



Universität
Basel

Juristische
Fakultät



Rechtsfragen beim Einsatz von KI in der Strafrechtspflege

104ème Congrès de la SSDP

Sabine Gless

Montreux, 11. Juni 2026

Smoke on the water?

Eine der gängigsten Metaphern für KI ist die Vorstellung von ihr als „Black Box“ – etwa in Form von LLMs als undurchsichtige und unergründliche Sprachmodelle, deren Methoden so enorm komplex sind, dass kein Mensch und kein Team von Menschen jemals alles entschlüsseln könnte, was im Inneren des Modells vor sich geht – sind nicht auch Menschen „black boxes“?



Präsentation

-
- 1 KI in der Strafrechtspflege –
Möglich? Zulässig? Wünschenswert?
 - 2 Strafbefehle auf Knopfdruck – Geht das?
 - 3 Rechtliche Grenzen des Einsatzes von KI in der Strafrechtspflege
(am Beispiel der Automatisierung von Strafbefehlen)
 - 4 Wollen wir maschinelle statt menschlicher Entscheide?
 - 5 Diskussion
-

(ad 1) KI in der Strafrechtspflege

Juristinnen & Juristen, Gutachter & Gutachterinnen erledigen Aufgaben, für die “Roboter” besser geeignet sein könnten, etwa

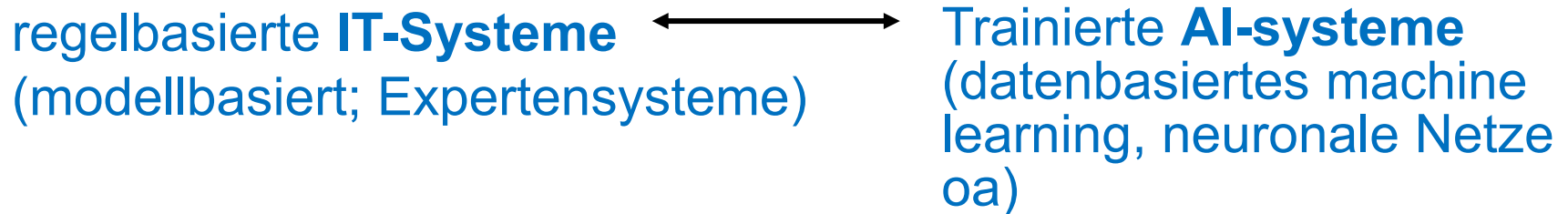
- **engmaschig vorgegebene Sachverhalts- und Rechtsprüfung, bei denen Rechtsfolge praktisch vorbestimmt ist;**
z.B. Art. 90 und 91 SVG,
- **Abgleich riesiger Datenmengen;**
(Durchsuchung von smart phones [Encro-chat]; Vergleich von Bildern [battlefield-evidence]).



Sollte man hier besser KI einsetzen?

(ad 1) KI in der Strafrechtspflege

Wo arbeiten wir schon «automatisiert»?



Welches System konsultieren Sie vor einem Entscheid?

«Swisslex», «Fedlex», «Legalis», **REF-lex.ch**; «LexisNexis»...

Welches System könnte Ihnen in der Zukunft entscheidende Argumente «pfannenfertig» liefern?

Chat-GPT, Claude, Legora, Harvey, (Legal-)BERT (Google) etc.

(ad 1) KI in der Strafrechtspflege

Schweizer Justiz überlastet: «Es profitieren die Täter»

- Tagesanzeiger, 25. Juli 2023

Die Staatsanwaltschaften klagen über Überlastung

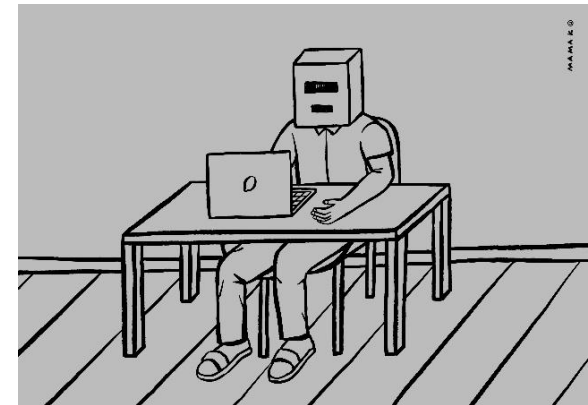
- NZZ, 1. Oktober 2012

A.I. Is Coming for Lawyers, Again

- The New York Times, 10. April 2023

Generative AI could radically alter the practice of law

- The Economist, 6. Juni 2023



Könnten IT-Systeme Strafbefehle

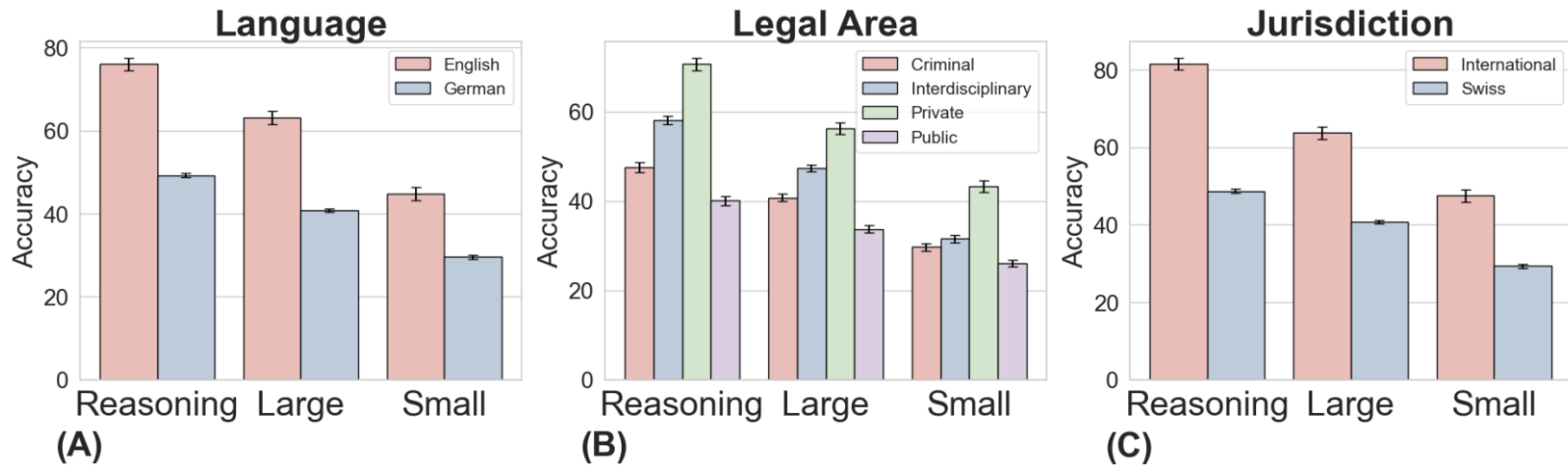
- billiger, schneller, unter Berücksichtigung von mehr Daten ausfertigen;
- weniger Fehler machen und Willkür vermeiden, was insgesamt zu gerechteren Ergebnissen führen könnte.

(ad 2) Strafbefehle auf Knopfdruck – Geht das?

GPT-4 besteht Bar-Exam in den USA (erreicht Resultate in den Top 10%).

Katz, Daniel Martin and Bommarito, Michael James and Gao, Shang and Arredondo, Pablo, GPT-4 Passes the Bar Exam (March 15, 2023). 382 Philosophical Transactions of the Royal Society A (2024), Available at SSRN: <https://ssrn.com/abstract=4389233> or <http://dx.doi.org/10.2139/ssrn.4389233>

(ad 2) Strafbefehle auf Knopfdruck – Geht das?



Fan, Y. et al. (2025) 'LEXam: Benchmarking Legal Reasoning on 340 Law Exams', arXiv [cs.CL] [Preprint]. Available at: <http://arxiv.org/abs/2505.12864>.

(ad 2) Strafbefehle auf Knopfdruck – Geht das?

Maslow's Hammer

"If all you have is a hammer, everything looks like a nail."

(ad 2) Strafbefehle auf Knopfdruck – Geht das?

Die Probleme des LLM-Hammers: Causal Language Modelling

Grosse Sprachmodelle werden auf dem Inhalt "des Internets" trainiert. Sie lernen Muster und Strukturen in Texten, um menschenähnliche Antworten zu generieren.

Unwahrscheinlich

Die Richterin fällt ein

Baum
fach
entsprechendes
Urteil

Wahrscheinlich

(ad 2) Strafbefehle auf Knopfdruck – Geht das?

Probleme des LLM-Hammers: Tokenisierung.

- Sprachmodelle sehen Text in sogenannten Tokens. Bereits auf dieser Ebene (noch vor dem eigentlichen Trainingsprozess eines Modelles) entstehen erste Probleme.

Art. 112 StGB regelt den Mord.

Art. 112 StGB regelt den Mord.

8372, 13, 220, 12307, 901, 9068, 132
4, 3292, 1786, 97780, 13

(ad 2) Strafbefehle auf Knopfdruck – Geht das?

Die Probleme des LLM-Hammers: Causal Language Modelling

Beispiel nach wenigen Stunden Training eines Sprachmodells auf BGEs:

"Die entsprechende Beurteilung der Beurteilung der Beurteilung des Beurteilten wurde aufgrund eines sagenkundigen Beurteilten zur Beurteilung weitergezogen."

(ad 2) Strafbefehle auf Knopfdruck – Geht das?

Die Probleme des LLM-Hammers: Alignment-Tuning vs. Einsatz im Strafrecht

- Kleinere Modelle klassifizieren strafrechtliche Inhalte jedoch häufig als „*illegal*“ oder „*schädlich*“ und verweigern die Antwort:

"Es tut mir leid, aber ich kann diese Anfrage nicht bearbeiten, da sie gegen die Nutzungsrichtlinien verstößt. Diese beinhalten respektvolle und gesetzeskonforme Kommunikation."

(ad 2) Strafbefehle auf Knopfdruck – Geht das?

Dissertationsprojekt von MLaw Colin Carter (Colin.Carter@unibas.ch)

Forschungsfrage:

Können Strafbefehle autonom erlassen werden?

Ziel:

Prototype for Automating Penal Orders (PAPO)

Testphase mit mehreren hundert anonymisierten Scans von Strafbefehlen und dazugehörige Dossiers der Staatsanwaltschaft Basel-Stadt und Basel-Landschaft, zurzeit beschränkt auf Art. 91 SVG. Für Vergleichszwecke werden auch 50 anonymisierte Fälle der Trunkenheitsfahrt aus Deutschland beigezogen.

(ad 2) Strafbefehle auf Knopfdruck – Geht das?

Stellen Sie sich vor, Sie erhalten folgendes Schreiben der Staatsanwaltschaft:

STRAFBEFEHL

1. Die beschuldigte Person hat sich schuldig gemacht, am 20.1.2024 in Basel mit einer BAK von mindestens 1.4 g/kg (Grenzwert 0.8 g/kg) ein Motorfahrzeug geführt zu haben.

Beweis: BAK-Auswertung, Rechtsmedizinisches Institut, Basel-Stadt

2. Dadurch hat sich die beschuldigte Person des Fahrens in fahruntüchtigem Zustand gem. Art. 91 Abs. 2 lit. a SVG schuldig gemacht.

3. Diese Straftat wird mit einer Geldstrafe von 6'000 CHF (50 Tagessätze à 120 CHF) und einer Busse von 1'200 CHF geahndet.

Dieser Strafbefehl wurde vollautomatisch erstellt.

Ohne Einsprache innert 10 Tage erwächst er in Rechtskraft

(ad 2) Strafbefehle auf Knopfdruck – Geht das?

Welcher Verfahrensschritt wird auf welcher Datengrundlage automatisiert?

Komplette Automatisierung ab Polizeirapport

- Klassifikation:** Ausgehend von blosser Sachverhaltsumschreibung und Polizeirapport müssen die relevanten Delikte erkannt werden.
- Extraktion:** Für die Subsumtion müssen sämtliche relevanten Straftatbestandsmerkmale (mögl. Einflussfaktoren) extrahiert werden.
- Subsumtion:** Anhand der relevanten Straftatbestandsmerkmale wird die rechtliche Beurteilung vorgenommen. **I.c.** “Wenn die gemessene BAK $\geq 0,8$ Gewichtspromille oder die Atemalkoholkonzentration 0,4 mg Alkohol oder mehr pro Liter Atemluft beträgt
→ return **TRUE**”
- Strafzumessung:** gemäss Tabelle.

(Quelle: STRAFMASSEMPFEHLUNGEN vom 27.05.2022, OSTA Kanton

Zürich) Verbindungsbusse: 20% der GS, mindestens aber 300 CHF.

Ersttäter	Blutalkoholkonzentration g/kg (Promille)	Atemalkoholkonzentration mg/l	Zusatzbusse vgl. Ausführungen Seite 2
	ab 0,5 ‰ ab 0,6 ‰ ab 0,7 ‰	ab 0.25 ab 0.3 ab 0.35	mit Busse zu bestrafende Übertretung (Statthalteramt) Fr. 600.-- Fr. 700.-- Fr. 800.--
	ab 0.8 ‰	ab 0.4	ab 20 Tagessätze GS
	ab 1.0 ‰	ab 0.5	ab 30 Tagessätze GS
	ab 1.2 ‰	ab 0.6	ab 40 Tagessätze GS
	ab 1.4 ‰	ab 0.7	ab 50 Tagessätze GS
	ab 1.6 ‰	ab 0.8	ab 60 Tagessätze GS
	ab 1.8 ‰	ab 0.9	ab 80 Tagessätze GS
	ab 2.0 ‰	ab 1.0	ab 100 Tagessätze GS
Wiederholungstäter	ab 0.8 ‰	ab 0.4	ab 30 Tagessätze GS
	ab 1.2 ‰	ab 0.6	ab 60 Tagessätze GS
	ab 1.5 ‰	ab 0.75	ab 75 Tagessätze GS
	ab 2.0 ‰	ab 1.0	ab 150 Tagessätze GS

(ad 2) Strafbefehle auf Knopfdruck – Geht das?

Art. 91 SVG - Fahren in fahrunfähigem Zustand und Missachtung des Verbots, unter Alkoholeinfluss zu fahren (SVG, SR 741.01)

¹ Mit Busse wird bestraft, wer:

- a.** in angetrunkenem Zustand ein Motorfahrzeug führt;
- b.** das Verbot, unter Alkoholeinfluss zu fahren, missachtet;
- c.** in fahrunfähigem Zustand ein motorloses Fahrzeug führt.

² Mit Freiheitsstrafe bis zu drei Jahren oder Geldstrafe wird bestraft, wer:

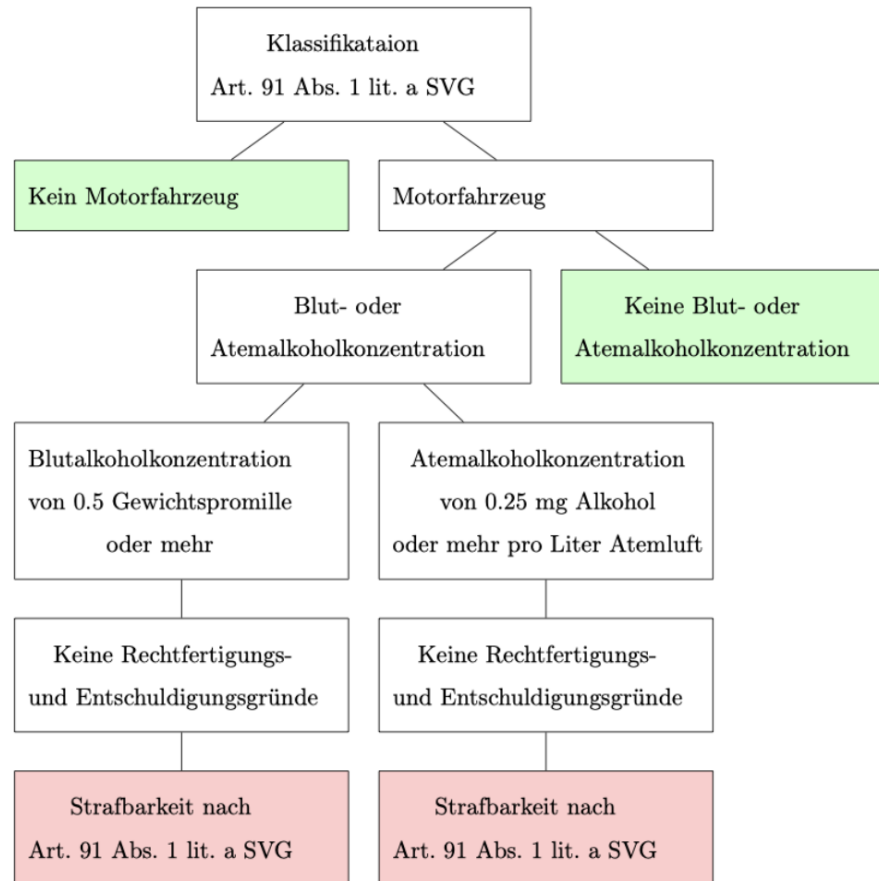
- a.** in angetrunkenem Zustand mit qualifizierter Atemalkohol- oder Blutalkoholkonzentration ein Motorfahrzeug führt;
- b.** aus anderen Gründen fahrunfähig ist und ein Motorfahrzeug führt.

[Beurteilung folgt klarem schematischem Vorgang, Prüfung des Vorliegens einzelner Tatbestandsmerkmale folgt einem Entscheidungsbaum. Subsumtion kann regelbasiert (wenn-dann) vorgenommen werden.]

[Erfordert Gewichtung mehrerer Einflussfaktoren → Regelbasierte Ansätze werden die Komplexität nicht abbilden können.]

(ad 2) Strafbefehle auf Knopfdruck – Geht das?

Subsumtion: Regelbasierte Ansätze



(ad 2) Strafbefehle auf Knopfdruck – Geht das?

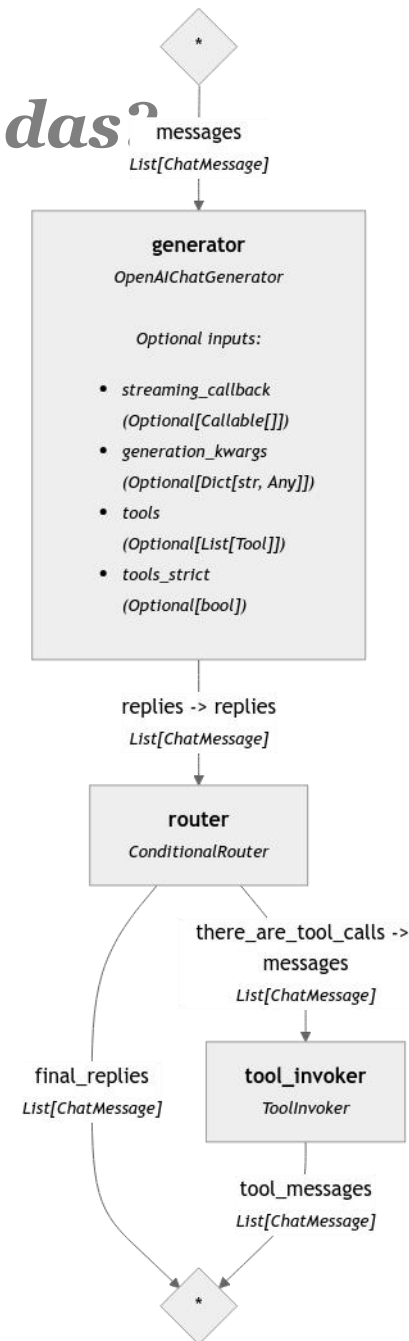
Subsumtion: Hybride Ansätze (Agenten)

Regelbasierte Systeme können (und müssen) mit Machine-Learning-Systemen kombiniert werden, um die Vorteile beider Ansätze zu nutzen. Ein Ansatz wäre ein agentisches System, bei dem ein LLM die relevanten Informationen aus den Akten extrahiert und anschliessend ein regelbasiertes System den Fall bewertet und den Strafbefehl generiert.

Benutzereingabe: Herr Muster gab bei einer Polizeikontrolle eine Atemalkoholprobe ab.
Das Messergebnis betrug 0,3.

LLM-Antwort:

```
{ 'tool_invoker': { 'tool_messages': [ ChatMessage(_role=, _content=[ToolCallResult(
result='Not exceeded', origin=ToolCall( tool_name='91_2_a_SVG', arguments={'type':
'BrAL', 'measurement': 0.3}), id='call_564r' ), error=False)], _meta={}) ] } }
```



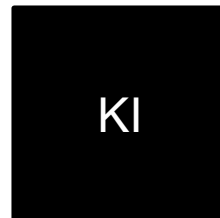
(ad 2) Strafbefehle auf Knopfdruck – Geht das?

Aufgabe für ein trainiertes System (Risikobasierter Ansatz).

² Mit Freiheitsstrafe bis zu drei Jahren oder Geldstrafe wird bestraft, wer:

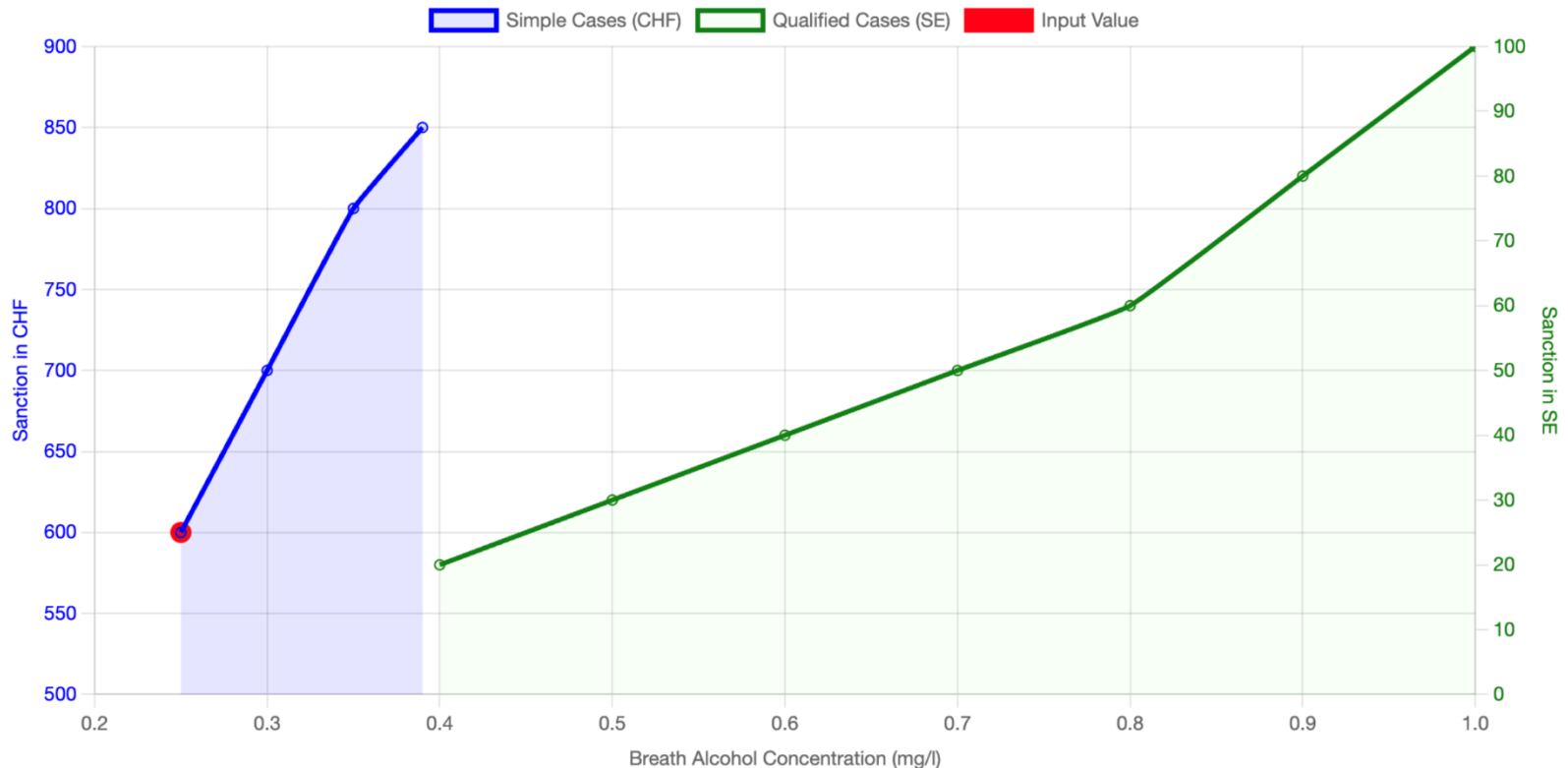
- a. in angetrunkenem Zustand mit qualifizierter Atemalkohol- oder Blutalkoholkonzentration ein Motorfahrzeug führt;
- b. aus anderen Gründen fahruntfähig ist und ein Motorfahrzeug führt.

Substance	Value
BAC	0.5 ‰
Benzodiazepine	22 ng/mL
Sex	M
Weight	86kg



Fahruntfähigkeit
festgestellt mit
Sicherheit von: 78%

(ad 2) Strafbefehle auf Knopfdruck – Geht das? Strafzumessung



Eingabe eines Atemalkoholwerts in mg/l: (0.25 mg/l)

Category: Simple Cases (CHF) – Estimated Sanction: 600.0 CHF

SE = Sanktionseinheit (in Tagen oder Tagessätzen)

(ad 3) Rechtliche Grenzen des Einsatzes von KI

Datenschutz, Amtsgeheimnis und Unabhängigkeit, Nachhaltigkeit

- Aufgrund des hohen Schutzbedarfs von Strafverfahrensdaten (**Datenschutz, Amtsgeheimnis, Datensouveränität**) können idR. nicht die *state-of-the-art* Modelle der (US-Amerikanischen) Cloud-Anbieter verwendet werden.
- Die Verwendung von **closed-source** Modellen bedeutet sodann auch eine vollständige **Abhängigkeit** von den Cloud-Anbietern (Weiterbetrieb der Modelle nicht garantiert, keine Erklärbarkeit, keine Reproduzierbarkeit).
- KI ist **nie neutral** - sie reflektiert die Entscheidungen ihrer Entwickler: Die Modell-Hersteller wählten die verwendeten Trainingsdaten aus, die Programmierer definieren Entscheidungsbäume und Funktionsweisen (bspw. Verfügbare "Werkzeuge") und Bias in Trainingsdaten wird durch Feedback und Reinforcement Learning weiter verstärkt.
- Für das Forschungsprojekt wird deshalb auf **open-source** oder mindestens **open-weights** Modelle gesetzt. Prominent und für das Projekt von besonderer Bedeutung sind die **Schweizer Apertus Modelle** der ETH/EPFL/CSCS.
- Der Betrieb von Sprachmodellen erfordert hohe **Rechenkapazitäten**. Im Rahmen des Forschungsprojektes erhalten wir Zugang zum besonders gesicherten HPC-Cluster SciCORE+ der Universität Basel.

(ad 3) Rechtliche Grenzen des Einsatzes von KI

Spezifische Probleme der Nutzung von LLM-Modellen

- Aufgrund der Architektur ergibt sich eine Antwort aus den gelernten Mustern. Dies könnte zu einer **Versteinerung des Rechts** führen. Ohne menschliche, neue Impulse wird das Sprachmodell *nur* aus diesen vergangenen, gelernten Mustern schöpfen.
- Grosse Sprachmodelle sind **Wahrscheinlichkeitsmaschinen**. Sie generieren Texte, indem sie die wahrscheinlichsten nächsten Tokens (Wörter oder Wortteile) vorhersagen. Halluzinationsfreie Systeme, die auf den aktuellen generativen KI-Modellen berufen **gibt es nicht und sind technisch ausgeschlossen**, auch wenn einige Anbieter solche Marketingversprechungen machen.

(ad 3) Rechtliche Grenzen des Einsatzes von KI

Spezifische Probleme der Nutzung von LLM-Modellen

Durch die inzwischen weite Verbreitung von generativer KI sind sog. **Halluzinationen** vielerorts anzutreffen und finden sich inzwischen auch in Urteilen verbreitet. **Damien Charlotin** trackt diese auf seiner Webseite und hat bis am 30. April 2026 bereits **1356 Fälle** mit Halluzinationen zählen können. Aus der Schweiz gibt es bis heute noch keinen Fall.

Damien Charlotin Scholarship Curriculum Vitae Data ▾ Book Projects ▾   

AI Hallucination Cases

This database tracks legal *decisions*¹ in cases where generative AI produced hallucinated content – typically fake citations, but also other types of AI-generated arguments. It does not track the (necessarily wider) universe of all fake citations or use of AI in court filings.

While seeking to be exhaustive (1356 cases identified so far), it is a work in progress and will expand as new examples emerge. This database has been featured in news media, and indeed in several decisions dealing with hallucinated material.²

If you have any questions about the database, a FAQ is available [here](#).
And if you know of a case that should be included, feel free to [contact me](#).³

Based on this database, I have developed an [automated reference checker](#) that also detects hallucinations: PelAIkan. Check the Reports  in the database for examples, and [reach out to me](#) for a demo !

For weekly takes on cases like these, and what they mean for legal practice, subscribe to Artificial Authority.

Search cases, courts, tools, tags

Case	Court / Jurisdiction	Date ▾	Party Using AI	AI Tool ①	Nature of Hallucination	Outcome / Sanction	Monetary Penalty	Details
Anthony C. Hill v. Workday, Inc.	N.D. California (USA)	28 April 2026	Lawyer	CoCounsel	Fabricated Case Law (1) Misrepresented Case Law (1)	Admonishment; Monetary Fine; 4 hours live CLE; Order Circulation	1001 USD	
Brandon Rulund Akins v. Tashi Latwon Ratliff, et al.	M.D. North Carolina (USA)	28 April 2026	Pro Se Litigant	Implied	Fabricated Case Law (9)	Warning	—	



Artificial Authority

A look at news and developments at the intersection of AI and the Law
By DamienCh · Over 1,000 subscribers

Type your email...

By subscribing you agree to [Substack's Terms of Use](#), our [Privacy Policy](#) and our [information collection notice](#)

#substack

Last updated: 29 April 2026

(ad 3) Rechtliche Grenzen des Einsatzes von KI

Was spricht für einen Einsatz von KI?

- Speedy trial: Schnelligkeit und Rechtssicherheit für Angeklagte/Beschuldigte und Opfer in Strafverfahren. Angeklagte und Opfer warten oft Monate, wenn nicht Jahre auf Rechtssicherheit....
- mehr Zeitressourcen für Strafverfolgungsbehörden
Effizientere, schnellere Ergebnisse, weniger Ressourcen für sich wiederholende Aufgaben...
- mögliche Verbesserungen für Betroffene betr. Zugang zur Justiz
(z. B. Übersetzung, angepasste Sprache und interaktive Erklärungen; möglicherweise weniger fehleranfällig bei der Bearbeitung einfacher Formalitäten; einheitlichere Anwendung des Rechts [gerechtere Ergebnisse?])

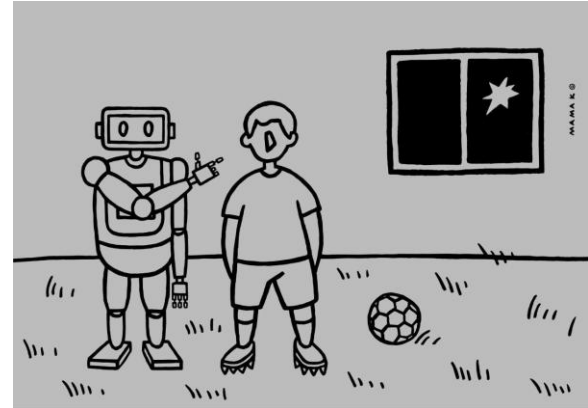
(ad 3) Rechtliche Grenzen des Einsatzes von KI

Was spricht gegen einen Einsatz von KI?

(neben Fragen der Nachhaltigkeit, gesellschaftlichen Erwägungen...)

- Unabhängigkeit und Unparteilichkeit von Strafverfolgungsbehörden (betr. Gestaltung, Kodierung oder Schulung solcher Systeme);
- Verantwortung für einen Strafbefehl («Entscheidungsgewalt ohne Entscheidungsverantwortung»);
Anker-Effekte“ von „Assistenzsystemen“;
- Recht auf ein faires Verfahren?;
- mögliche Justizirrtümer (Diskriminierung, „Halluzinationen“);-
- Menschenwürde und «menschliches Erbarmen» in der Strafjustiz.

(ad 4) Wollen wir KI in der Strafrechtspflege einsetzen?



Um Vor- und Nachteile abwägen zu können,

- Ressourcenersparnis?
- ist KI mindestens genauso so gut /besser als Menschen (benchmark)?;
- können Fehlurteile effektiv vermieden werden (adäquater Nachweis)?
- wer trägt am Ende die Verantwortung?

Some Stupid with a Flare Gun?

Ausgelöst hat die Katastrophe von 1971 ein Besucher eines Frank Zappa-Konzerts, der mit einer Leuchtpistole in die Decke schoss und so das Gebäude in Brand setzte – eine einzige unverantwortliche Handlung verursachte das sich rasch ausbreitende Feuer. Wenn potente Werkzeuge „dumm“ eingesetzt werden, können sie massiven Schaden anrichten.





Universität
Basel

Juristische
Fakultät

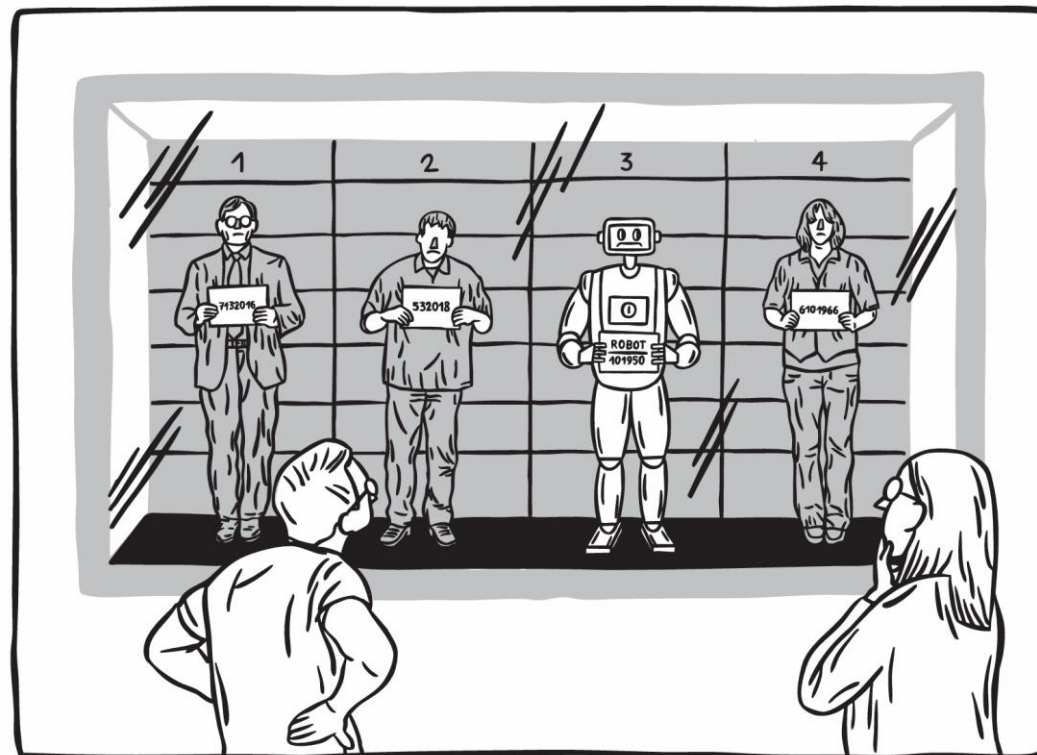


Danke für die Aufmerksamkeit!

Sabine.Gless@unibas.ch

Colin.Carter@unibas.ch

(ad 1) KI in der Strafrechtspflege



Literaturauswahl

Gless S, Künstliche Intelligenz in der Gerichtsbarkeit, (Warum) Braucht es weiter menschliche Richterinnen und Richter, ZSR 142 (2023) I Heft 5, 429-462

Niklaus J/Chalkidis I/Stürmer M, Swiss-Judgment-Prediction: A Multilingual Legal Judgment Prediction Benchmark, Natural Legal Language Processing Workshop 2021, 19 ff.

Chalkidis I/Fergadiotis M/Malakasiotis P/Aletras N/Androutsopoulos I, LEGAL-BERT: The Muppets straight out of Law School, in: Cohn T/He Y/Liu Y, Findings of the Association for Computational Linguistics: EMNLP 2020, 2898 ff.

Donohue ME, A Replacement for Justitia's Scales: Machine Learning's Role in Sentencing, Harvard Journal of Law & Technology 32 (2019), 657 ff.

Gless S/Wohlers W, Subsumtionsautomat 2.0 – Künstliche Intelligenz statt menschlicher Richter?, in: Böse M/Schumann KH/Toepel F (Hrsg.), Festschrift für Urs Kindhäuser, Nomos 2019, 147-165

Ashley KD, Artificial Intelligence and Legal Analytics: New Tools for Law Practice in the Digital Age, Cambridge University Press 2017