Vertrauenswürdige KI in der juristischen Praxis

Rechtliche Rahmenbedingungen für eine effektive Integration von Legal Tech

Florian Wienert, LL.B.
Sigmund Freud PrivatUniversität Wien

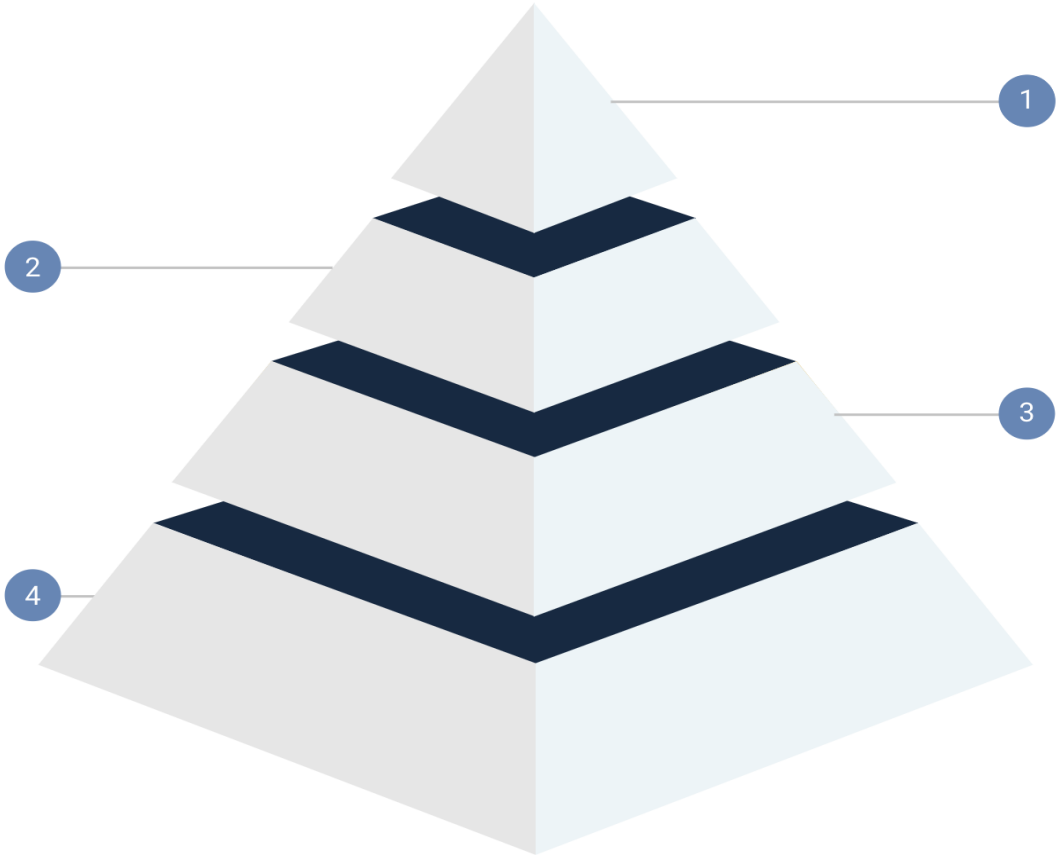
1 Einführung

- **Legal Tech**: KI-gestützte juristische Recherche & automatisierte Entscheidungsfindung
- **AI-Act**: Rechtsrahmen für Nutzung von KI-Systemen in der EU

- **Forschungsfragen**:
  - *Welche rechtlichen Rahmenbedingungen legt die KI-VO fest, um eine vertrauenswürdige und zuverlässige Nutzung von KI-Systemen in der juristischen Praxis zu gewährleisten?*
  - *Welche technischen Rahmenbedingungen sind erforderlich, um die rechtlichen Vorgaben der KI-VO in der Praxis umzusetzen?*

2 AI-Act

Klassifizierung von KI-Systemen in Risikostufen



1 **Inakzeptables Risiko** (Art 5 AI-Act)
→ *zB Social-Scoring*
→ **verboten**

2 **Hohes Risiko** (Art 6 AI-Act)
→ *zB kritische Infrastruktur, Justiz*
→ **strenge Anforderungen**

3 **Begrenztes Risiko** (Art 50 AI-Act)
→ *zB Chatbots*
→ **Transparenzpflichten**

4 **Minimales Risiko**
→ *zB Spamfilter*
→ keine speziellen Anforderungen

2 AI-Act

Wesentliche Anforderungen bei Hochrisiko-Systemen

- Konformitätsbewertungsverfahren (Art 43)

→ **Transparenz** & (hinreichende) **Nachvollziehbarkeit** (Art 11 ff)

→ **Menschliche Kontrolle** (Art 14)

→ Betreiberpflichten (Art 16 ff)

Abschnitt 2	–	**Anforderungen an Hochrisiko-KI-Systeme**
Art. 8 KI-VO	–	Einhaltung der Anforderungen
Art. 9 KI-VO	–	Risikomanagementsystem
Art. 10 KI-VO	–	Daten und Daten-Governance
Art. 11 KI-VO	–	Technische Dokumentation
Art. 12 KI-VO	–	Aufzeichnungspflichten
Art. 13 KI-VO	–	Transparenz und Bereitstellung von Informationen für die Betreiber
Art. 14 KI-VO	–	Menschliche Aufsicht
Art. 15 KI-VO	–	Genauigkeit, Robustheit und Cybersicherheit
Abschnitt 3	–	**Pflichten der Anbieter und Betreiber von Hochrisiko-KI-Systemen und anderer Beteiligter**
Art. 16 KI-VO	–	Pflichten der Anbieter von Hochrisiko-KI-Systemen
Art. 17 KI-VO	–	Qualitätsmanagementsystem
Art. 18 KI-VO	–	Aufbewahrung der Dokumentation
Art. 19 KI-VO	–	Automatisch erzeugte Protokolle
Art. 20 KI-VO	–	Korrekturmaßnahmen und Informationspflicht
Art. 21 KI-VO	–	Zusammenarbeit mit den zuständigen Behörden
Art. 22 KI-VO	–	Bevollmächtigte der Anbieter von Hochrisiko-KI-Systemen
Art. 23 KI-VO	–	Pflichten der Einführer
Art. 24 KI-VO	–	Pflichten der Händler
Art. 25 KI-VO	–	Verantwortlichkeiten entlang der KI-Wertschöpfungskette
Art. 26 KI-VO	–	Pflichten der Betreiber von Hochrisiko-KI-Systemen
Art. 27 KI-VO	–	Grundrechte-Folgenabschätzung für Hochrisiko-KI-Systeme

3 Technische Anforderungen

Ableitung aus den rechtlichen Vorgaben des AI-Act

- **Hohes Risiko** (*zB Legal Prediction in Justiz*): erhöhte Erklärbarkeit, strenge Validierung, Betreiberpflichten
- **Begrenztes Risiko** (*zB Recherchetools*): bloße Transparenzpflicht

- <u>Problem der Zuverlässigkeit</u>; abhängig von Aufgabentyp und Komplexität

→ Zuverlässigkeit/Vertrauenswürdigkeit ist insb **technische Herausforderung**

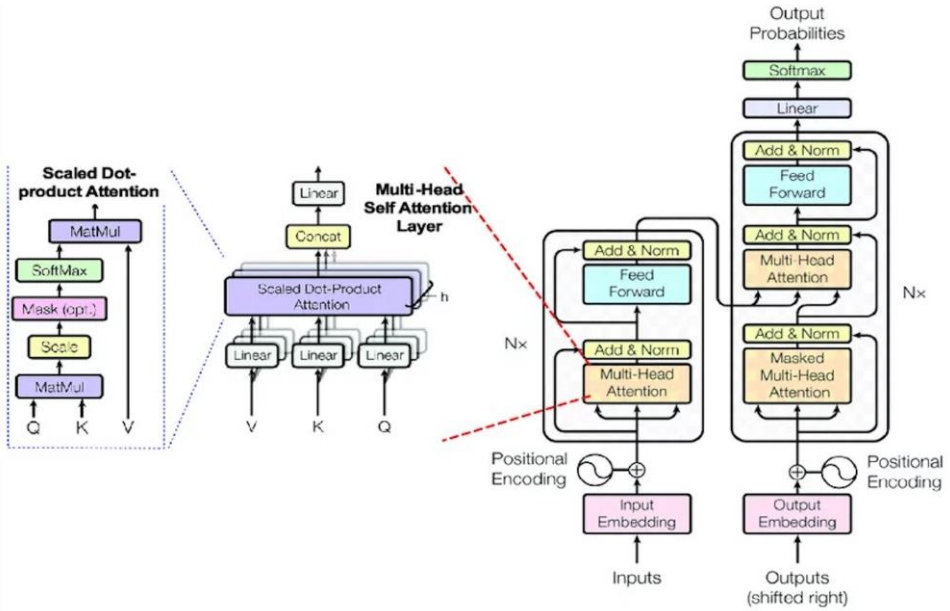
→ Fokus auf **Erhöhung der Vertrauenswürdigkeit** bei universaler (juristischer) Anwendung

3 Technische Anforderungen

Funktionsweise eines LLMs

- **Grundprinzip**: Transformer-Architektur
  - Sprache wird mithilfe von *Self-Attention*-Mechanismen in diverse Schichten verarbeitet
  - Modell berechnet auf Basis von Wahrscheinlichkeiten, welches Wort (Token) als nächstes folgt

→ *Hochdimensionales Modell der Wahrscheinlichkeitsverteilung*



höhere Anzahl an Parametern → präziserer Output
höhere Anzahl an Parametern → erschwerte Nachvollziehbarkeit (Black-Box-Problem)

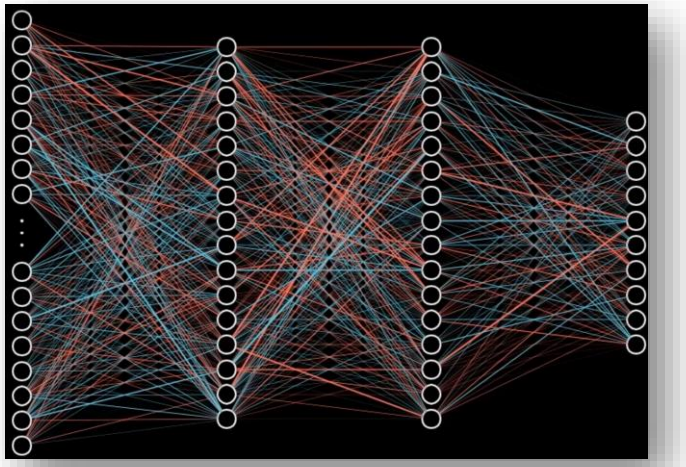
3 Technische Anforderungen

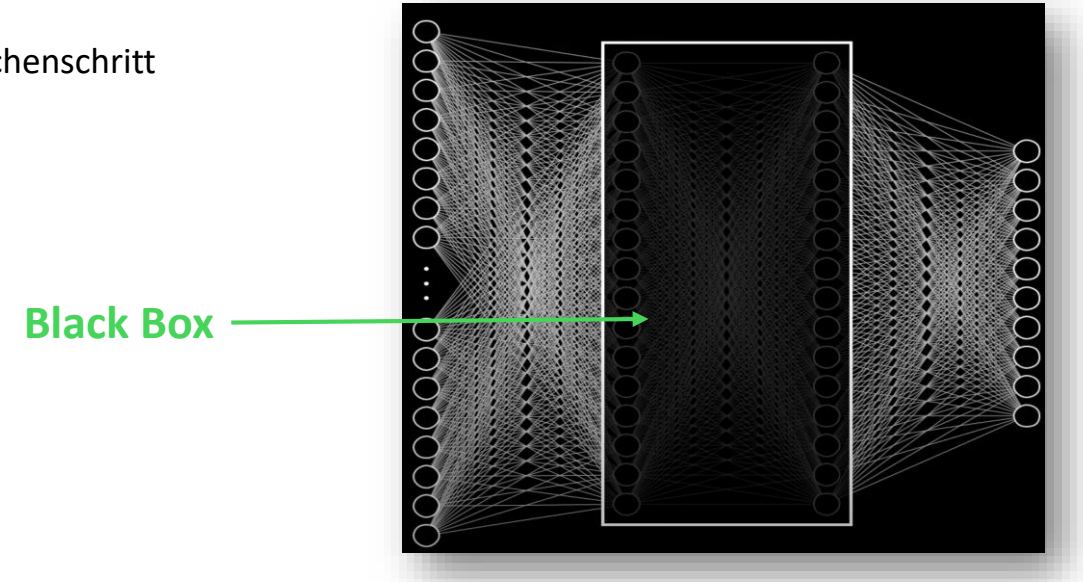
Lösungsansätze: <u>Transparenz und Erklärbarkeit</u>

Stichwort: *hinreichende Nachvollziehbarkeit*

- ***Ex-post*-Begründungen**
  - Extra-Module
  - Begründung muss nicht (und kann auch nicht) intern in jedem Rechenschritt der KI liegen

- **Dokumentation des Prompts**
- (Kennzeichnungspflichten)





3 Technische Anforderungen

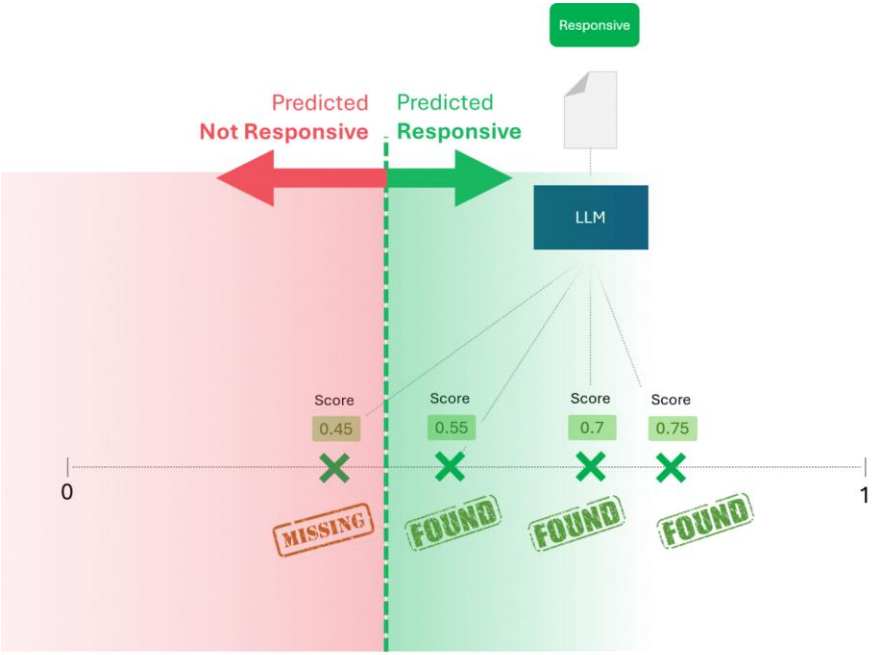
Lösungsansätze: <u>Umgang mit Unsicherheiten</u>

Modelle lernen nur *Korrelationen*, keine *Kausalität*

- **Confidence Scores** „Wie sicher bist du dir?“
  - unterschiedliche Interpretationen im Recht → LLM versucht aber “eine Wahrheit” zu erzwingen

    → Angabe mehrerer möglicher Labels + Confidence Score

    → Gefahr von *false Negatives*



- **Differenzierung zwischen „einfachen“ und „schwierigen“ Fragen**

3 Technische Anforderungen

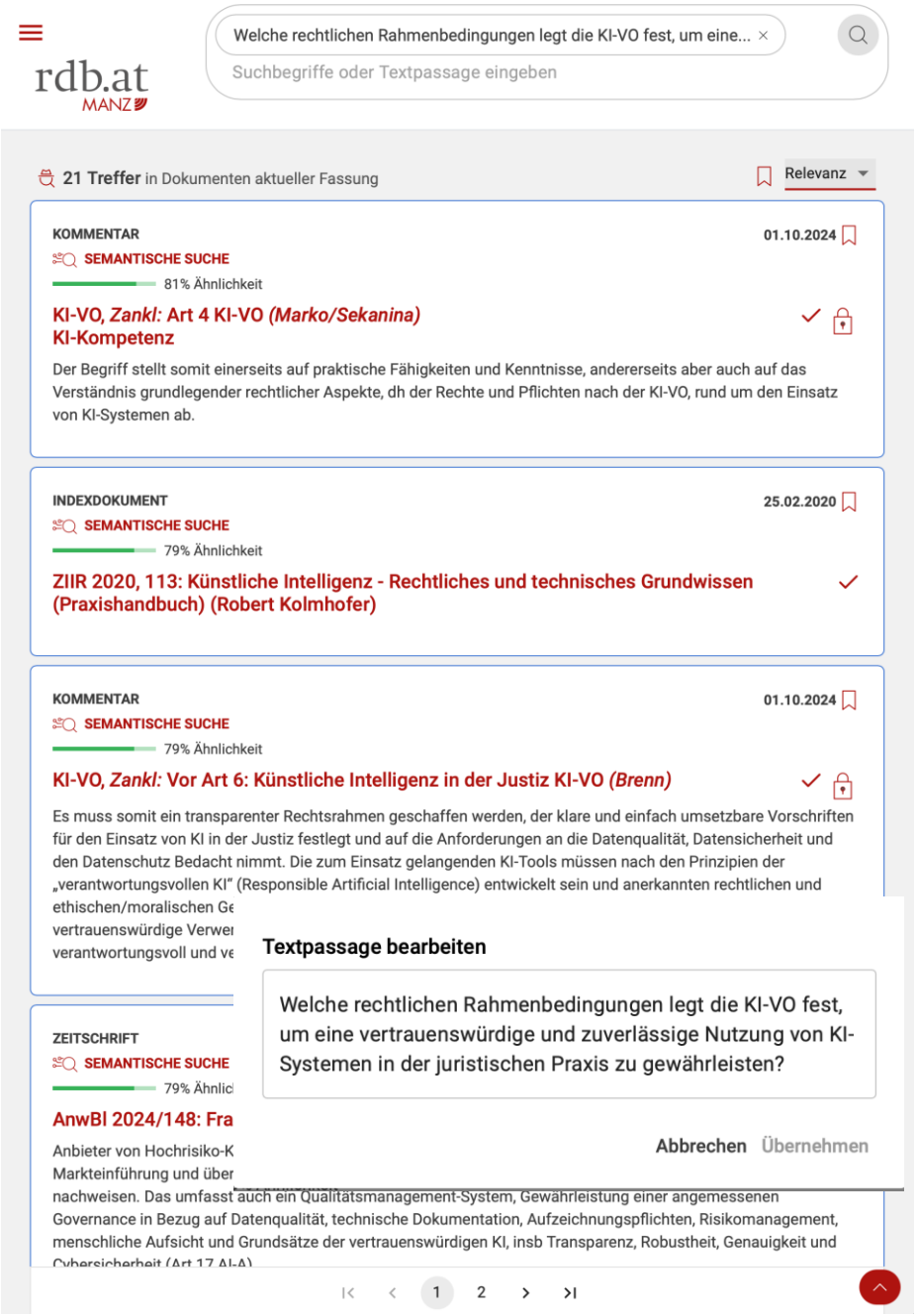
Lösungsansätze: <u>Vermeidung von Halluzinationen</u>

Kontextualisierung ist erforderlich

- ***Retrieval-Augmented Generation (RAG)***

  → Hybridansatz*:* LLM + externe Datenbanken

- <u>aktuelle Tools:</u> **Recherche** → Vektordatenbanken/**semantische Suche**



3 Technische Anforderungen

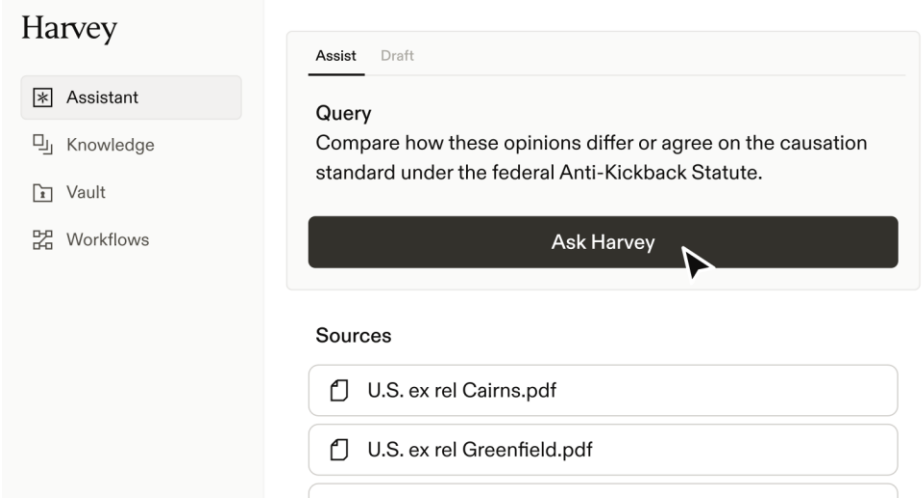
Lösungsansätze: <u>Vermeidung von Halluzinationen</u>

Kontextualisierung ist erforderlich

- ***Retrieval-Augmented Generation (RAG)***

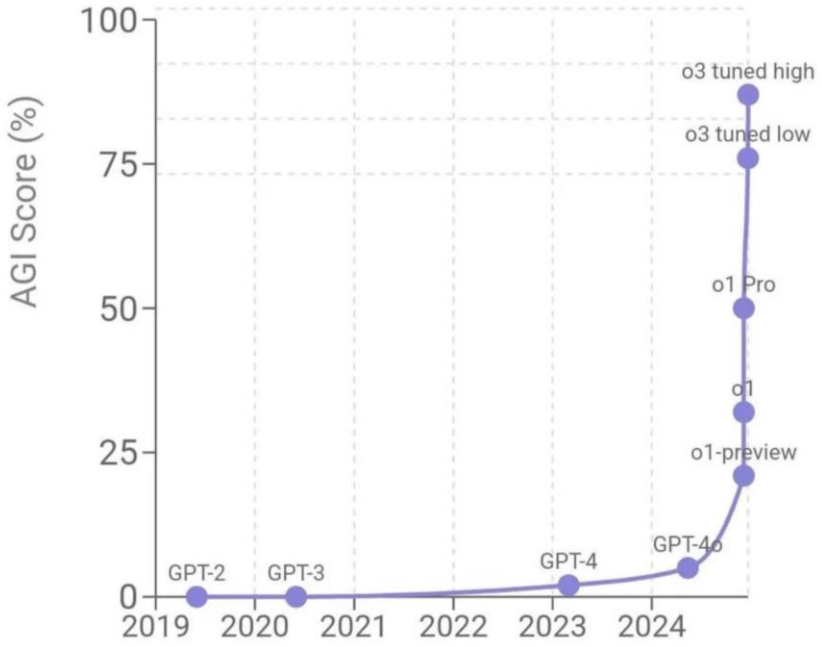
  → Hybridansatz*:* LLM + externe Datenbanken

- <u>aktuelle Tools:</u> **Universelle Anwendungen → spezialisierte LLMs**



3 Technische Anforderungen

ARC AGI Score Progression



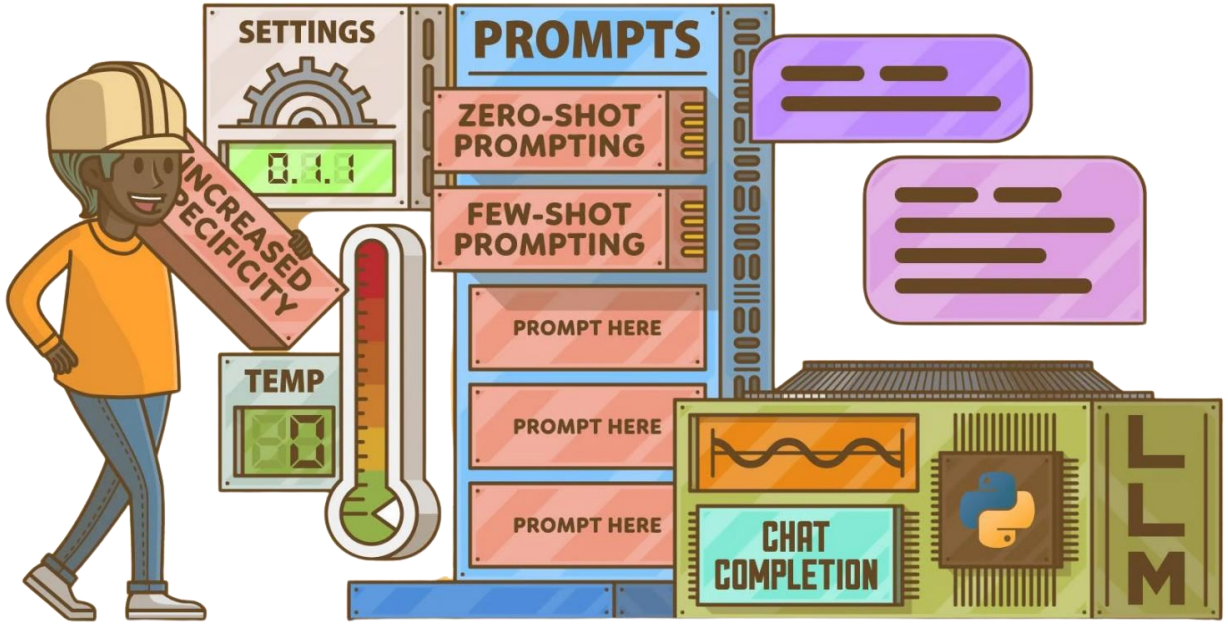
ARC AGI Score bewertet die Fähigkeit eines LLMs zur **adaptiven allgemeinen Intelligenz**

→ **Fähigkeit zu konzeptionellen/logischen Denken**

3 Technische Anforderungen

Lösungsansätze: <u>Prompt Engineering</u>

- **präzise, juristisch korrekte Anweisungen**
  - Persona-Setups
  - explizite Normen-Angabe
  - Bereitstellung von Kontext (→ siehe *RAG*)
- **Chain-of-Thought**
- **Parametersteuerung**

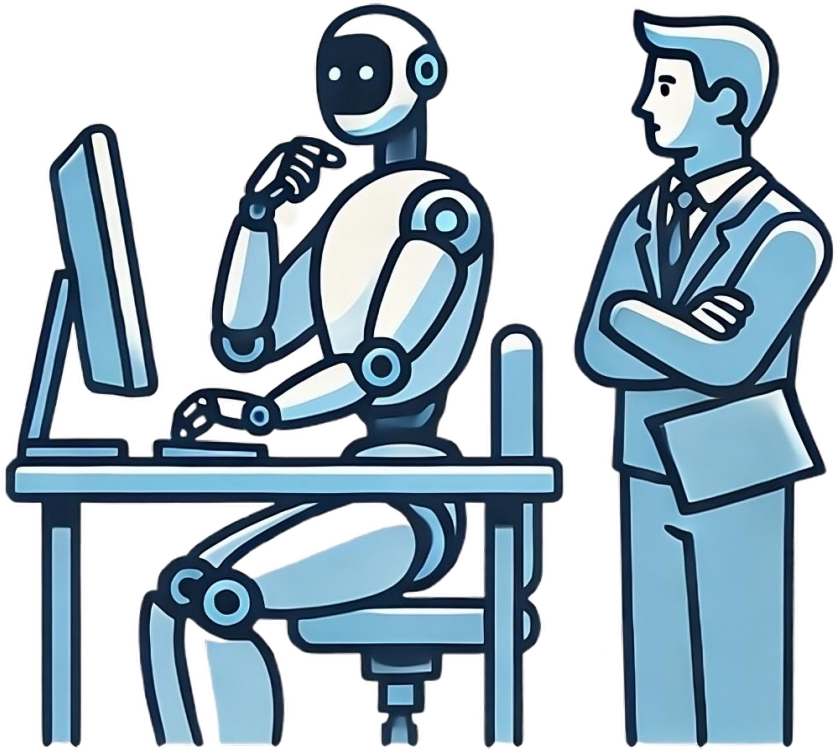


3 Technische Anforderungen

Lösungsansätze: <u>Menschliche Aufsicht</u>

- **Endkontrolle durch Menschen**
  → Menschlicher Eingriff als „Sicherungsebene“
  → Qualitätsicherung, speziell in komplexen/Hochrisiko-Fällen

- **Konformität mit AI-Act**



4 Fazit

- vielfältige Einsatzmöglichkeiten von KI in juristischer Praxis eröffnet

- **Effizienz hoch, aber Qualität unsicher**
  → hoher Bedarf an technischen Mechanismen zur Verlässlichkeit
  → korreliert mit rechtlichen Vorgaben (AI-Act)

- **Rechtliche Perspektive**
  → jedenfalls Transparenzpflichten
  → größere Hürden für Hochrisiko-Anwendungen

→ für **Anwendungen mit begrenztem Risiko** (private Anwendungen) können LLMs **jetzt schon sehr nützlich** sein (sofern technische Limits beachtet werden)

→ für **Anwendungen mit hohem Risiko** (insb Justiz) bleibt technische Entwicklung und Anpassung an AI-Act abzuwarten

Vielen Dank!

Fragen?

Florian Wienert, LL.B.
Sigmund Freud PrivatUniversität Wien