

Ein Gesetz, viele Sprachen: NLP für gerichtliche Unterstützung in der Schweiz

21. Juni 2024

Joel Niklaus

Joel Niklaus

Education		
2022 – 2023	Research Visit	@ Stanford University
2020 – 2023	PhD in NLP	@ University of Bern
2013 – 2019	BSc and MSc in Computer Science and Sports Science	@ University of Bern & University of Exeter & Nanyang Technological University
Teaching		
2021 – 2024	Lecturer	@ Bern University of Applied Sciences
2015 – 2024	Lecturer & Teaching Assistant	@ Institute for Computer Science, University of Bern
2016 – 2020	High School Teacher	@ Kantonsschule Rychenberg Winterthur & Freigymn Bern
Industry		
2023	PhD Resident	@ (Google) X
2022	Applied Research Intern	@ Thomson Reuters Labs
2020	Data Scientist	@ ARTORG Center & University of Fribourg
2017 – 2018	Web Developer	@ fastforward websolutions GmbH
Since 2012	Independent Software Developer & AI Consultant	@ JoelNiklaus.ch & niklaus.ai



www.niklaus.ai

joel@niklaus.ai



[joelniklaus](https://twitter.com/joelniklaus)



[joelniklaus](https://www.linkedin.com/in/joelniklaus)

Ko-Autoren



Ilias Chalkidis



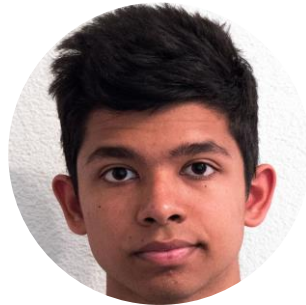
Srinanda Bose



Daniel E. Ho



Veton Matoshi



Vishvaksenan Rasiah



Ronja Stern



Matthias Stürmer

Die Justiz ist heillos überfordert

Überall auf der Welt!

USA us

Backlogged Courts and Years of Delays Await Many Migrants

A severe shortage of immigration judges has led to long delays for asylum cases. The backlog is expected to grow now that pandemic-era border restrictions have been lifted.

Vereinigtes Königreich GB

Britain | Family division

Britain's family-court system is overwhelmed

The justice system is leaving children in limbo

Frankreich FR

À 50 ans, le tribunal de Bobigny lutte pour garder la tête hors de l'eau

Bobigny. Bobigny - Devenu malgré lui l'emblème d'un système judiciaire débordé et manquant de moyens, le tribunal de Bobigny en Seine-Saint-Denis, qui marque mardi ses 50 ans, espère que l'avenir lui apportera un peu d'air dans l'un des départements les plus pauvres et criminogènes de France.

Deutschland DE

Ein Berliner Oberstaatsanwalt schlägt Alarm: «Einbrecher werden hier in 97 Prozent der Fälle nicht verurteilt»

Ralph Knispel ist seit fast drei Jahrzehnten Staatsanwalt und sorgt sich um Deutschlands Justizsystem. Ein Gespräch über realitätsferne Politik, eingeschüchterte Zeugen und einen bundesweit bekannten Clan-Boss.

Schweiz CH

Überlastete Schweizer Justiz

110'000 offene Verfahren sind pendent – jetzt setzen Regierungsräte Taskforce ein

Die Justiz stehe kurz vor dem Kollaps, sagten Strafverfolger vor einem halben Jahr dieser Redaktion. Nun werden die kantonalen Justizdirektoren aktiv. Die Situation sei unhaltbar.

Indien IN

‘A Lifelong Nightmare’: Seeking Justice in India’s Overwhelmed Courts

With 50 million criminal and civil cases pending, it would take 300 years to clear the country’s judicial backlog.

Brasilien BR

Brazil's Overwhelmed Judiciary, Desperate for Help, Turns to Artificial Intelligence

Dozens of programs have been rolled out across Brazil's courts as innovators attempt to advance justice in one of the world's most-clogged legal systems.

Können Sprachmodelle wie ChatGPT unterstützen?

Sprachmodelle (LLMs) können vieles schon unglaublich gut!



Grammar correction

Convert ungrammatical statements into standard English.



Parse unstructured data

Create tables from unstructured text.



Calculate time complexity

Find the time complexity of a function.



Keywords

Extract keywords from a block of text.



Python bug fixer

Find and fix bugs in source code.



Summarize for a 2nd grader

Simplify text to a level appropriate for a second-grade student.



Emoji Translation

Translate regular text into emoji text.



Explain code

Explain a complicated piece of code.



Product name generator

Generate product names from a description and seed words.



Spreadsheet creator

Create spreadsheets of various kinds of data.

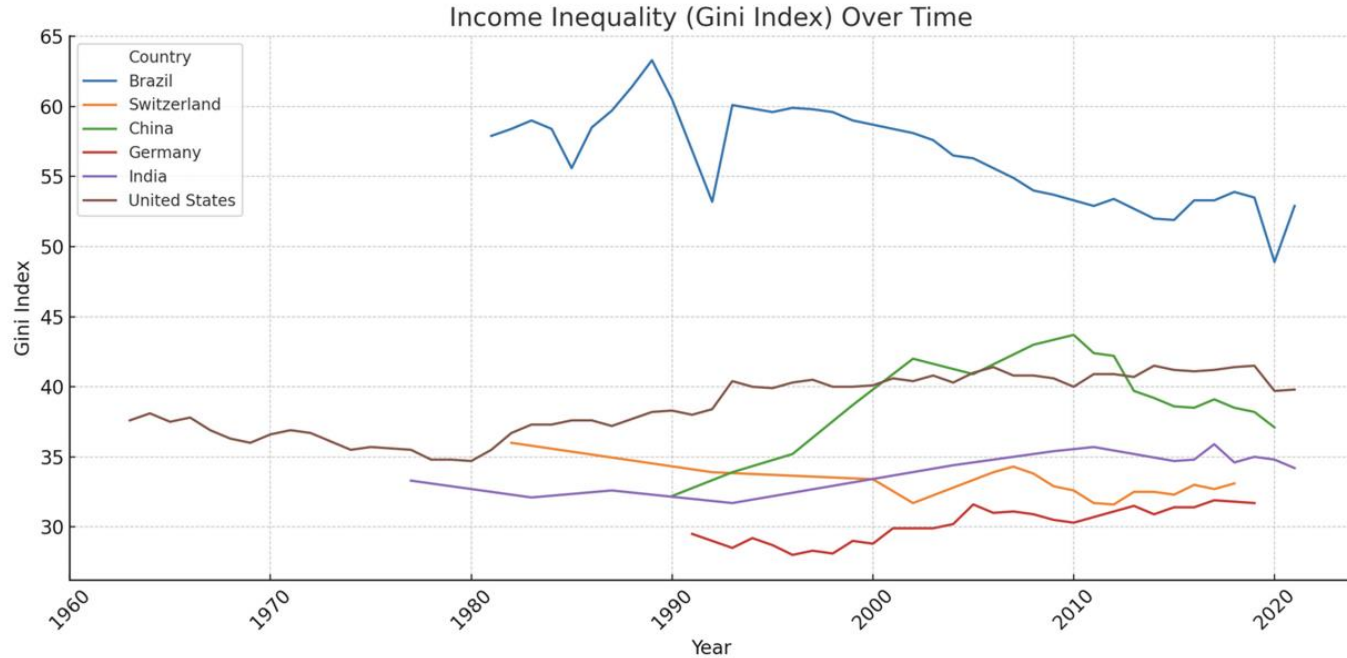
J You



API_SI.POV.GINI_DS2_en_csv_v2_6299...

Spreadsheet

Take this data on income inequality from the world bank and generate a plot comparing the following countries: Switzerland, Germany, USA, China, India, and Brazil. Make the plot readable by only displaying the year on the x-axis every 10 years. Make sure that the colors are following a clear color palette making it easy to distinguish the different lines.



The plot visualizes the Gini Index over time for Switzerland, Germany, the United States, China, India, and Brazil. The Gini Index, a measure of income inequality, is presented on the y-axis, and the years are displayed on the x-axis at 10-year intervals. Each country's trend is represented by a distinct color for clarity. [↗]

Aber sie machen immer noch viele Fehler.
Vor allem im juristischen Bereich!

These lawyers used ChatGPT to save time. They got fired and fined.

Artificial intelligence is changing how law is practiced, but not always for the better

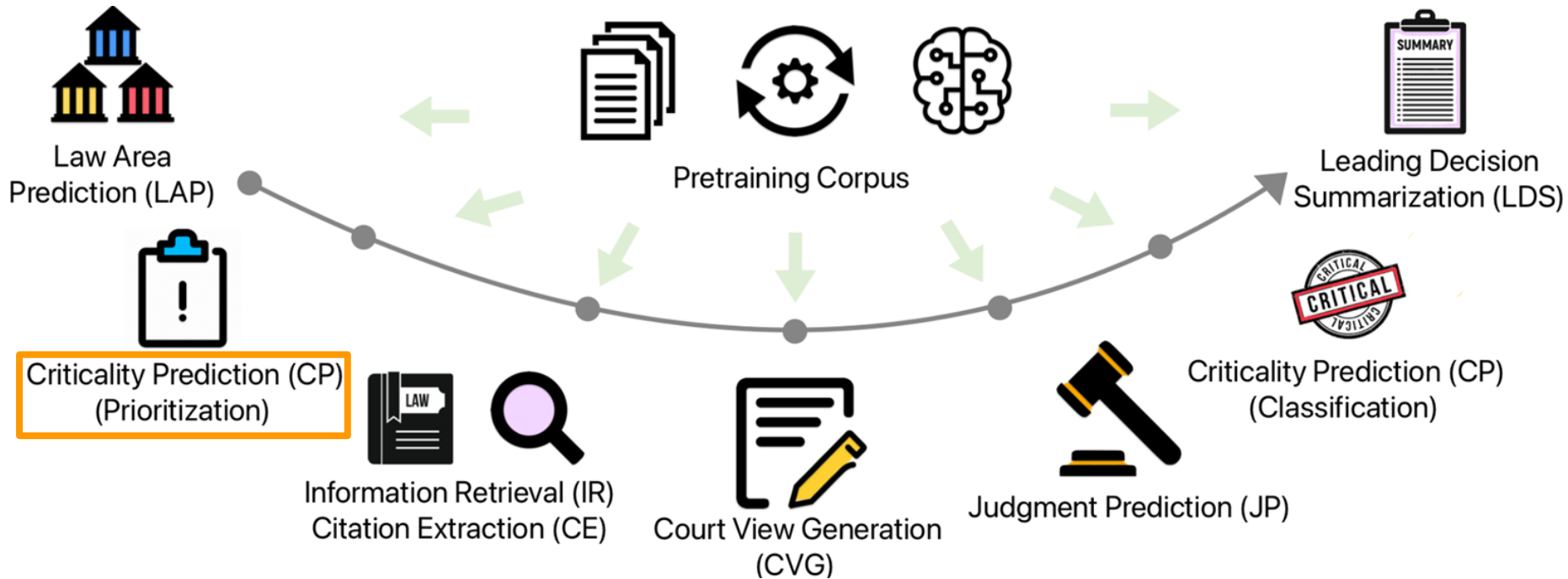
FORBES > BUSINESS

BREAKING

Judge Fines Two Lawyers For Using Fake Cases From ChatGPT

Wo und wie kann KI
unterstützen?

Task Übersicht



Was ist “Criticality Prediction (Prioritization)”?

= Priorisierung

Ziel: Wichtige Fälle priorisieren

Input: Sachverhalt eines Urteils

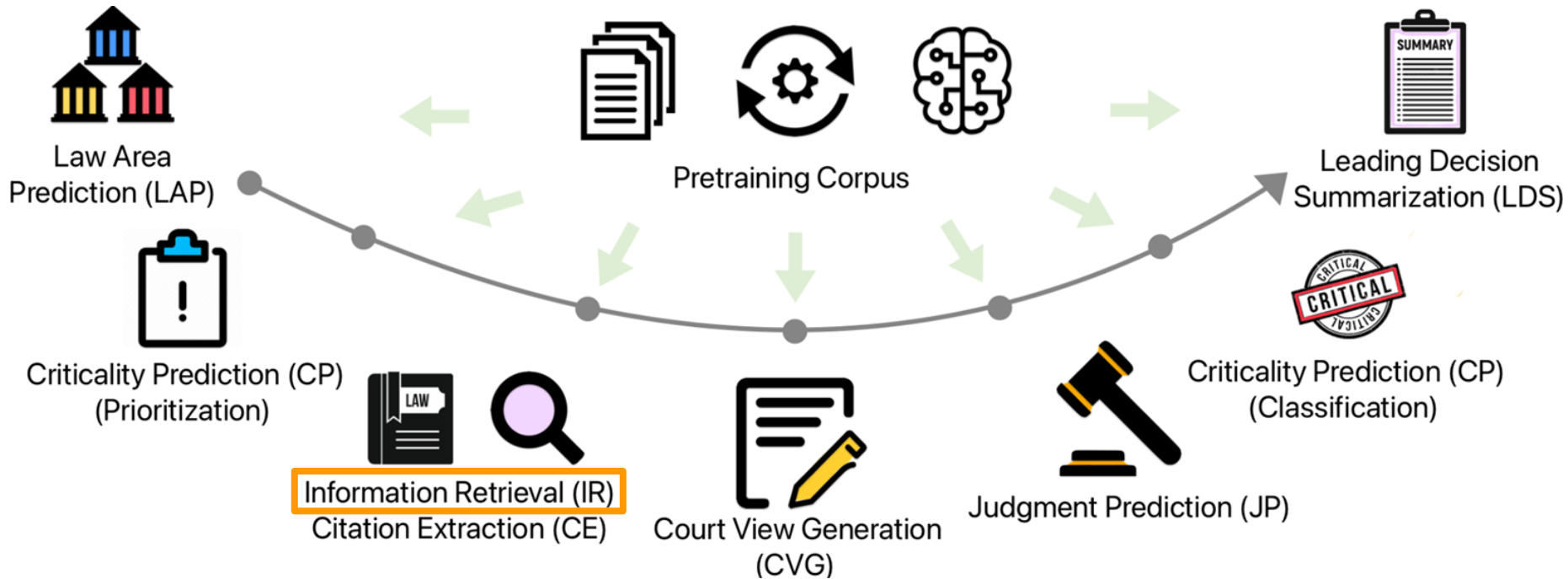
Output: Kritikalität / Wichtigkeit



Justice delayed is justice denied.

William E. Gladstone

Task Übersicht



Was ist “Information Retrieval”?

= Suche

Ziel: Rechtsspezifische Suchen unterstützen

Input: Sachverhalt eines Urteils

Output: Relevante Urteile und Gesetzesartikel

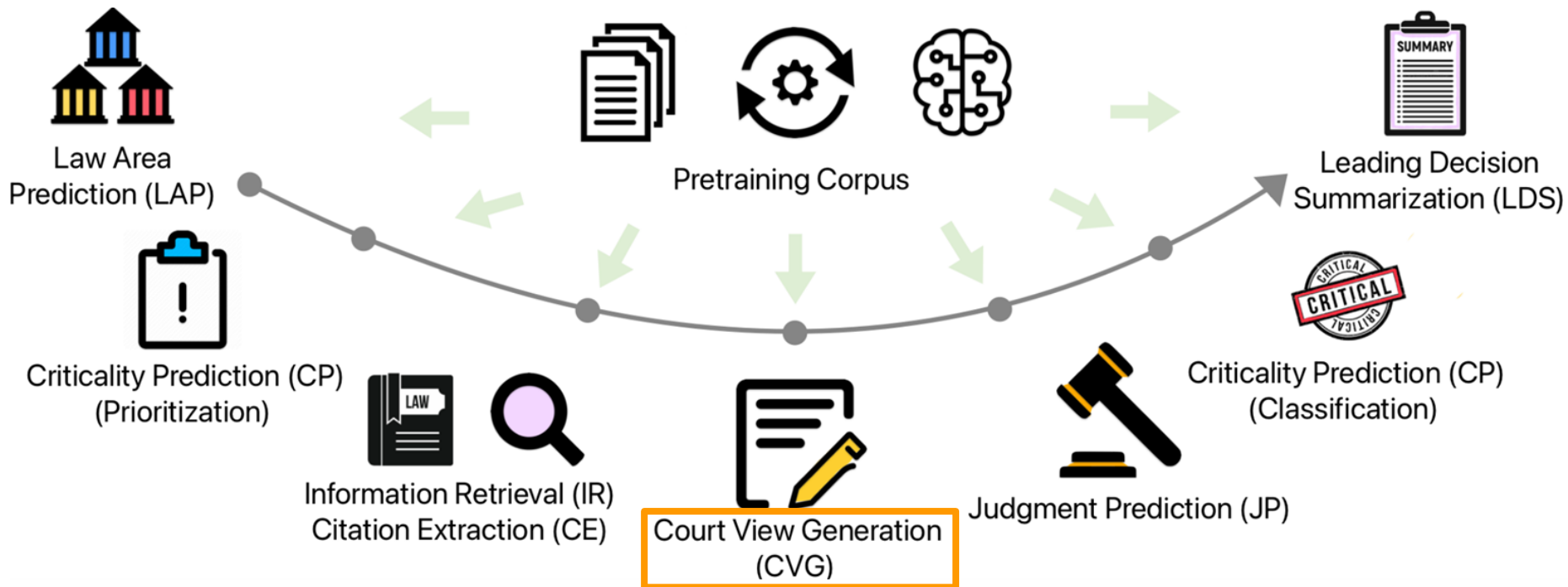


THOMSON REUTERS

WESTLAW™



Task Übersicht



Was ist “Court View Generation”?

= Entwurf

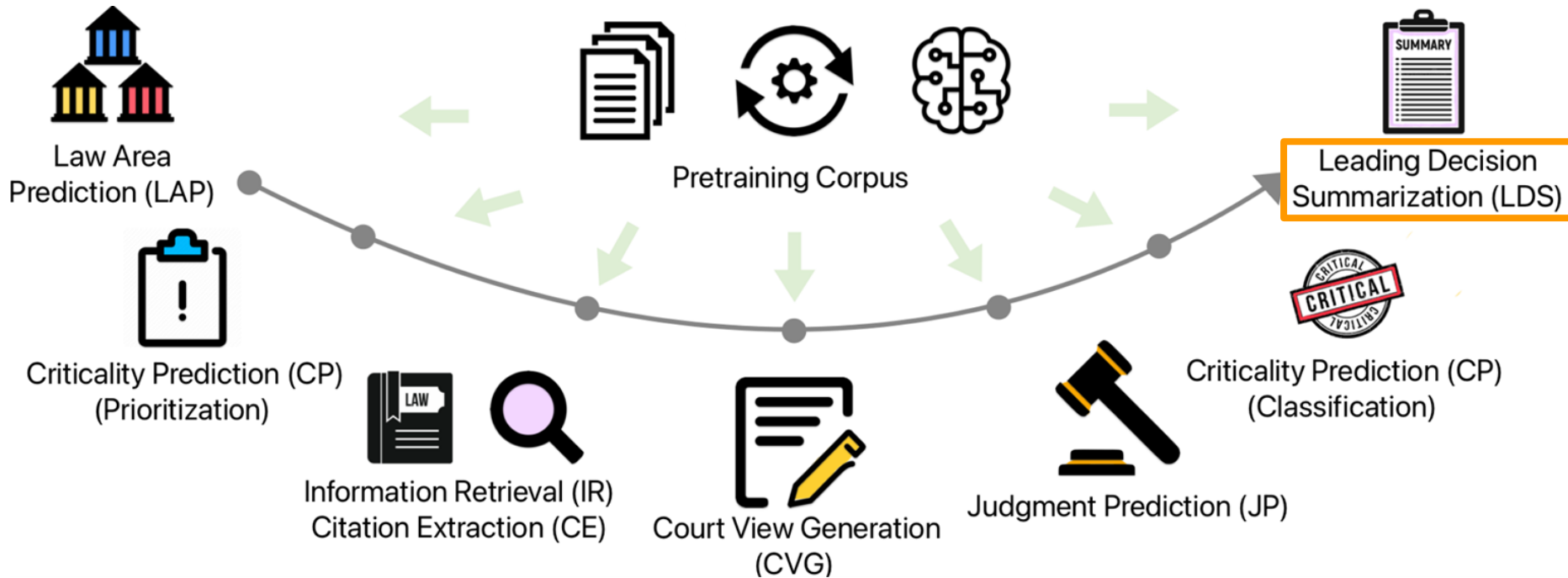
Ziel: Gerichtsschreiber im Verfassen von Urteilen unterstützen

Input: Sachverhalt, relevante Urteile und Gesetzesartikel, Urteil der Vorinstanz

Output: Erwägungen



Task Übersicht



Was ist “Leading Decision Summarization”?

= Zusammenfassung

Ziel: Rechtssprechung zugänglicher machen

Input: Gesamter Text eines Leitentscheids

Output: Zusammenfassung (**wichtige Zitierungen**, **Stichwörter** aus dem

Thesaurus, **Zusammenfassung im**

Fleishtext)

Regeste

Art. 19 Ziff. 2 lit. a BetmG; Handel mit Ecstasy; mengenmässig schwerer Fall.

Ecstasy ist keine harmlose Droge. Nach dem derzeitigen Wissensstand ist Ecstasy aber nicht geeignet, die körperliche oder seelische Gesundheit in eine naheliegende und ernstliche Gefahr zu bringen. Die Annahme eines **mengenmässig schweren Falles** scheidet deshalb aus. Vorbehalt der Änderung der Rechtsprechung, falls wesentliche neue Erkenntnisse zu den Gefahren von Ecstasy gewonnen werden sollten (E. 3).

Fazit

Es gibt sehr viele Möglichkeiten für Automatisierung in der Justiz

- Priorisierung
- Suche
- Entwürfe
- Zusammenfassung

Aktuelle Modelle erzielen schon gute Resultate

Gerichte sollten Unterstützung durch KI prüfen

⇒ Das Bundesgericht ist sehr fortgeschritten und testet moderne Methoden

Fragen? 

Backup Slides

Database Creation

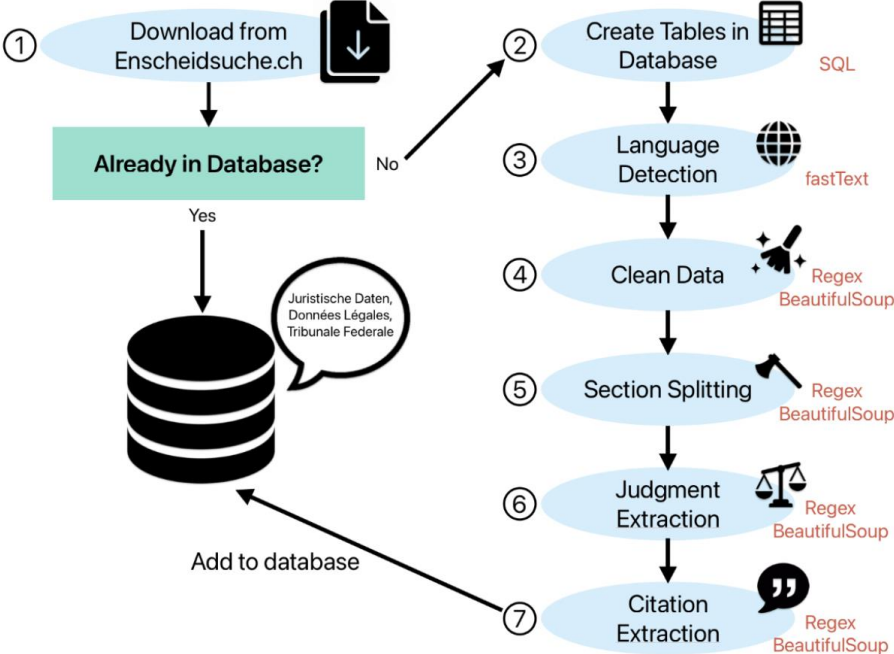


Figure 2: Database Creation Pipeline

Multilingual Doc2Doc Information Retrieval

Table 7: Results on Information Retrieval with best scores per section in **bold**. Abbreviations: **distil**use_{Base}-multilingual-cased-v1, joelito/**swiss**-legal-roberta_{Base}

Model	RCap@ 1 / 10 / 100 \uparrow	NDCG@ 1 / 10 / 100 \uparrow
BM25 (fr lang analyzer)	11.37 / 7.74 / 16.54	11.37 / 8.34 / 11.51
SBERT distil	0.90 / 0.75 / 2.64	2.06 / 1.70 / 3.31
SBERT distil + pos	4.40 / 3.92 / 12.64	10.11 / 8.76 / 16.16
SBERT distil + pos + h-neg	3.97 / 4.46 / 13.36	9.12 / 9.21 / 16.87
SBERT swiss + pos	3.97 / 3.47 / 12.28	9.12 / 7.76 / 15.16
SBERT distil eval on de queries	4.22 / 4.49 / 15.21	8.21 / 8.15 / 15.86
SBERT distil eval on fr queries	1.88 / 2.20 / 9.19	5.77 / 6.22 / 13.94
SBERT distil eval on it queries	0.22 / 0.24 / 0.79	5.43 / 5.74 / 11.44

Court View Generation

Table 5: Results of Court View Generation task. 'In Len' denotes input length in tokens. **Bold**: best within model; underlined: best overall.

Model	In Len \uparrow	BERT	BLEU	MET	R1 / R2 / RL
mT5 _{Large}	2048	<u>75.74</u>	<u>66.92</u>	<u>34.44</u>	<u>34.91/15.58/33.53</u>
mT5 _{Large}	1024	75.56	66.68	34.02	34.26/14.72/32.87
mT5 _{Large}	512	75.27	66.12	33.48	33.61/14.26/32.21
mT5 _{Base}	2048	75.01	65.48	32.89	33.23/13.57/31.89
mT5 _{Base}	1024	75.15	65.73	33.15	33.49/13.96/32.18
mT5 _{Base}	512	74.89	65.55	32.66	32.66/13.16/31.35
mT5 _{Small}	2048	74.13	63.97	30.96	31.29/11.01/29.90
mT5 _{Small}	1024	74.00	63.70	30.68	31.05/10.77/29.64
mT5 _{Small}	512	73.92	63.83	30.57	30.58/10.35/29.20

Text Classification

Table 4: Results on the Text Classification datasets. Macro F1 score is reported. The 'F' or 'C' following the dash represents input based on 'Facts' or 'Considerations' respectively. 'CPB' and 'CPC' refer to the CP task using BGE and Citation labels, respectively, while 'SLAP' denotes Sub Law Area Prediction. Note: Seeds that yielded very high evaluation losses were considered as failed and therefore excluded from the analysis.

Model	CPB-F	CPB-C	CPC-F	CPC-C	SLAP-F	SLAP-C	JP-F	JP-C	Agg.
MiniLM	54.7	65.8	9.8	20.8	59.7	61.1	58.1	78.5	32.4
DistilBERT	56.2	65.4	19.6	22.1	63.7	65.9	59.9	75.5	42.1
mDeBERTa-v3	55.1	69.8	21.0	17.5	63.8	59.3	60.6	77.9	40.2
XLM-R _{Base}	57.2	65.9	21.3	23.7	67.2	73.4	60.9	79.7	44.6
XLM-R _{Large}	56.4	67.9	24.4	<u>28.5</u>	65.1	78.9	60.8	80.9	<u>48.4</u>
X-MOD _{Base}	56.6	67.8	20.0	20.6	63.9	64.4	60.5	79.1	41.9
SwissBERT (xlm-vocab)	56.2	67.3	<u>25.7</u>	23.0	62.2	75.1	61.4	79.4	44.6
Legal-ch-R _{Base}	57.6	<u>72.8</u>	23.1	22.5	<u>81.6</u>	83.0	64.0	86.4	46.9
Legal-ch-R _{Large}	57.4	70.8	21.3	23.3	80.4	<u>84.9</u>	62.8	<u>87.0</u>	46.2
Legal-ch-LF _{Base}	<u>58.1</u>	70.8	21.4	17.4	80.2	83.5	<u>65.4</u>	86.4	42.8

Leading Decision Summarization

Table 6: Results of Leading Decision Summarization task. 'In Len' denotes input length in tokens. **Bold**: best within model; underlined: best overall.

Model	In Len \uparrow	BERT \uparrow	BLEU \uparrow	MET \uparrow	R1 / R2 / RL \uparrow
mT5 _{Large}	4096	OOM	OOM	OOM	OOM
mT5 _{Large}	2048	73.10	27.21	21.88	31.47 / 12.22 / 29.94
mT5 _{Large}	512	70.67	26.89	18.31	24.76 / 6.15 / 23.48
mT5 _{Base}	4096	73.33	30.81	23.50	32.43 / 12.78 / 30.87
mT5 _{Base}	2048	72.45	30.13	21.94	30.09 / 10.79 / 28.71
mT5 _{Base}	512	70.60	27.10	18.31	24.72 / 6.15 / 23.55
mT5 _{Small}	4096	72.04	28.68	21.29	29.61 / 10.31 / 28.12
mT5 _{Small}	2048	71.38	24.64	19.28	27.88 / 9.19 / 26.54
mT5 _{Small}	512	69.66	20.73	15.95	22.91 / 5.36 / 21.85

Best model outputs: mT5 Large 2048

Target

considerations: Erwägungen: 1. 1.1. Die Beschwerde in öffentlich-rechtlichen Angelegenheiten kann wegen Rechtsverletzungen gemäss Art. 95 und 96 BGG erhoben werden. Das Bundesgericht wendet das Recht von Amtes wegen an (Art. 106 Abs. 1 BGG). [...]

BERTScore: 73.3

METEOR: 28.4

BLEU: 64.0

Rouge 1/2/L: 30.7/11.1/30.7

Predicted

considerations: Erwägungen: 1. 1.1 Der angefochtene Entscheid betrifft Familienzulagen. Er stellt eine Streitigkeit, auf die das Bundesgericht verzichtet hat (Art. 82 lit. a BGG; **BGE 132 V 393 E. 1.2 S. 395**). Da die Beschwerde in öffentlich-rechtlichen Angelegenheiten angefochten werden kann, richtet sich das Verfahren noch nach dem neuen Bundesgesetz vom 16. Dezember 2005 über die Alters- und Hinterlassenenversicherung (AHVG; **SR 831.20**). [...]

[Federal law about invalidity insurance](#)

Best model outputs: mT5 Base 4096

Target

regeste: Regeste Art. 30 BV und 6 Ziff. 1 EMRK;
Ablehnung einer Kindesschutzbehörde;
"Freundschaft" auf Facebook mit einer
Prozesspartei. "Freundschaft" auf Facebook als
Ablehnungsgrund (E. 4).

BERTScore: 74.6

METEOR: 10.5

BLEU: 11.0

Rouge 1/2/L: 24.7/7.1/24.7

Predicted

regeste: Regeste Art. 29 Abs. 1 BV und Art. 314
Abs. 1 ZGB ; Art. 47 Abs. 1 lit. f ZPO ; Streit
vom Entscheid, mit dem ein Direktor einer
Rechtspflegebehörde zu einem Kinde zugestellt
wird. Ein Beziehung oder Einbeziehung
zwischen einem Richter und eine Partei bildet
ein solches Vertrauensverhältnis jedoch nur
gemäss analogischen Rechtshilfegrund (E. 4).
Bei der Berufungsfreiheit können Streitigkeiten,
im konkreten Fall gegeben sind, nach Art. 47
Abs. 1 lit. f ZPO abgewiesen werden (E. 4.3 und
Art. 6 Abs. 1 EMRK), auch wenn jemand auf
dem öffentlichen Internet seine "Bekanntschaft"
gemacht hat (E. 4.4).