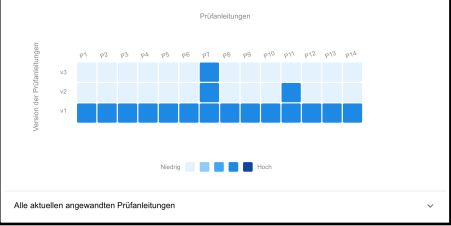
Dieses Jahr

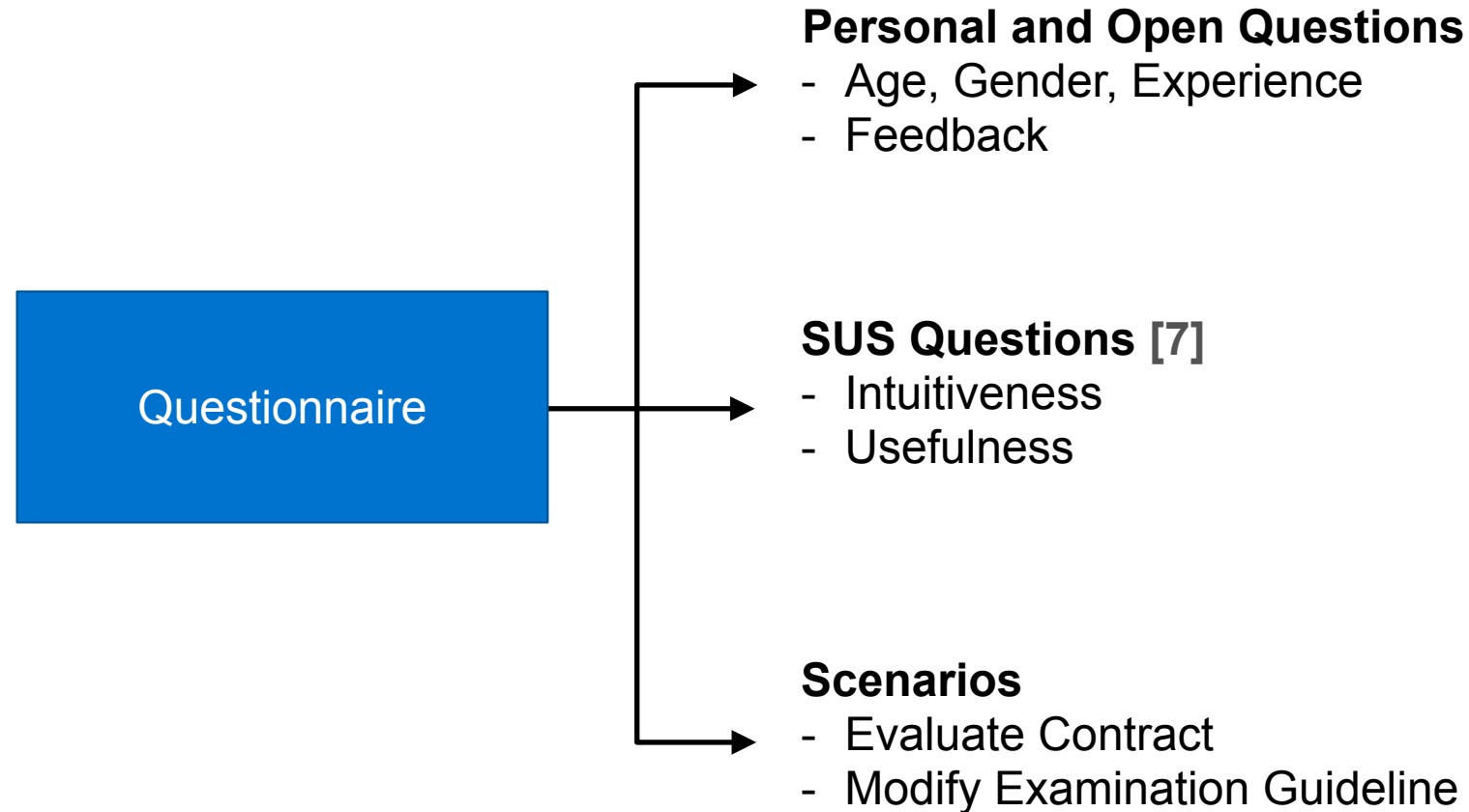
1.1.2024 - 10.12.2024

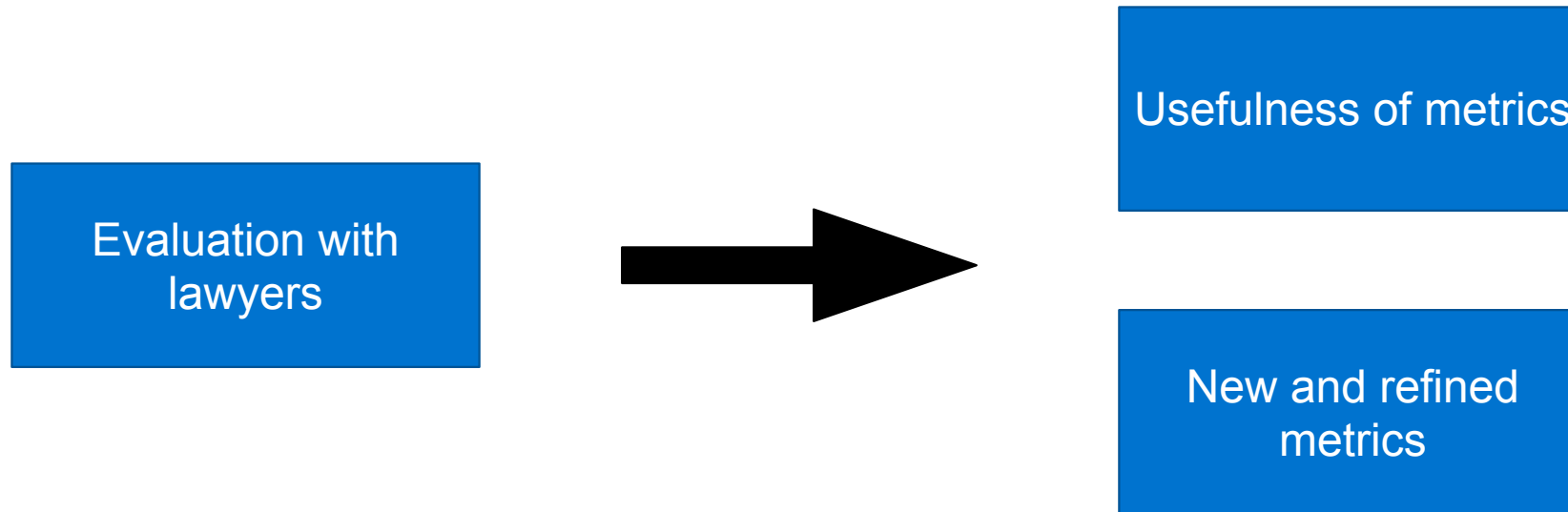
Genauigkeit: 69.0% (153/223)

LLM \ Answert	valid	unfair	void
valid	95.45 %	66.20 %	40.91 %
unfair	3.79 %	29.58 %	31.82 %
void	0.76 %	4.23 %	27.27 %

Confidence Score der Prüfanleitungen ?







No Evaluation with
Lawyers

Single Researcher
Bias

Time Constraint



Prof. Dr.

Florian Matthes

Technical University of Munich (TUM)
TUM School of CIT
Department of Computer Science (CS)
Chair of Software Engineering for Business
Information Systems (sebis)

Boltzmannstraße 3
85748 Garching bei München

+49.89.289.

17132
www.matthes.in.tum.de
matthes@in.tum.de



- [1] Gartner, Legal Technology, accessed January 2025, <https://www.gartner.com/en/newsroom/press-releases/2024-04-25-gartner-predicts-global-legal-technology-market-will-reach-50-billion-by-2027-as-a-result-of-genai>
- [2] Ver.di, accessed January 2025, <https://www.verdi.de/themen/arbeit/++co++d4ff4502-5cd5-11ec-9ee8-001a4a16012a>
- [3] OpenAI, Structured Output, accessed January 2025, <https://platform.openai.com/docs/guides/structured-outputs>
- [4] Oracle. Slowly Changing Dimensions, accessed January 2025, https://www.oracle.com/webfolder/technetwork/tutorials/obe/db/10g/r2/owb/owb10gr2_gs/owb/lesson3/slowlychangingdimensions.htm
- [5] K. Tian et al. Just Ask for Calibration: Strategies for Eliciting Calibrated Confidence Scores from Language Models Fine-Tuned with Human Feedback, <https://arxiv.org/abs/2305.14975>
- [6] K. Shen and M. Kejriwal. A Formalism and Approach for Improving Robustness of Large Language Models Using Risk-Adjusted Confidence Scores, <https://arxiv.org/abs/2310.03283>
- [7] J. Lewis. “The System Usability Scale: Past, Present, and Future”. In: International Journal of Human-Computer Interaction (Mar. 2018), pp. 1–14
- [8] P. Khayyatkhooshnevis, S. Tillberg, E. Latimer, T. Aubry, A. Fisher, and V. Mago. Comparison of Moderated and Unmoderated Remote Usability Sessions for Web-Based Simulation Software: A Randomized Controlled Trial.



Prof. Dr.

Florian Matthes

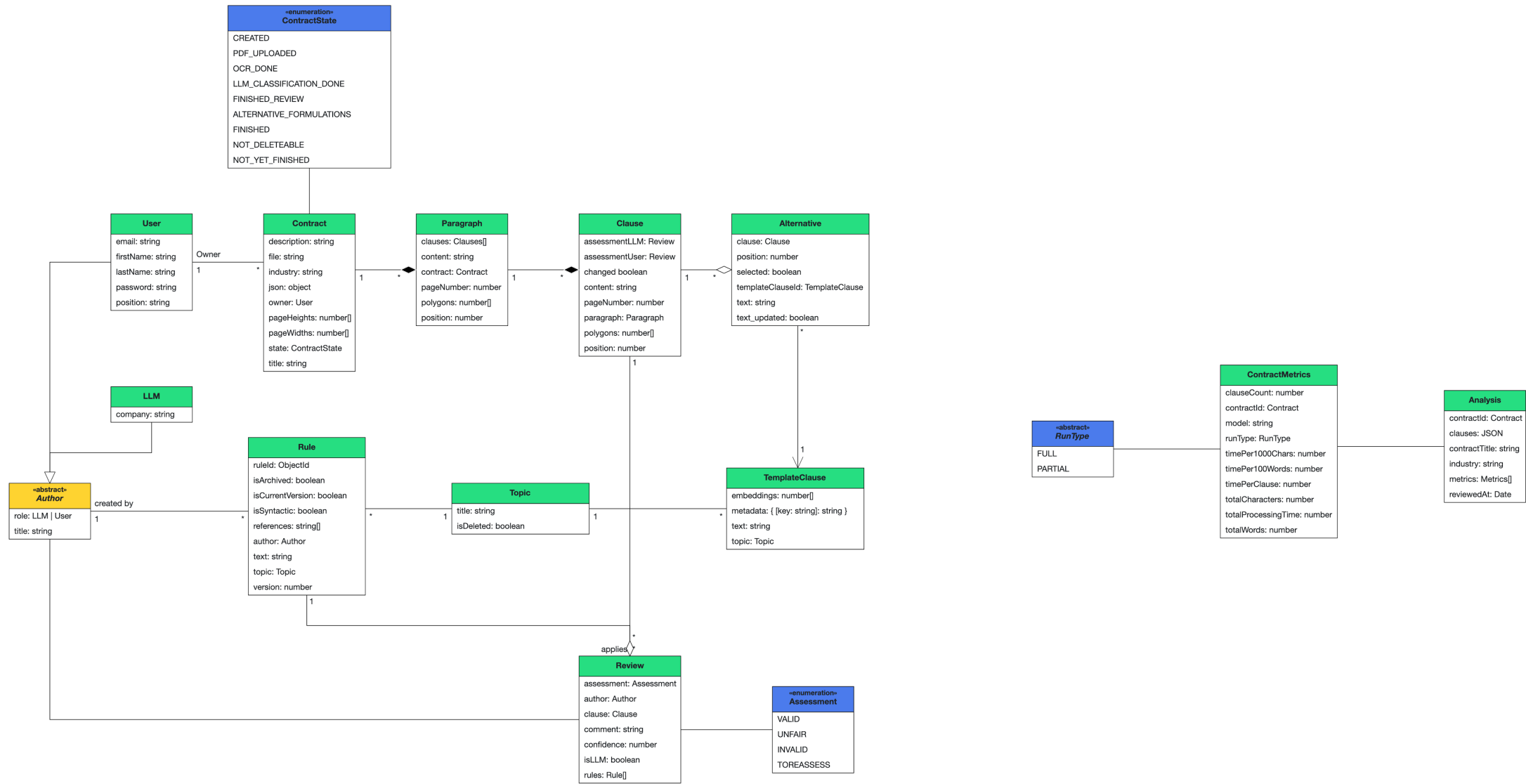
Technical University of Munich (TUM)
TUM School of CIT
Department of Computer Science (CS)
Chair of Software Engineering for Business
Information Systems (sebis)

Boltzmannstraße 3
85748 Garching bei München

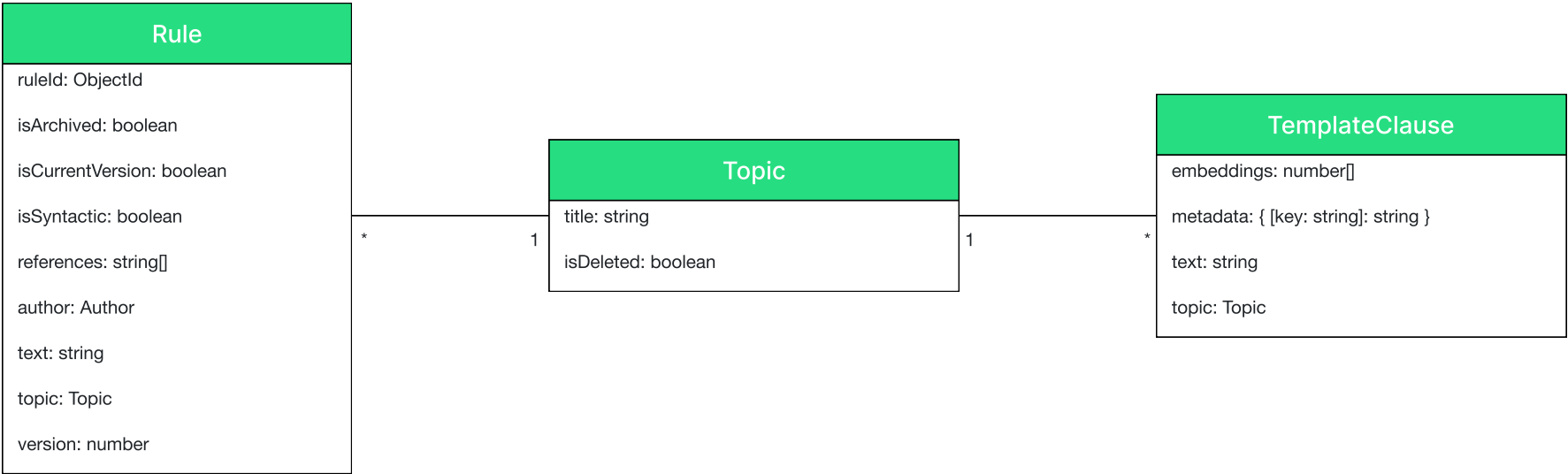
+49.89.289.

17132
www.matthes.in.tum.de
matthes@in.tum.de





2024



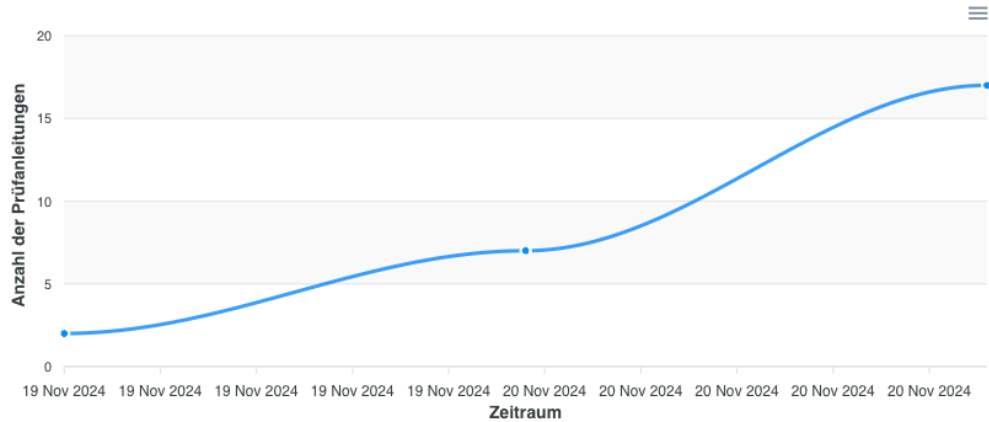
Results

System Information Metrics

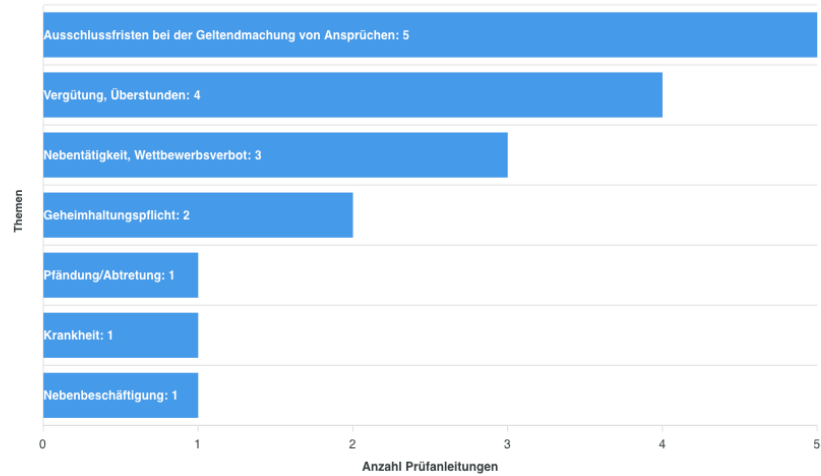


00:00

Anzahl der Prüfanleitungen im Zeitverlauf ?

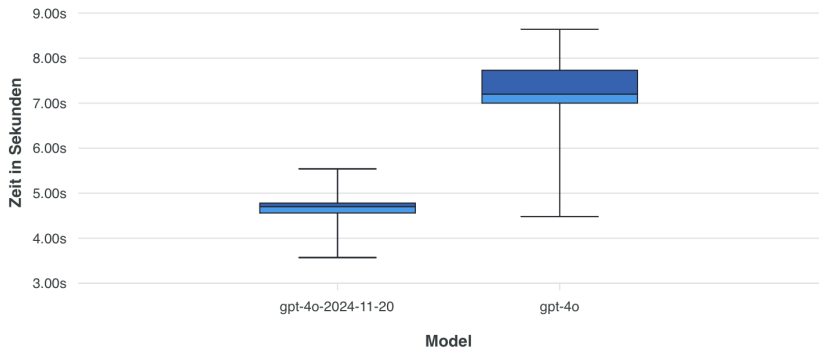


Prüfanleitungen pro Thema ?



Durchschnittliche Zeit pro Vertrag (LLM only) ?

Verarbeitungszeit pro 100 Wörter nach Modell



* Nur vollständige Durchläufe berücksichtigt

Vertrag 5

Erstellt von: John Graf von Hardenberg Erstellt am: 24.11.24

1 Fairness-Bewertung

2 Alternative Formulierungen

BACK

NEXT

IN ORDNUNG 21

UNFAIR 11

UNZULÄSSIG 6

NEUBEWERTUNG 1

ZURÜCKSETZEN

KI-ANALYSE

ANWALTlich ABGENOMMEN

Einleitung

Vertrag neu analysieren

Alle Klassifizierungen zurücksetzen

Anwalt Reviews zurücksetzen

Nicht anwaltlich abgenommen

der Anstellungsvertrag geschlossen:

§ 19 Schlussbestimmungen

(1) Dieser Vertrag ist inhaltlich vollständig. Alle Änderungen und Ergänzungen dieses Vertrages - ausgenommen Abreden, die nachweislich individuell ausgehandelt worden sind - bedürfen zu ihrer Wirksamkeit der Schriftform oder der schriftlichen Bestätigung durch den Arbeitgeber. Dies gilt auch für die Aufhebung dieser Schriftformklausel.

AI Assessment

Die Klausel ist unfair, da sie eine strengere Form als die gesetzlich vorgegebene Textform vorschreibt und gleichzeitig die Abbedingung der Schriftform ebenfalls schriftlich fordert, was den Vertragspartner unangemessen benachteiligen könnte.

Angewandte Regeln:

- Es darf keine strengere Form als die Textform verlangt werden

Ihre Bewertung

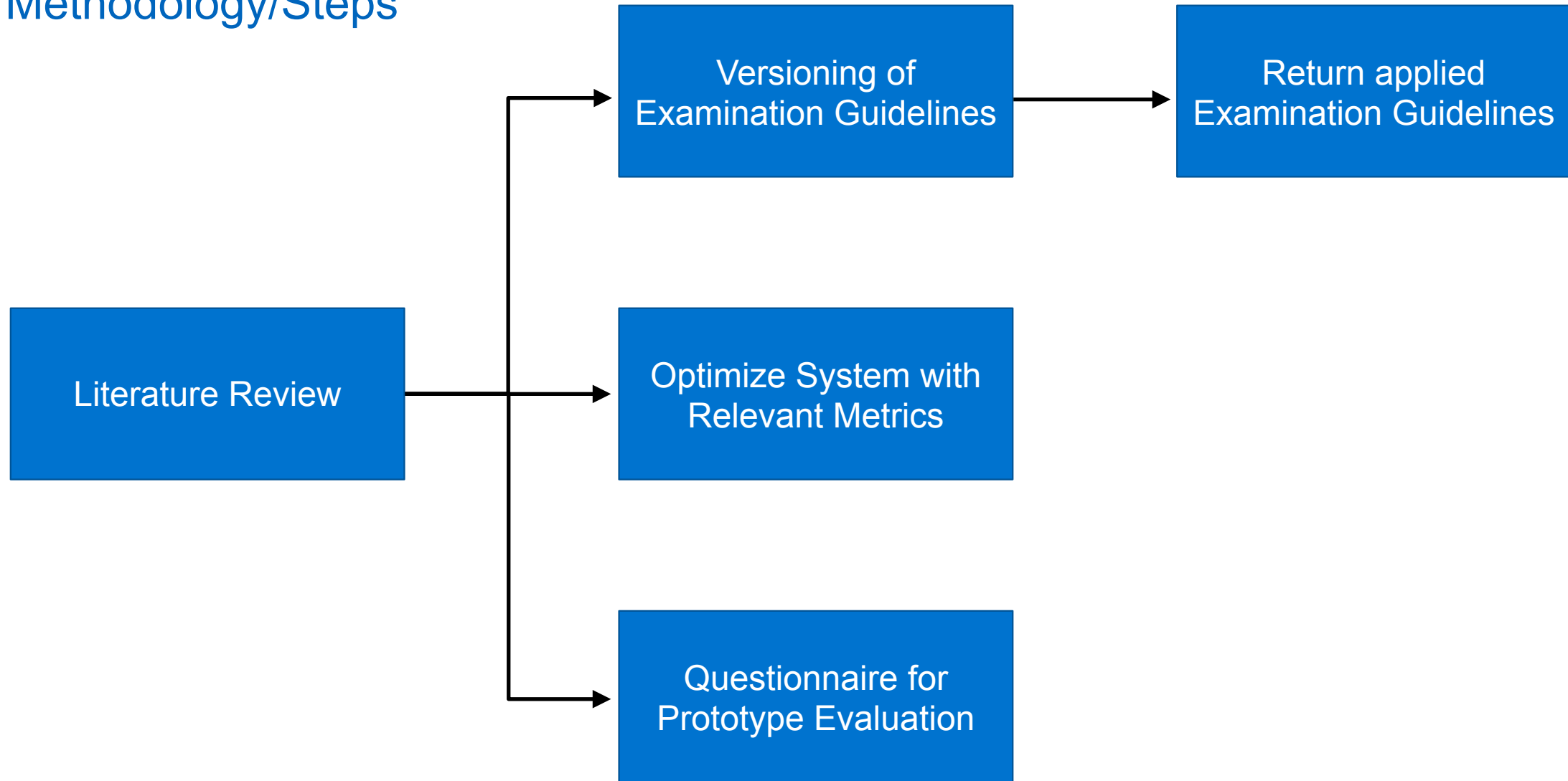
FERTIG

250113 John Graf von Hardenberg Controlling the Evolution of a Legal Knowledge Base for an In-Context Learning AI System - Design and Implementation

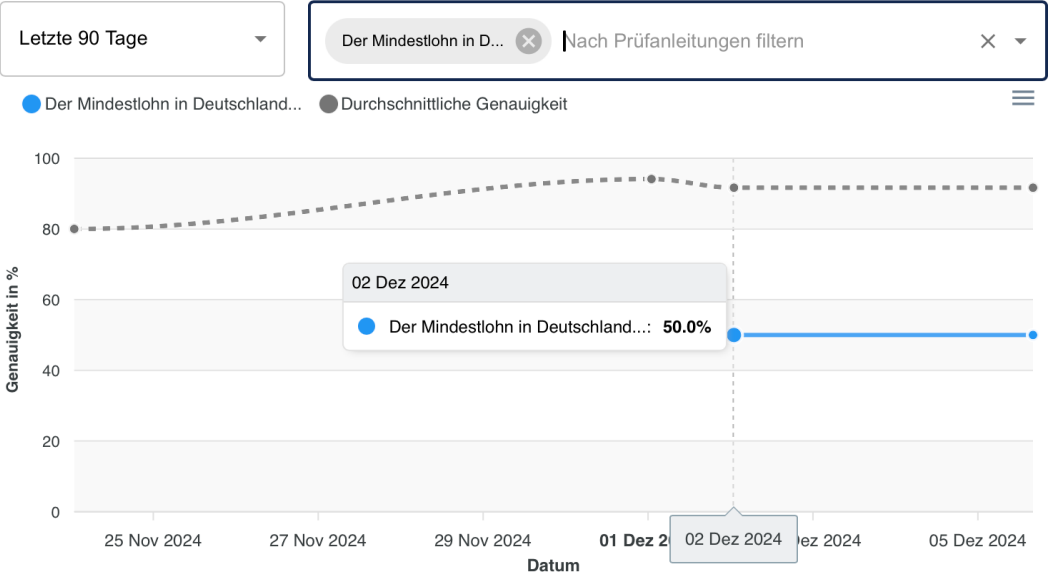
© sebis

27

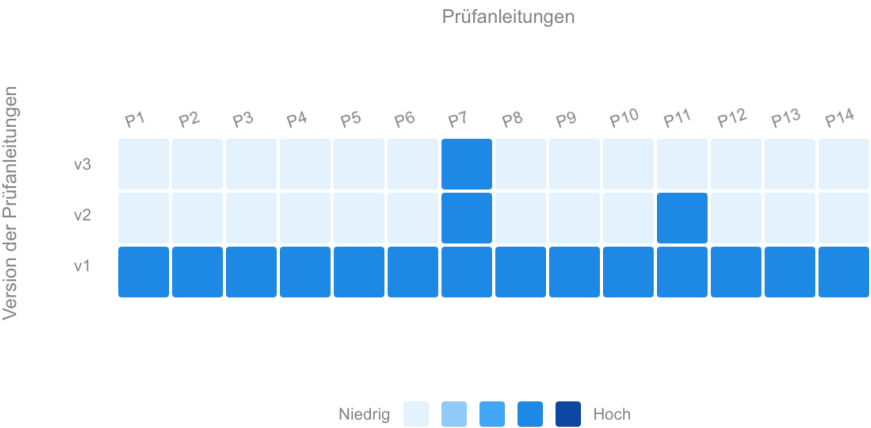
Methodology/Steps



Genauigkeit der LLMs im Zeitverlauf



Confidence Score der Prüfanleitungen



Alle aktuellen angewandten Prüfanleitungen

Confusion Matrix

Dieses Jahr

1.1.2024 - 10.12.2024

Genauigkeit: 68.0% (153/225)

LLM \ Anwalt	valid	unfair	void
valid	95.45 %	66.20 %	40.91 %
unfair	3.79 %	29.58 %	31.82 %
void	0.76 %	4.23 %	27.27 %

[5,6]