

LEGAL TECH IM RICHTERZIMMER? STREIFLICHTER AUS WISSENSCHAFT UND PRAXIS ZUM KI-EINSATZ BEI FLUGVERSPÄTUNGEN

Andreas Sesing-Wagenpfeil / Matthias Biniok / Felix Kares /
Lena Kästner / Markus Langer / Christian Metz /
Tatyana V. Peshteryanu / Ulla Wessels

Akademischer Rat a.Z., Universität des Saarlandes, Institut für Rechtsinformatik
Campus A5 4, 66123 Saarbrücken, DE
andreas.sesing@uni-saarland.de, www.rechtsinformatik.saarland

Manager IBM Client Engineering DACH – Public Sector, IBM Deutschland GmbH, IBM Client Engineering
Französische Str. 24, 10117 Berlin, DE
matthias.biniok@de.ibm.com, www.ibm.com/de-de

Wissenschaftlicher Mitarbeiter, Universität des Saarlandes, Arbeitseinheit Arbeits- und Organisationspsychologie
Campus A1.3, 66123 Saarbrücken, DE
felix.kares@uni-saarland.de, www.uni-saarland.de/fr/lehrstuhl/koenig/personen/felix-kares-msc.html

Universitätsprofessorin, Universität Bayreuth, Institut für Philosophie, Professur für Philosophie, Informatik und KI
95447 Bayreuth, DE
lena.kaestner@uni-bayreuth.de, www.lenakaestner.com

Juniorprofessor, Universität Marburg, Arbeitseinheit Digitalisierung in psychologischen Handlungsfeldern
Gutenbergstraße 18, 35032 Marburg, DE
markus.langer@uni-marburg.de

Data Scientist, IBM Deutschland GmbH, IBM Client Engineering
Französische Str. 24, 10117 Berlin, DE
christian.metz@ibm.com, www.ibm.com/de-de

Richterin am Amtsgericht, Amtsgericht Frankfurt am Main, IT-Koordinatorin
Gerichtsstraße 2, 60313 Frankfurt am Main, DE
tatyana.peshteryanu@ag-frankfurt.justiz.hessen.de, www.ag-frankfurt-justiz.hessen.de

Universitätsprofessorin, Universität des Saarlandes, Philosophisches Institut, Professur für Praktische Philosophie
Campus A2 3, 66123 Saarbrücken, DE
ulla.wessels@uni-saarland.de, www.uwessels.de

Schlagworte: *Künstliche Intelligenz; Gerichtliche Entscheidungen; Urteilsconfigurator; Entscheidungsunterstützungssystem; Flugverspätung; Automation Bias; ex post-Rationalisierung; gesetzlicher Richter; Unabhängigkeit der Richter; Urteilsbegründung; KI-Gesetz; AI Act*

Abstract: *Deutsche Amtsgerichte im Einzugsgebiet von Großflughäfen ächzen unter der Klagelast wegen Flugverspätungen. Während es für LegalTech-Anbieter ein Leichtes ist, mithilfe von automatisierten Systemen die voraussichtliche Begründetheit einer Flugverspätungsklage vorherzusagen und eine entsprechende Klage zu erheben, entscheiden Richterinnen und Richter der betroffenen Amtsgerichte – schon von Gesetzes wegen – durch „manuelle“ Prüfung. Der Begriff der „Waffengleichheit im Zivilprozess“ bekommt hierdurch eine völlig neue Facette: In Frage steht nicht die formelle Gleichheit von Klägerin und Beklagtem; in Frage steht vielmehr die technologische Gleichheit von Parteien und Gericht. Der Beitrag führt Wissenschaft und Praxis am Beispiel des KI-Einsatzes bei Flugverspätungsklagen zusammen und zeigt auf, dass der Einsatz von KI in der Ziviljustiz gleichermaßen Chancen und Risiken birgt.*

Hinweis: *Das vorliegende Manuskript ist – soweit es die Beiträge der Autorinnen Kästner und Wessels sowie der Autoren Kares, Langer und Sesing-Wagenpfeil betrifft – im Rahmen des von der Volkswagen Stiftung geförderten Projekts „Explainable Intelligent Systems (EIS)“ (AZ 9B 830; AZ 98 511–98 513) entstanden. Zudem wurde die Arbeit des Autoren Langer unterstützt durch den DFG grant*

389792660 als Teil des Sonderforschungsbereichs 248 „Grundlagen verständlicher Software-Systeme“. Im Übrigen gibt der Beitrag – ungeachtet einer dienstlichen Befassung – ausschließlich die persönliche Auffassung der Autorin Peshteryanu sowie der Autoren Biniok und Metz wieder.

1. Einführung

Gerichtliche Massenverfahren in gleichgelagerten Fallgestaltungen nehmen in jüngerer Zeit drastisch zu. Ein Umstand, der zu dieser Entwicklung beiträgt, ist der Einsatz von LegalTech-Systemen: Hierauf spezialisierte Diensteanbieter prüfen anhand einfacher Kurz-Fragebögen, ob Rechtsuchenden Ansprüche gegen Fluggesellschaften wegen Flugverspätungen,¹ gegen Automobilhersteller wegen der Manipulation von Diesel-Abgaswerten² oder gegen Wohnungsgesellschaften wegen gesetzwidrig überhöhter Mietpreise³ zustehen.

Die Anbieter – häufig Inkassounternehmen – bewegen sich dabei in Deutschland seit der Grundsatzentscheidung „wenigermiete.de“⁴ des Bundesgerichtshofs vom 27.11.2019 im zulässigen Rahmen, da der Einsatz von LegalTech-Instrumenten, auch gepaart mit der in der Branche üblichen Vereinbarung von Erfolgshonoraren, mit den Vorgaben des Rechtsdienstleistungsgesetzes im Einklang steht.⁴

Demgegenüber kann und darf die Justiz nicht mit gleichen Waffen zurückschlagen: Richterinnen und Richter sind (auch) von Gesetzes wegen als menschliche Wesen aus Fleisch und Blut konzipiert, die Ersetzung durch bloße „Subsumtionsautomaten“ ist daher (noch) eher Dystopie als ein kurzfristig bevorstehendes, realistisches Szenario. Dies führt dazu, dass etwa alle beim Amtsgericht Frankfurt am Main – das Gericht ist aufgrund des Erfüllungsorts (§ 29 ZPO) zuständiges Gericht für verspätete Flüge mit Abflug- oder Ankunftsartort Frankfurt⁵ – eingehenden Klagen wegen verspäteter oder annullierter Flüge bis auf weiteres „manuell“ bearbeitet werden müssen. Der hiermit verbundene Arbeitsaufwand ist immens: So wurden am Amtsgericht Frankfurt am Main in den Spitzenjahren 2019 und 2020 allein ca. 15.000 Verfahren in Reisesachen anhängig gemacht, denen – gemessen an ihren Arbeitskraftanteilen – rund 40 Richterinnen und Richter gegenüberstanden.⁶ Rechnerisch betrachtet müssen damit – auf eine volle Richterstelle bezogen – etwa 30 zusätzliche Streitverfahren pro Monat einer (streitigen) Entscheidung zugeführt werden, wobei im gesamten Verfahrensgang mit jeder Entscheidung auch das nichtrichterliche Personal (z.B. Serviceeinheiten, Schreibkräfte, Kostenbeamte usw.) gebunden wird.

Der Bedarf zur Effizienzsteigerung durch den Einsatz von geeigneten (technischen) Werkzeugen in der Justiz ist damit mehr als offensichtlich, vereinzelt werden entsprechende Systeme in Deutschland bereits erprobt. Ein Beispiel hierfür ist der „Frankfurter Urteils-Konfigurator, elektronisch“ – kurz: „FRAUKE“ –, der bei der Urteilsbegründung in Flugverspätungsfällen zur Unterstützung zum Einsatz kommen soll. Aber auch in anderen Bereichen der Jurisprudenz hat die Automatisierung unlängst Einzug erhalten: So hilft etwa die „Frankfurter regelbasierte intelligente Dokumentenerstellungs-Assistent*in“ – kurz: „FRIDA“ – bei der Erstellung von Protokollen und Urteilen in Bußgeldsachen. Ein weiteres Beispiel ist der IBM-Massenverfahrensassistent, der seit November 2022 operativ am Oberlandesgericht Stuttgart zur effizienten Bearbeitung von Dieselabgasverfahren eingesetzt wird.⁷

¹ Prominentes Beispiel ist etwa die Flightright GmbH (www.flightright.de), hierneben sind aber auch zahlreiche weitere Anbieter am Markt tätig.

² Ein Anbieter, der bereits höchstrichterliche Urteile in Deutschland erstritten hat, ist die Financialright claims GmbH (www.financialright.com); siehe hierzu etwa BGH, Urt. v. 13.6.2022, Az. VIa ZR 418/21 = NJW 2022, 3350 ff.

³ Die ersten Verfahren angestrengt hat dabei die Lexfox GmbH mit ihrem Portal wenigermiete.de, das heute von der Conny GmbH fortgeführt wird.

⁴ BGH, Urt. v. 27.11.2019, Az. VIII ZR 285/18 = BGHZ 224, 89 ff.; bestätigt durch BGH, Urt. v. 8.4.2020, Az. VIII ZR 130/19 = NJW-RR 2020, 779 ff.; BGH, Urt. v. 19.1.2022, Az. VIII ZR 123/21 = NJW-RR 2022, 376 ff.

⁵ BGH, Urt. v. 18.1.2011, Az. X ZR 71/10, Rn. 24 ff. = NJW 2011, 2056, 2057 f.; ausführlich zur (auch: internationalen) Zuständigkeit des Gerichts bei Ansprüchen wegen Flugverspätungen *Lehmann/Wancke*, BRJ 2017, 107 ff.; zur Zuständigkeit bei Verspätungen von Teilflügen EuGH, Urt. v. 3.2.2022, Rs. C-20/21.

⁶ Presseinformation zur Jahrespressekonferenz des Amtsgerichts Frankfurt am Main vom 08.05.2020, S. 8, 22; Presseinformation zur Jahrespressekonferenz des Amtsgerichts Frankfurt am Main vom 19.03.2021, S. 12, 27.

⁷ Vgl. die Pressemitteilung von IBM, abrufbar unter <https://de.newsroom.ibm.com/2022-12-07-OLG-Stuttgart-pilotiert-IBM-Massenverfahrensassistenten-zur-Fallbearbeitung-in-Dieselabgasverfahren>.

Der nachfolgende Beitrag bietet einen interdisziplinären Überblick über die Herausforderungen für die Justiz, die aus dem massenhaften Anfall von Klagen erwachsen. Rechtlich orientieren sich die Ausführungen dabei fallstudienartig an den rechtlichen und tatsächlichen Gegebenheiten bei der Bearbeitung von Klagen wegen Flugverspätungen. Dabei werden zunächst die für das Anwendungsbeispiel der Flugverspätungsschädigung die einschlägigen materiell-rechtlichen Grundlagen (2.) und sodann das am Amtsgericht Frankfurt erprobte Entscheidungsunterstützungssystem „FRAUKE“ (3.) dargestellt. Im Anschluss an die Beleuchtung von „Risiken und Nebenwirkungen“ von KI im Entscheidungsfindungsprozess aus psychologischer Sicht (4.) und mögliche Vorbehalte der Philosophie (5.) werden abschließend einige Anforderungen an derartige Systeme aus rechtlicher Sicht formuliert (6.).

2. Materiell-rechtliche Ausgangslage für Verspätungsschädigungen

Die materiell-rechtliche Ausgangslage für Klagen, mit denen Fluggäste Verspätungsschädigungen von Luftfahrtunternehmen verlangen, ist durchaus überschaubar: Zentrale Grundlage ist in der Europäischen Union die Fluggastrechte-Verordnung (FGVO),⁸ nach der im Fall von Flugverspätungen pauschalierte Ersatzansprüche bestehen.⁹ Die Verordnung unterscheidet im Grundsatz drei verschiedene Leistungsstörungstatbestände, namentlich die Nichtbeförderung eines Fluggastes (Art. 4 FGVO), die Annullierung eines Fluges (Art. 5 FGVO) sowie eine erhebliche Abflugverspätung (Art. 6 FGVO). Luftfahrtunternehmen werden sodann – abhängig vom jeweils erfüllten Leistungsstörungstatbestand – zu Ausgleichszahlungen (Art. 7 FGVO), zu Erstattung oder Ersatzbeförderung (Art. 8 FGVO) sowie zur Erbringung von Betreuungsleistungen wie z.B. Mahlzeitengestellung oder Hotelunterbringung (Art. 9 FGVO) verpflichtet.

Insoweit sieht Art. 7 FGVO eine von der Länge der Flugstrecke abhängige, pauschalierte Ausgleichszahlung vor, die allerdings – jedenfalls nach dem Wortlaut der Verordnung – ausschließlich im Fall der Nichtbeförderung und der Annullierung eines Fluges zu zahlen ist. Für den Fall einer Abflugverspätung (Art. 6 FGVO) ist eine vergleichbare Entschädigung nicht vorgesehen, im Fall einer bloßen Ankunftsverspätung ohne Flugannullierung sieht die Verordnung grundsätzlich ebenfalls keine Entschädigungszahlungen vor. Nach der Rechtsprechung des EuGH in der grundlegenden „Sturgeon“-Entscheidung vom 19.11.2009 ist jedoch die Verordnung dahin auszulegen, dass der in Art. 7 FGVO vorgesehene Ausgleichsanspruch auch dann besteht, wenn Fluggäste wegen eines verspäteten Fluges einen Zeitverlust von drei Stunden oder mehr erleiden, mithin eine erhebliche Ankunftsverspätung vorliegt.¹⁰

Auf diesen Anspruch findet allerdings nach der Rechtsprechung des EuGH der in Art. 5 Abs. 3 FGVO geregelte Anspruchsausschluss entsprechende Anwendung, wonach ein ausführendes Luftfahrtunternehmen nicht verpflichtet ist, Ausgleichszahlungen nach Art. 7 FGVO zu leisten, wenn es nachweisen kann, dass die Annullierung auf außergewöhnliche Umstände zurückgeht, die sich auch dann nicht hätten vermeiden lassen, wenn alle zumutbaren Maßnahmen ergriffen worden wären.¹¹ Die außergewöhnlichen Umstände lassen sich dabei in unterschiedlichen Fallgruppen zusammenfassen; als taugliche Gründe für einen Anspruchsausschluss anerkannt sind etwa extreme Wetterereignisse,¹² Terrorangriffe¹³ und politische Instabilität,¹⁴ die Einordnung

⁸ Verordnung (EG) Nr. 261/2004 des Europäischen Parlaments und des Rates vom 11. Februar 2004 über eine gemeinsame Regelung für Ausgleichs- und Unterstützungsleistungen für Fluggäste im Fall der Nichtbeförderung und bei Annullierung oder großer Verspätung von Flügen und zur Aufhebung der Verordnung (EWG) Nr. 295/91, ABl. EU L 46, S. 1 ff.

⁹ Hierneben besteht mit dem Übereinkommen zur Vereinheitlichung bestimmter Vorschriften über die Beförderung im internationalen Luftverkehr vom 28. Mai 1999 – dem sog. Montrealer Übereinkommen – eine Grundlage für die individuelle Wiedergutmachung von Schäden; zur Abgrenzung EuGH, Urt. v. 10.1.2006, Rs. C-344/04, Rn. 34 ff.

¹⁰ EuGH, Urt. v. 19.11.2009, verb. Rs. C-402/07 und C-432/07, Rn. 40–66.

¹¹ EuGH, Urt. v. 19.11.2009, verb. Rs. C-402/07 und C-432/07, Rn. 67, 69.

¹² Vgl. für den Ausbruch des Vulkans Eyjafjallajökull im März 2010 EuGH, Urt. v. 31.1.2013, Rs. C-12/11.

¹³ STEINRÖTTER, in: BeckOGK, Art. 5 FGVO Rn. 84.

¹⁴ STEINRÖTTER, in: BeckOGK, Art. 5 FGVO Rn. 97.

von Streik ist im Einzelfall umstritten;¹⁵ technische Defekte hingegen fallen regelmäßig in die Risikosphäre des Luftfahrtunternehmens.¹⁶

Zusammengefasst setzt damit das Bestehen eines Entschädigungsanspruchs wegen einer Ankunftsverspätung einerseits eine Ankunftsverspätung am Zielort von mindestens drei Stunden voraus, und andererseits darf diese Verspätung nicht auf außergewöhnliche Umstände zurückgehen, die durch das Luftfahrtunternehmen nicht zumutbar hätten vermieden werden können.

Automatisierungslösungen – neben den bereits erwähnten LegalTech-Angeboten nicht zuletzt auch das so gleich vorgestellte System „FRAUKE“ – machen sich dabei die einfache Formalisierbarkeit und den vergleichsweise geringen Subsumtionsaufwand zunutze. Aus denselben Gründen wird in der Literatur mitunter gar erwogen, Entschädigungszahlungen wegen Flugverspätungen mithilfe selbstvollziehender Verträge („smart contracts“) zu automatisieren.¹⁷

3. Der technische Ansatz: Das Entscheidungsunterstützungssystem „FRAUKE“

Das System „FRAUKE“ unterstützt, wie eingangs erwähnt, die Richterschaft bei der Erstellung von Urteilen im Bereich der Fluggastrechte. Aus technischer Sicht besteht das System aus zwei Kernkomponenten: Einem KI-Modell zur automatischen Extraktion von relevanten Falldaten und einem Vorschlagsalgorithmus für zum Urteil passenden Textbausteine. Dazu ist von Seiten des Endnutzers eine Reihe von Angaben erforderlich, wozu insbesondere die Urteilsformel, der Tenor des Urteils, gehört (vgl. Abbildung 1). Auch die in der konkreten Akte maßgebliche Fallgruppe für einen möglichen Ausschluss ist auszuwählen¹⁸ – z.B. „Wetterphänomene“, wenn es im zu bearbeitenden Fall um eine Flugverspätung aufgrund schlechter Wetterbedingungen geht.¹⁹ Das System selbst schlägt kein Urteil vor, sondern liefert lediglich Textbausteinvorschläge passend zur richterlichen Entscheidung.

The image shows a screenshot of the 'FRAUKE' system's input mask. The main window is titled 'Akte auswählen' and contains the following elements:

- Header:** 'Akte auswählen' with a close button (X).
- Instructions:** 'Wählen Sie hier ihre Akte aus, zu der Sie Ihr Urteil verfassen möchten und beantworten Sie die unten stehenden Fragen, damit die passenden Textbausteine zugeordnet werden können.'
- Buttons:** 'Manueller input einfügen' (top right) and 'Urteil verlassen' (bottom right).
- Form Fields:**
 - 'Akte auswählen' (dropdown menu)
 - 'Standardfall' (dropdown menu)
 - 'Tenor - Hauptsache' with radio buttons for 'stattgegeben' and 'abgewiesen'.
 - 'Tenor - Kosten' with radio buttons for 'Klägerseite', 'Beklagenseite', and 'Andere Verteilung'.
 - 'Tenor - Vorläufige Vollstreckbarkeit' with radio buttons for 'Paragraph 709 ZPO' and 'Paragraph 711 ZPO'.
- Footer:** 'Stornieren' (bottom left).

Three callout boxes on the left side of the screenshot point to specific parts of the form:

- 'Auswahl des Aktenzeichens' points to the 'Akte auswählen' dropdown.
- 'Auswahl des Standardfalls' points to the 'Standardfall' dropdown.
- 'Definition des Tenors' points to the 'Tenor - Hauptsache' radio buttons.

A fourth callout box on the right side points to the 'Urteil verlassen' button:

- 'Weiter zum Verfassen des Urteils' points to the 'Urteil verlassen' button.

Abbildung 1: Eingabemaske des Systems FRAUKE

¹⁵ Gegen eine Beachtlichkeit „wilden“ Streiks etwa EuGH, Urt. v. 17.4.2018, Rs. C-195/17; für eine Beachtlichkeit des Streiks fluggesellschaftseigenen Personals BGH, Urt. v. 21.8.2012, Az. X ZR 138/11.

¹⁶ EuGH, Urt. v. 22.12.2008, Rs. C-549/07.

¹⁷ Ausführlich TAWAKOLI, ZRP 2020, 46 ff.; am Rande auch FRIES, NJW 2019, 901, 902.

¹⁸ In einer weiteren Projektausbaustufe ist geplant, dass das System auch die Fallgruppe automatisch aufgrund der Schriftsatzanalyse vorschlägt.

¹⁹ Siehe zu den maßgeblichen Ausschlussgründen für Ansprüche auf Verspätungsschädigungen soeben 2.

Nach Eingabe der Parameter wird in der momentanen Ausbaustufe des Systems die Klageschrift von einem sogenannten Named-Entity-Recognition Modell (NER-Modell) analysiert, welches auf der IBM Watson Technologie für natürliches Sprachverständnis basiert. Das Modell beruht auf einem vortrainierten Basismodell, welches mittels Labeling von entsprechenden Trainingsdokumenten auf die Domänensprache angepasst wird. Es ist dadurch in der Lage, die für den Fall relevanten Informationen, wie z.B. Start- und Zielflughafen, die Flugentfernung sowie den Streitwert, kontextabhängig zu extrahieren und die Semantik zu berücksichtigen. Auch komplizierte Zusammenhänge wie Beziehungen zwischen den Entitäten können antrainiert werden. Das Training erfolgt mittels des Programms IBM Watson Knowledge Studio (WKS), welches eine No-Code Benutzeroberfläche zur Verfügung stellt. Das gesamte Modell-Management über Training, Testung und Evaluierung der Modellperformance wird mit WKS durchgeführt, bis die gewünschte Genauigkeit erreicht ist. Das Programm ermöglicht durchgehend die Einsicht in die dem Modell zugrunde liegenden, gelabelten Daten (die sogenannte Ground-Truth), wodurch die Modell-Ergebnisse zu jeder Zeit nachvollzogen und überprüft werden können. Die Qualität der Trainingsdaten wird dabei im Vorfeld durch die Auswahl repräsentativer Dokumente sichergestellt. Die Bereitstellung des Modells zur Analyse erfolgt über eine REST-API Schnittstelle. Zeitgleich zur Textanalyse durch das KI-Modell erfolgt die regelbasierte Auswahl der passenden Textbausteine zum Urteil. Die Textbausteine basieren auf vergangenen Urteilen und wurden durch die Frankfurter Richterschaft zur Verfügung gestellt. Im weiteren Projektausbau und bei komplexeren Fallkategorien und Parameterkonstellationen kann der regelbasierte Algorithmus durch einen KI-basierten Vorschlag ersetzt werden. Dies ermöglicht eine höhere Komplexität der Fallkonstellationen und zudem einen „Human-In-The-Loop“-Ansatz, bei dem das System durch eine Bewertung der Textbausteinvorschläge durch den Nutzenden ständig mitlernt. Basierend auf den Textbausteinvorschlägen und den extrahierten Falldaten öffnet sich nach der Analyse das eigentliche Bearbeitungsfenster, welches in Abbildung 2 dargestellt ist. Die extrahierten Falldaten sind bereits in den Textbausteinen hinterlegt. Über das Zauberstab-Symbol lassen sich weitere, zum jeweiligen Abschnitt passende Textbausteine anzeigen. Dies ist insbesondere bei der Individualisierung der Entscheidungsgründe hilfreich. Dadurch wird das Auffinden von passenden Textbausteinen aus vorhandenen Quellen – wie zum Beispiel einer persönlichen Textbausteinsammlung – sehr erleichtert. Jegliches Feld ist außerdem frei editierbar und ermöglicht somit die Finalisierung des Urteils.

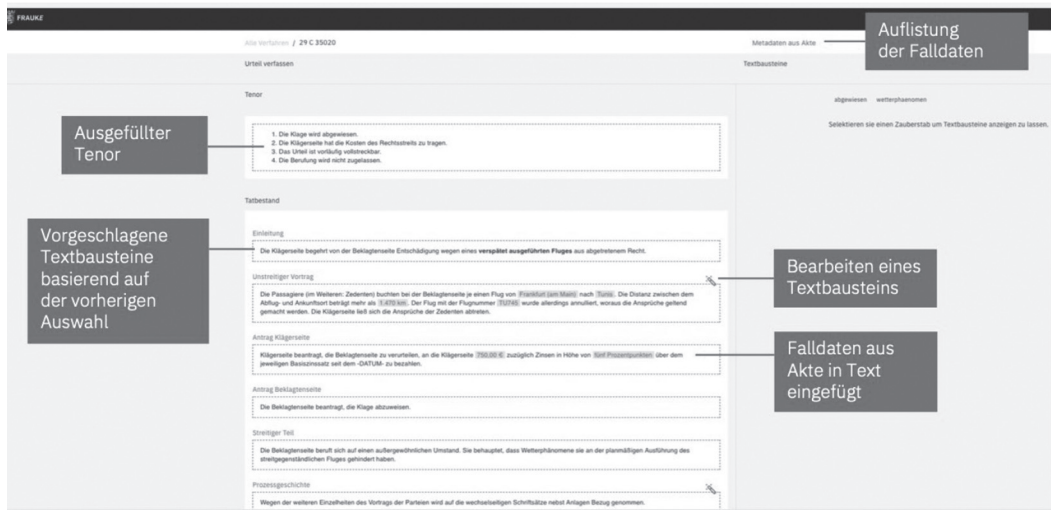


Abbildung 2: Bearbeitungsfenster von FRAUKE

Die Datengrundlage der Textbausteine kann sich aus verschiedensten Quellen speisen: Aus persönlichen Sammlungen einzelner Richterinnen und Richter, aus kollaborativ erstellten Sammlungen oder aus einer systematischen Analyse bereits erlassener Urteile. Dadurch kann auch eine Harmonisierung der Rechtsprechung erfolgen und schneller auf sich änderndes Recht reagiert werden, indem die Textbausteine und der Vorschlagsalgorithmus angepasst werden. Um die Qualität der Daten zu gewährleisten, wird dabei eng mit der Richterschaft zusammengearbeitet.

Basierend auf dieser Funktionalität kann das System dank offener Schnittstellen in einer weiteren Ausbaustufe an die bestehenden Verschriftungssysteme angeschlossen und in die Geschäftsprozesse integriert werden. Denkbar ist auch eine zyklische Integration neuer Daten zur kontinuierlichen Verbesserung des Systems.

4. Risiken und Nebenwirkungen aus psychologischer Sicht

Algorithmische Entscheidungsunterstützungssysteme wie FRAUKE haben zum Ziel, Teile von Entscheidungsprozessen zu automatisieren und somit Entscheiderinnen und Entscheider in ihrer Aufgabenerfüllung zu unterstützen. Unabhängig von den technischen Grundlagen des Systems (z.B. händisch programmiert oder auf Methoden des maschinellen Lernens basierend) gibt es eine schier unbegrenzte Zahl an Möglichkeiten, welche Aufgaben solche Systeme unterstützen könnten: Die Extraktion von wichtigen Informationen aus unübersichtlichem Datenmaterial, um die Informationen für Menschen greifbar zu machen, das Unterbreiten von Entscheidungsvorschlägen oder das selbstständige Treffen von Entscheidungen seien hier genannt.²⁰ Auch die Zusammenarbeit von Mensch und System unterliegt einer Vielzahl an Gestaltungsmöglichkeiten, die sich auf einem Kontinuum zwischen vollständiger menschlicher Kontrolle bis hin zur vollen Automatisierung von Prozessen erstrecken. Systeme könnten Menschen zusätzliche Informationen darbieten, System und Mensch könnten parallel an Aufgaben arbeiten oder Entscheidungen treffen, die am Ende zusammengeführt werden, Systeme könnten Entscheidungsalternativen anbieten, aus denen Menschen auswählen, oder Menschen könnten nur noch dazu da sein zu überprüfen, ob Systementscheidungen „in Ordnung“ sind.²¹ In anderen Worten: die Entwicklung und Implementierung algorithmischer Entscheidungsunterstützungssysteme unterliegt einer Vielzahl an Designentscheidungen, die weitreichende Implikationen haben können.

Designentscheidungen bei der Entwicklung und Implementierung von Systemen sollten – neben grundlegenden Fragen der Rechtskonformität oder der eigentlichen Güte und Nutzbarkeit des Systems – auch im Hinblick auf mögliche psychologische Risiken beim Einsatz getroffen werden. Auf alle möglichen Risiken einzugehen ist hier aus Platzgründen nicht möglich. Wir möchten daher nur auf einige Risiken beispielhaft eingehen: Ankereffekte, Über- und Untervertrauen sowie Hindsight Bias.

4.1. Ankereffekte

Systemoutputs können als sog. Anker wirken, die menschliche Entscheidungen in eine gewisse Richtung beeinflussen und teils wie Fremdbestimmung anmuten können. Ankereffekte wurden in der Rechtsprechung beispielsweise von English u.a. nachgewiesen, die zeigten, dass auch erfahrene Richterinnen und Richter allein durch die vorausgehende Präsentation von zufälligen Zahlen in der Bestimmung von Strafmaßen beeinflusst sind.²² Systemoutputs können eine ähnliche Wirkung haben. Ein Beispiel stellt der amerikanische COMPAS-Algorithmus dar. Dieser gibt die Gefahr für Rückfälligkeit bei Straftätern als Zahl an. Dieses Vorgehen kann einen Ankereffekt nach sich ziehen, bei dem sich Beteiligte selbst bei unrealistisch scheinenden Werten übermäßig an vorgegebenen Zahlen orientieren. Weil diese Outputs von einem System stammen, das

²⁰ PARASURAMAN/SHERIDAN/WICKENS, IEEE Transactions on Systems, Man, and Cybernetics – Part A: Systems and Humans, 30(3), S. 286 ff.

²¹ KABER/ENDSLEY, Theoretical Issues in Ergonomics Science, 5(2), S. 113 ff.

²² ENGLISH/MUSSWEILER/STRACK, Personality and Social Psychology Bulletin, 32(2), 188 ff.

eine dritte Partei entwickelt hat, besteht die Gefahr der Fremdbestimmung und Einflussnahme auf die Rechtsprechung.

Im Fall von FRAUKE könnte das Liefern von Begründungen durch das System, je nach seiner genauen Beschaffenheit und Implementierung, dazu verleiten, nicht oder weniger eigenständig über andere mögliche Urteilsbegründungen nachzudenken, als man es ohne Unterstützungssystem getan hätte. Insbesondere eine zu schematische Beschreibung von bestimmten Fallgruppen (z.B. „technischer Defekt“, „Streik“) in vorgefertigten Textbausteinen könnte Richterinnen und Richter dazu verleiten, die zugrunde liegenden Erwägungen im Einzelfall nicht ausreichend zu hinterfragen und Besonderheiten des Einzelfalles zu übergehen.

4.2. Über- und Untervertrauen

Ähnliche Effekte können beim Thema Übervertrauen erwartet werden. Übervertrauen bedeutet, dass Menschen sich zu stark auf Systeme verlassen, die ihrerseits jedoch fehlerbehaftet sein können. Übervertrauen wird beispielsweise auch im Entwurf eines KI-Gesetzes (AI Act) der Europäischen Kommission vom 21.4.2021²³ unter dem Namen „Automation Bias“ als Risiko hervorgehoben.²⁴ Das Risiko besteht darin, dass inkorrekte Systemoutputs unerkannt bleiben, wodurch menschliche Entscheidungen unter Zuhilfenahme eines Systems schlechter ausfallen, als sie es ohne das System wären. Außerdem können so zusätzliche Anstrengungen erforderlich werden, um mangelhafte Entscheidungen aufzudecken und deren Konsequenzen zu beheben. Übervertrauen untergräbt damit die erhoffte Risikominimierung durch eine menschliche Aufsicht von automatisierten Systemen („human oversight“), weil Systemoutputs nicht hinreichend auf ihre Korrektheit geprüft werden und Korrekturmöglichkeiten, die den Überwachenden zur Verfügung stehen, nicht in Anspruch genommen werden. Eine Gefahr von Übervertrauen besteht im Rahmen von FRAUKE beispielsweise, wenn Richterinnen und Richter ungeprüft Vorschläge für Textbausteine als eigene Argumentationslinie übernehmen. Besonders bei hohem Zeitdruck und bei insgesamt hoher Güte der Systemempfehlungen kann möglicherweise Übervertrauen entstehen. Im Falle von FRAUKE könnten so ggf. überholte Begründungen übernommen werden oder unschlüssige Argumentationslinien entstehen, die zu (ggf. unanfechtbaren) Fehlentscheidungen führen und sich in der Arbeit anderer Richterinnen und Richter verstetigen.

Doch auch Untervertrauen stellt ein Risiko dar. Untervertrauen meint, dass sich Menschen zu wenig auf Systeme und deren Outputs verlassen. Zunächst stellt sich die Frage, warum ein (womöglich kostspieliges) System überhaupt eingesetzt wird, wenn Menschen es nicht nutzen. Des Weiteren bleiben erhoffte Effizienz- und Qualitätssteigerungseffekte aus, wenn Menschen Systemoutputs nicht oder unangemessen nutzen. Zuletzt können durch Untervertrauen neue Fehler entstehen, wenn Menschen eigentlich korrekte Systemoutputs ignorieren oder durch ihren Eingriff verschlechtern. Im Beispiel von FRAUKE könnte Untervertrauen bedeuten, dass Richterinnen und Richter die Textbausteine des Systems ignorieren oder selbstständig so anpassen, dass am Ende kaum Zeit eingespart wird und Argumentationslinien sich möglicherweise sogar verschlechtern.

4.3. Hindsight Bias

Hindsight Bias²⁵ beschreibt die Tendenz, Ereignisse im Nachhinein vorhersehbarer einzuschätzen, als sie in der eigentlichen Situation waren. Zukünftige Systeme könnten in der Rechtsprechung so eingesetzt werden, dass Erklärungen generiert werden, die aufzeigen, welche Entscheidungen und Aktionen beispielsweise in einem autonomen Fahrzeug später zu einem Unfall führten. Die Präsentation solcher Informationen kann dazu führen, dass negative Konsequenzen als vermeidbarer eingestuft werden, als sie eigentlich waren und

²³ Vorschlag für eine Verordnung des Europäischen Parlaments und des Rates zur Festlegung harmonisierter Vorschriften für Künstliche Intelligenz (Gesetz über Künstliche Intelligenz) vom 21.4.2021, COM(2021) 206 final.

²⁴ Vgl. Art. 14 Abs. 4 lit. b) des Vorschlags.

²⁵ Z.B. ROESE/VOHS, Perspectives on Psychological Science, 7(5), S. 411 ff.

beispielsweise Sorgfaltspflichten mit höherer Wahrscheinlichkeit als verletzt eingestuft werden. Im Fall von Flugverspätungsklagen könnte dieser Effekt im Rahmen der Frage von Relevanz sein, ob die Folgen eines außergewöhnlichen Umstandes für die Fluggesellschaft mit zumutbarem Aufwand vermeidbar gewesen wären – mit der Folge, dass ein Haftungsausschluss unberechtigterweise nicht in Betracht gezogen wird. Im Beispiel von FRAUKE könnte etwa ein ohne vollständigen Kontext dargestellter Vorschlag für einen möglicherweise passenden Textbaustein dazu führen, dass eine Flugverspätung irrtümlich als vermeidbar angesehen wird.

4.4. Zwischenfazit

Insgesamt bietet die Implementierung algorithmischer Entscheidungsunterstützung in der Rechtsprechung Chancen und Risiken, von denen einige psychologischer Natur sind. Manche psychologische Risiken in der Rechtsprechung sind bereits ohne algorithmische Systeme bekannt, könnten aber durch diese verstärkt auftreten, während andere sich möglicherweise erst durch den Einsatz solcher Systeme ergeben. Das Wissen um solche Risiken sollte bereits bei Konzeption und Implementierung von Systemen in den Arbeitsalltag bedacht werden. FRAUKE stellt hier ein gutes Beispiel dar: Anstatt das System Entscheidungen generieren zu lassen, die den Richterinnen und Richtern direkt präsentiert werden, entschied man sich dazu, das System zur Produktion von Textblöcken einzusetzen. Damit wird beispielsweise Ankereffekten bei der Urteilsbildung vorgebeugt, auch wenn für die Auswahl von Begründungen weiterhin ein Ankereffekt bestehen bleiben könnte (z.B. wird von den vorgegebenen Begründungen kaum mehr abgewichen). Auch mit Hinblick auf Über- und Untervertrauen bleiben Risiken bestehen, die womöglich einer empirischen Überprüfung im Alltag von Richterinnen und Richtern bedürfen. Chancen einer solchen empirischen Überprüfung könnten die Verbesserung der Gestaltung des Systems und der Gestaltung der Zusammenarbeit von Mensch und System sein.

5. Was sagt die Philosophie: Sollte der Einsatz von KI in der Justiz zulässig sein?

Wie eingangs erwähnt, sind Richterinnen und Richter menschliche Wesen aus Fleisch und Blut; als solche sind sie Wesen, die normative Gründe verstehen und auf sie reagieren können.²⁶ Sie müssen von Gesetzes wegen sogar solche Wesen sein (dazu unten 6.1.).

Nun mutieren Richterinnen und Richter, wenn sie sich eines Urteilsconfigurators bedienen, allein dadurch nicht zu „Subsumtionsautomaten“: Sie sind weiterhin diejenigen, die die Urteile fällen, und sie sind auch weiterhin diejenigen, die selbige begründen (müssen). Letztgenanntes tun sie mithilfe von Urteilsconfiguratoren wie FRAUKE nur deutlich schneller – und möglicherweise sogar besser. Denn ohne FRAUKE arbeiten sie, statt mit den Textbausteinen dieses Configurators, häufig mit ihren eigenen, die als solche aber keiner externen Kontrolle unterliegen und folglich nur so gut sind, wie die einzelnen Richterinnen und Richter sie pflegen.²⁷ Es spricht mithin einiges *für* den Einsatz von FRAUKE, jedenfalls dann, wenn FRAUKE bestimmte Qualitätsstandards erfüllt. Zu den betreffenden Standards gehört etwa, (i) dass FRAUKE mit juristisch korrekten Textbausteinen bestückt wird, (ii) dass die verwendeten Textbausteine aus zuverlässigen Quellen stammen, die prinzipiell auch für die Nutzer von FRAUKE zugänglich sind und (iii) dass FRAUKE, wo nötig, kontinuierlich und unverzüglich aktualisiert wird (z.B. Anpassung an geänderter höchstrichterliche Rechtsprechung). Darüber hinaus sollten mit dem Configurator erstellte Urteilsbegründungen solchen, die Richterinnen und Richter manuell erstellen, in nichts nachstehen; z.B. in puncto Klarheit, Ausführlichkeit, Nachvollziehbarkeit etc. Sind diese Qualitätsstandards de facto *nicht* oder nur *teilweise* erfüllt, sollten wir skeptischer sein. Es gilt dann zu prüfen, ob FRAUKE die geforderten Qualitätsstandards überhaupt erfüllen kann und, wenn ja, wie

²⁶ PARFIT, On What Matters, Kap. 1.1: „We are animals that can both understand and respond to reasons.“

²⁷ Einer externen Kontrolle unterliegen allenfalls die mit ihnen begründeten Urteile, wenn sie mit Rechtsmitteln überprüft werden. Dieser Weg steht jedoch in manchen Fällen nicht offen. Dies gilt regelmäßig für den hier diskutierten Fall von Klagen wegen Flugverspätungen, deren Streitwert häufig die Berufungssumme nicht erreicht.

zu erreichen ist, dass sie tatsächlich erfüllt werden. Erst wenn entsprechende Änderungen implementiert sind, spricht dann tatsächlich einiges für den Einsatz.

Doch angenommen, wir können die Einhaltung bestimmter Qualitätsstandards als gesichert ansehen. Urteils-konfiguratoren scheinen dann also zunächst einmal einen positiven Beitrag zur Entlastung von Richterinnen und Richtern in Massenverfahren zu leisten. Allerdings bergen sie genau dadurch aber auch gewisse Gefahren.

5.1. Richterliche Reflexion und die Gefahr der *ex post*-Rationalisierung

Zunächst einmal reflektieren Richterinnen und Richter normalerweise ihre Urteile, wenn und während sie sie zu begründen versuchen – in Abhängigkeit u.a. von der Komplexität der Fälle, mit denen sie es zu tun haben, von dem, was für die Beteiligten auf dem Spiel steht, oder auch von ihren vorgängigen Erfahrungen mit vergleichbaren Fällen. Sie vergegenwärtigen sich die einschlägigen normativen Gründe für mitunter intuitiv gewordene Expertenentscheidungen; sie prüfen, ob die Gründe ihre Urteile stützen; und stellen sie fest, dass die Gründe ihre Urteile *nicht* stützen, revidieren sie die Gründe, ggf. sogar das komplette Urteil mitsamt seinem Ergebnis. Genau dieses Reflektieren, dieses Er- und Abwägen von Gründen, drohen Urteilskonfiguratoren zu untergraben. Sie drohen das gewissenhafte Ringen um Begründungen und das sorgfältige Abfassen derselben durch ein quasi-mechanisches Aneinanderreihen von Textbausteinen zu ersetzen – und damit auch die Urteile nicht mehr zu *fundieren*, sondern nur noch *ex post* zu *rationalisieren*. Hierbei können u.a. die oben bereits erwähnten psychologischen Nebenwirkungen (Ankereffekte, Übervertrauen, Hindsight Bias) eine wichtige Rolle spielen; es besteht die Gefahr, dass bestehende Biases verstärkt, Hintergrundannahmen nicht sorgfältig geprüft und Entscheidungen, von denen Richterinnen und Richter nicht vollends überzeugt sind, stark gemacht werden, weil vermeintlich gute Gründe vom System vorgeschlagen werden.

Nun mag man argumentieren, dass diese Angst übertrieben oder gar unbegründet ist. Immerhin fällen Richterinnen und Richter ihre Urteile zunächst selbst und nutzen FRAUKE dann lediglich bei der Abfassung ihrer Begründungen. Insofern ist der Fall analog zu dem im Rechtssystem ausdrücklich vorgesehenen Fall, dass ein Urteil direkt vom Richterstuhl gefällt, die Begründung aber erst später geliefert wird.²⁸ Bei solchen Stuhlurteilen steht, sogar stärker als beim Einsatz von FRAUKE, das Urteil bereits fest, wenn die Begründung formuliert wird. Handelt es sich nur deshalb bei der Begründung bereits um eine *Ex-post*-Rationalisierung des Urteils? Sicherlich gilt das nicht immer. Aber die Gefahr, dass das Urteil eben doch *ex post* rationalisiert wird, dürfte umso größer sein, je mehr Zeitdruck herrscht, je mehr Begründung man bereits ohne eigene Deliberation präsentiert bekommt und je weniger standardisiert und überschaubar die Fälle sind.²⁹ In Auseinandersetzung mit standardisierten Klagen wegen verspäteter oder annullierter Flüge mag diese Gefahr noch gering sein; aber schon etwa bei Wohnraum betreffenden Räumungsklagen (welche ebenfalls vergleichsweise stark standardisiert sind) wird eine Vielzahl von unterschiedlichen Faktoren relevant, die sorgfältig gegeneinander abgewogen werden müssen, damit das gefällte Urteil als gerecht gelten darf.

Noch ist der Einsatz von Urteils-konfiguratoren sehr begrenzt. Aber sobald ein System wie FRAUKE sich einmal bewährt hat, wird es sicher schnell Nachfragen für weitere Anwendungsfelder geben. Die Frage, die wir uns dann stellen müssen, ist, unter genau welchen Bedingungen wir automatisierte Urteilsbegründungen zulassen wollen, wenn hier die Gefahr besteht, dass Richterinnen und Richter doch zu dem mutieren, was sie eben nicht sein dürfen: Subsumtionsautomaten, deren eigene Überprüfung und Begründung von Urteilen durch den Einsatz von Urteils-konfiguratoren ungewollt untergraben wird.

²⁸ Aus der Perspektive des deutschen Zivilprozessrechts etwa stellt das „Stuhlurteil“ den vom Gesetz zuerst genannten „Idealfall“ dar, die Praxis hingegen verfährt regelmäßig anders (vgl. ELZER, in: BeckOK ZPO, § 310 Rn. 10).

²⁹ Bei der Entscheidung der RichterIn, ob sie ein Stuhlurteil fällt oder einen gesonderten Verkündungstermin bestimmt, ist daher regelmäßig der Vorteil einer schnellen Entscheidung gegen die Gefahr abzuwägen, im Einzelfall wichtige Gesichtspunkte übersehen zu haben (MUSIELAK, in: Musielak/Voit, ZPO, § 310 Rn. 4).

5.2. Die Notwendigkeit empirischer Untersuchungen

Um zu klären, wie groß die Gefahr ist, könnte es sich lohnen, einmal empirisch zu untersuchen, wie genau Richterinnen und Richter tatsächlich mit Urteilskonfiguratoren für komplexere Fälle arbeiten würden. Lassen sie sich womöglich unterschiedliche Urteile begründen, explorieren und reflektieren diese im Anschluss, um dann den Konfigurator erneut zu starten? In dem Fall könnte die Software eine Reflexionshilfe sein, die nicht die eigene Reflexion ersetzt. Aber ob Richterinnen und Richter so arbeiten würden und, wenn ja, ob sie dann noch Zeit gegenüber dem „klassischen“ Verfassen einer Urteilsbegründung einsparen würden, das sind empirische Fragen, die wir hier nicht beantworten können.

Allgemein bleibt aber festzuhalten, dass wir nicht nur die oben bereits angerissenen Anforderungen an die Qualität der Urteilsbegründung als Endprodukt stellen (sollten), sondern auch Anforderungen an den Prozess bzw. die Genese dieser Begründung bestehen, die Urteilskonfiguratoren nicht untergraben sollten. In bestimmten (einfachen) Fällen mag dies freilich weniger relevant sein als in anderen (komplexeren).

5.3. Zwischenfazit

Unter der Voraussetzung, dass FRAUKE die nötigen Qualitätsstandards erfüllt, spricht einiges für den Einsatz und wenig dagegen. Es spricht deshalb wenig dagegen, weil die Gefahren, die Urteilskonfiguratoren bergen, bei FRAUKE als vernachlässigbar gelten können. Die Fälle von Flugverspätungen und -annullierungen sind hoch standardisiert, und der Streitwert ist im Regelfall auf wenige hundert Euro begrenzt. Doch wo die Gefahren von *Ex-post*-Rationalisierungen größer werden, sollte auf den Einsatz von Konfiguratoren verzichtet werden.

Wo nun allgemein die Linie zu ziehen ist, die die Bereiche, in denen Urteilskonfiguratoren sich eignen, von solchen trennt, in denen darauf verzichtet werden sollte, bedarf neben einer genaueren Abwägung von Kosten und Nutzen auch Forschung zum Verhalten im Umgang mit Konfiguratoren. Solange keine empirischen Daten hierzu vorliegen, ließe sich etwa folgende Daumenregel erwägen: Der Einsatz von Urteilskonfiguratoren ist in der Regel in solchen Fällen angemessen, in denen erfahrene Richterinnen und Richter eine Entscheidung im Wege eines Stuhlurteils in der Sache als unbedenklich ansehen würden.³⁰ Der Vorteil einer solchen Regel bestünde darin, dass für den Einsatz von KI-Systemen auf vorhandenem Expertenwissen aufgebaut werden kann und nicht zwangsläufig eine neue, explizit ausformulierte Regulierung erforderlich wird.

6. Grundlegende rechtliche Anforderungen an KI in der Justiz

Nachfolgend können aus der Perspektive der Rechtswissenschaft lediglich einige grundlegenden Anforderungen genannt werden, die an den Einsatz von KI-Systemen in der Justiz zu stellen sind.³¹ Interessanterweise sind die aus rechtlicher Sicht anzutreffenden Bedenken häufig solche, die hier bereits aus psychologischem Blickwinkel betrachtet wurden:³² So wird etwa eine zunehmende Fremdbestimmung aufgrund eines „Automation Bias“ durch die unreflektierte Übernahme von KI-Empfehlungen befürchtet,³³ auch der bereits angesprochene begrenzte Nutzen menschlicher Aufsicht für den Fall einer (zu) zögerlichen Nutzung entsprechender Befugnisse wird als mögliche Schwachstelle des Konzepts identifiziert.³⁴ Auch die Forderung aus der Philosophie, KI-Systeme nur bei Erfüllung von qualitativen Mindeststandards und im Einklang mit

³⁰ Die hier in Bezug genommene „Unbedenklichkeit“ eines Stuhlurteils bezieht sich ausdrücklich nicht auf die rechtliche Zulässigkeit, sondern auf die Frage, ob im Einzelfall tatsächlich die Gefahr einer Fehlentscheidung so gering ist, dass der Vorteil einer zügigen Entscheidung insgesamt überwiegt.

³¹ Monographisch hierzu NINK, *Justiz und Algorithmen*, 2021; ROLLBERG, *Algorithmen in der Justiz*, 2020.

³² Ausführlich zu ungewollten rechtsfremden Einflüssen NINK, *Justiz und Algorithmen*, S. 45 ff.

³³ KEVEKORDES, in: MMR-Hdb., Teil 29.1 Rn. 44.

³⁴ BORUTTA/HAAG/HOFFMANN/KEVEKORDES/VOGT, MMR-Aktuell 2020, 427809.

normativen Werten einzusetzen, findet sich in der rechtswissenschaftlichen Diskussion – nicht zuletzt in den Anforderungen des geplanten KI-Gesetzes – wieder.

6.1. Verfassungsrecht: Menschliche und unabhängige Richterinnen und Richter

Das deutsche Verfassungsrecht stellt zwei zentrale Anforderungen auf, die im hiesigen Kontext von besonderem Interesse sind: Die rechtsprechende Gewalt ist, so legt es Art. 92 GG fest, den „Richtern“ anvertraut. Darüber hinaus bestimmt Art. 97 Abs. 1 GG, dass die Richter unabhängig und nur dem Gesetze unterworfen sind. Einigkeit besteht zunächst darin, dass „Richter“ im Sinne von Art. 92 GG – wie bereits einleitend erwähnt – nur menschliche Richter sein können.³⁵ Die Idee eines „Robo-Richters“ oder eines „Subsumtionsautomaten“ ist daher mit dem geltenden Verfassungsrecht nicht vereinbar. Hieraus folgt zugleich eine Minimalanforderung an den Einsatz von KI-Systemen in der Justiz: Die letzte Entscheidungshoheit muss in jedem Fall – gleichviel, wozu KI-Systeme tatsächlich fähig wären³⁶ – einem Menschen überlassen bleiben. Indes: Das deutsche Zivilprozessrecht kennt mit der maschinellen Durchführung des Mahnverfahrens (vgl. § 689 Abs. 1 S. 2 ZPO) schon heute ein stark technologiegetriebenes Verfahren; der hohe Grad an Automatisierung wird hier allerdings durch die Möglichkeit kompensiert, eine Entscheidung durch einen menschlichen Richter zu erzwingen. Werden KI-Systeme hingegen nicht in einem eigens hierfür geschaffenen, sondern im üblichen Klageverfahren eingesetzt, muss es bei der Letztentscheidungsbefugnis der Richterin und des Richters verbleiben; dabei ist allerdings unklar, ob und wie sich etwa Ankereffekte oder Übervertrauen rechtlich auswirken. Ähnlich wie beim datenschutzrechtlichen Verbot automatisierter Entscheidungsfindung (Art. 22 DSGVO) ist zu fordern, dass die Entscheidungsbefugnis des Menschen nicht bloß auf dem Papier besteht.³⁷ FRAUKE jedenfalls wahrt die Entscheidungshoheit der Richterinnen und Richter auf den ersten Blick durch eine naheliegende Designentscheidung: Das Ergebnis des zu konfigurierenden Urteils wird vorab festgelegt, die (für ungewollte Einflüsse besonders anfällige) Entscheidungsbegründung ist zeitlich nachgelagert.

Die zweite, im Zusammenhang mit Entscheidungsunterstützungssystemen vieldiskutierte Frage ist, wie die Unabhängigkeit der Richterinnen und Richter gewährleistet werden kann. Von Bedeutung sind insoweit zwei zentrale Aspekte: Einerseits sind die Entscheidenden von einer unmittelbaren Beeinflussung der Sachentscheidung freizuhalten, andererseits sind auch faktische Zwänge zu vermeiden, die sich auf die freie Entscheidungsfindung negativ auswirken können. Während eine Beeinflussung der Sachentscheidung durch FRAUKE in der derzeitigen Ausgestaltung des Systems eher fernliegen dürfte, ist insbesondere angesichts der hohen Fallzahlen zu beachten, dass Richterinnen und Richter nicht – auch nicht faktisch – zur Nutzung des Systems verpflichtet werden dürfen.³⁸ Ein faktischer Nutzungszwang kann dabei schon dadurch entstehen, dass das anfallende Arbeitspensum ohne die Nutzung des Systems nicht bzw. mit einem erhöhten Rechtfertigungsdruck bewältigt werden kann.³⁹ So ist beim Vorschlag von Textbausteinen aus kollaborativ erstellten Sammlungen sicherzustellen, dass diese hinreichend repräsentativ sind und nicht – wengleich unbewusst und ungewollt – eine umstrittene Rechtsmeinung gegenüber einer beachtlichen Gegenauffassung bevorzugen. Vor diesem Hintergrund bedarf es geeigneter Mechanismen, um eine hinreichende Diversität in den genutzten Daten sicherzustellen. Insoweit ist es begrüßenswert, dass im Fall von FRAUKE zunächst die Bearbeitung einfach gelagerter Fallgruppen unterstützt wird.

³⁵ NINK, *Justiz und Algorithmen*, S. 261 ff.; ROLLBERG, *Algorithmen in der Justiz*, S. 88 f.

³⁶ Deutlich etwa FRIES, *NJW-Editorial vom 9. Juni 2022*: „Frauke will kein Urteilsroboter sein, ist aber nicht weit davon entfernt. Sie könnte, wenn man sie ließe.“

³⁷ Vgl. MARTINI, in: Paal/Pauly, *DS-GVO*, Art. 22 Rn. 19.

³⁸ Zulässig ist indes die Bereitstellung zur freiwilligen Nutzung, vgl. ROLLBERG, *Algorithmen in der Justiz*, S. 125 f.

³⁹ Siehe hierzu NINK, *Justiz und Algorithmen*, S. 297.

6.2. Anforderungen an die Pflicht zur Urteilsbegründung

Ein weiterer Aspekt betrifft die Frage, ob eine KI-assistierte oder gar vollständig KI-generierte Urteilsbegründung den rechtlichen Mindestanforderungen an eine Urteilsbegründung genügt. Im deutschen Zivilprozessrecht ist hierzu geregelt, dass die Entscheidungsgründe eines Urteils grundsätzlich eine kurze Zusammenfassung der Erwägungen enthalten muss, auf denen die Entscheidung in tatsächlicher und rechtlicher Hinsicht beruht (§ 313 Abs. 3 ZPO); fehlt eine entsprechende Begründung, ist das Urteil (formell) rechtsfehlerhaft. Der Zweck der Urteilsbegründung, dass die Parteien das Urteil nachvollziehen können sollen,⁴⁰ gebietet naturgemäß eine konkrete Bezugnahme auf den jeweiligen Einzelfall. Hiermit ist es grundsätzlich vereinbar, wenn zur Begründung auf Textbausteine zurückgegriffen wird, um vom Einzelfall losgelöste Urteilspassagen zu erstellen, die sodann durch Inbezugnahme der Fakten des Einzelfalls subsumiert werden; eine Begründung, die ausschließlich aus Textbausteinen besteht, genügt diesen Anforderungen demgegenüber nicht.⁴¹

Urteilskonfiguratoren wie FRAUKE werfen damit die Frage auf, in welchem Umfang eine Automatisierung der Entscheidungsbegründung zulässig ist. Man mag geneigt sein, eine Parallele zu dem bereits höchstrichterlich entschiedenen Fall zu ziehen, dass Vertragsentwürfe mithilfe eines „Vertragsgenerators“ vollautomatisiert generiert werden. Der BGH hat hierzu entschieden, dass die Erstellung eines Vertragsentwurfs mithilfe eines digitalen Rechtsdokumentengenerators keine – einem Rechtsanwalt vorbehaltene – Rechtsdienstleistung i.S.v. § 2 Abs. 1 RDG darstellt, wenn im Anschluss an die Beantwortung eines Fragebogens lediglich standardisierte Vertragsklauseln abgerufen werden.⁴² Zur Begründung hat der BGH angeführt, dass das Bereitstellen eines softwarebasierten Vertragsentwurfs keine auf einen konkreten Sachverhalt zugeschnittene Tätigkeit darstellt.⁴³

Wäre diese Argumentation auf eine KI-generierte Begründung einer Gerichtsentscheidung übertragbar, müsste die insoweit maßgebliche Kernanforderung – die Begründung der Entscheidung im konkreten Einzelfall – als nicht erfüllt angesehen werden. Der Unterschied zum „Vertragsgenerator“ besteht allerdings darin, dass dieser ohne jede Beaufsichtigung durch eine rechtskundige Person arbeitet, während FRAUKE durch eine Richterin oder einen Richter bedient wird. Überdies bietet FRAUKE – wie dargestellt – die Möglichkeit, die Entscheidungsgründe manuell zu bearbeiten, was ebenfalls für eine einzelfallbezogene Begründung spricht.

6.3. Ausblick: Der Entwurf eines KI-Gesetzes

Schließlich soll ein kurzer Blick auf die europäische Perspektive des KI-Einsatzes im Richterzimmer geworfen werden. Hier ist festzustellen, dass der Einsatz von KI-Systemen durch die Justiz vom Entwurf des KI-Gesetzes⁴⁴ als Hochrisiko-KI eingestuft wird.⁴⁵ Daraus folgt zweierlei: Zunächst lautet die wichtige Nachricht, dass derartige Systeme nicht etwa gänzlich unzulässig sind, sondern grundsätzlich eingesetzt werden dürfen. Darüber hinaus bedeutet dies aber auch, dass der Einsatz von KI-Systemen künftig den strengen Anforderungen des KI-Gesetzes unterworfen sein wird.

Der Einsatz eines Urteilskonfigurators wie FRAUKE mag dann zwar nach den Vorgaben des nationalen Verfassungs- und Prozessrechts ohne gesonderte gesetzliche Grundlage zulässig sein; nach unionsrechtlichen Vorgaben indes müssen derartige Systeme künftig die Anforderungen an Hochrisiko-KI-Systeme erfüllen, um sie auch auf der Grundlage des KI-Gesetzes rechtskonform einsetzen zu dürfen. Hierzu zählen insbesondere die Sicherstellung einer hinreichenden Datenqualität (bei FRAUKE etwa durch Auswahl geeigneter

⁴⁰ BVerwG, Urt. v. 13.6.2022, Az. VIa ZR 418/21 Beschl. v. 5.6.1998, Az. 9 B 412-98 = NJW 1998, 3290; ELZER, in: BeckOK ZPO, § 313 Rn. 221.

⁴¹ OLG Celle, Beschl. v. 30.11.1989, Az. 12 UF 176/89 = FamRZ 1990, 419.

⁴² BGH, Urt. v. 9.9.2021, Az. I ZR 113/20 = NJW 2021, 3125.

⁴³ BGH, Urt. v. 9.9.2021, Az. I ZR 113/20, Rn. 31 ff. = NJW 2021, 3125, 3128.

⁴⁴ Siehe den Fundstellennachweis oben Fn. 23.

⁴⁵ Vgl. Anhang III Ziffer 8 des Entwurfs vom 21.4.2021.

Textbausteine) sowie die Ergreifung von Maßnahmen zur Vermeidung eines möglichen Automation Bias, die insbesondere in der Bereitstellung aussagekräftiger Informationen der Nutzerinnen und Nutzer über die vom Systemeinsatz ausgehenden Risiken bestehen können.

7. Zusammenfassung

Der vorliegende Beitrag hat gezeigt, dass mit der automatisierten Durchsetzung von Fluggastrechten eine überaus praxisrelevante Fallgestaltung besteht, bei der ein Rückgriff auf KI-Systeme in vielerlei Hinsicht vorteilhaft erscheint: Die Arbeitslast ist zahlenmäßig gewaltig, das Automatisierungspotenzial aufgrund des überschaubaren rechtlichen Rahmens immens, die Konsequenzen von Fehlentscheidungen gering. Kurzum: Die zahlreichen Risiken, die mit dem Einsatz von KI-Systemen in der Justiz assoziiert werden, scheinen hinnehmbar und dürften durch die offensichtlichen Vorteile aufgewogen werden.

Zugleich zeigt sich, dass dieses zurückhaltend-positive Fazit den besonderen Umständen der hier untersuchten Fallgestaltung geschuldet ist; auf andere Anwendungsfälle lässt sich dies nicht, jedenfalls nicht unbesehen übertragen. Stets sind die hier aufgezeigten Risiken gegen den Nutzen eines Systems abzuwägen; dabei ist auch die Frage aufgeworfen, ob diese Abwägungsentscheidung – wie bislang – von den Gerichten autonom entschieden werden sollte, oder ob es womöglich – über das geplante KI-Gesetz hinaus – gesetzlicher Leitplanken für den KI-Einsatz in der Justiz bedarf.

Eine Kernanforderung jeglichen Einsatzes von KI-Systemen in der Justiz muss – das hat dieser Beitrag ebenfalls hervorgebracht – die Gewährleistung einer effektiven Evaluation mit dem Ziel einer reflektierenden Selbstkontrolle sein. Die Unterminierung einer unabhängigen, durch menschliche Richterinnen und Richter ausgeübten Justiz durch den Einsatz von Künstlicher Intelligenz geschähe sehr wahrscheinlich schleichend; dem mittels geeigneter Mechanismen vorzubeugen, ist eine der zentralen Herausforderungen, die beim KI-Einsatz in der Justiz zu bewältigen sein wird. Dabei schwingt freilich stets die Frage mit, ob der KI-Einsatz in der Sache wünschenswert oder lediglich zulässiger „Notnagel“ ist; möglicherweise verdienen in bestimmten Konstellationen – nicht zuletzt in den hier untersuchten Fällen von Flugverspätungsschädigungen – andere Konzepte als ein „technologisches Wettrüsten“ zwischen Justiz und Rechtsdienstleistern den Vorzug, um Fluggästen schneller und effektiver zu ihrem Recht zu verhelfen.

8. Literatur

- BORUTTA, YANNIK/HAAAG, MATTHIAS/HOFFMANN, HANNA/KEVEKORDES, JOHANNES/VOGT, VERENA, Strategiepapiere der EU-Kommission zur Künstlichen Intelligenz (KI), MMR-Aktuell 2020, 427809.
- ELZER, OLIVER, Kommentierung v. §§ 310, 313 ZPO, in: Vorwerk/Wolf (Hrsg.), Beck'scher Online-Kommentar ZPO, 46. Edition, Stand: 46. Edition (1.9.2022).
- ENGLISH, BIRTE/MUSSWEILER, THOMAS/STRACK, FRITZ, Playing Dice With Criminal Sentences: The Influence of Irrelevant Anchors on Experts' Judicial Decision Making, Personality and Social Psychology Bulletin, 32(2), 2016, S. 188–200, abrufbar unter: <https://doi.org/10.1177/0146167205282152>.
- FRIES, MARTIN, Frauke richtet, NJW-Editorial vom 9. Juni 2022, abrufbar unter: <https://rsw.beck.de/aktuell/daily/magazin/detail/frauke-richtet>.
- FRIES, MARTIN, Schadensersatz ex machina, Neue Juristische Wochenschrift (NJW) 2019, S. 901–905.
- HAUSMANN, LUDWIG, Europäische Fluggastrechte im Fall der Nichtbeförderung und bei Annullierung und großer Verspätung von Flügen, Jena 2012 (zit.: Fluggastrechte).
- HOEREN, THOMAS/SIEBER, ULRICH/HOLZSNAGEL, BERND (Hrsg.), Handbuch Multimediarecht, Stand: 58. Ergänzungslieferung (März 2022), München 2022 (zit.: *Bearbeiter*, in: MMR-Hdb.).
- KABER, DAVID B./ENDSLEY, MICA R., The effects of level of automation and adaptive automation on human performance, situation awareness and workload in a dynamic control task, Theoretical Issues in Ergonomics Science, 5(2), 2007, S. 113–153, abrufbar unter: <https://doi.org/10.1080/1463922021000054335>.

LEHMANN, MATTHIAS/WANCKE, ANSELM-LEANDER, Annullierte oder verspätete Flüge: Die Frage des Gerichtsstands, Bonner Rechtsjournal (BRJ) 02/2017, S. 107–113.

MARTINI, MARIO, Kommentierung v. Art. 22 DSGVO, in: Paal/Pauly (Hrsg.), DS-GVO/BDSG, 3. Aufl., München 2021.

MUSIELAK, HANS-JOACHIM, Kommentierung v. § 310 ZPO, in: Rauscher/Krüger (Hrsg.), Münchener Kommentar zur ZPO, 6. Aufl., München 2020.

NINK, DAVID, Justiz und Algorithmen, Berlin 2021.

PARASURAMAN, RAJA/SHERIDAN, TOM B./WICKENS, CHRISTOPHER D., A model for types and levels of human interaction with automation, IEEE Transactions on Systems, Man, and Cybernetics – Part A: Systems and Humans, 30(3), 2000, S. 286–297, abrufbar unter: <https://doi.org/10.1109/3468.844354>.

PARFIT, DEREK, On What Matters, Oxford, 2011.

ROESE, NEAL J./VOHS, KATHLEEN D., Hindsight Bias, Perspectives on Psychological Science, 7(5), 2012, S. 411–426, abrufbar unter: <https://doi.org/10.1177/1745691612454303>.

ROLLBERG, CHRISTOPH, Algorithmen in der Justiz, Baden-Baden 2020.

STEINRÖTTER, BJÖRN, Kommentierung v. Art. 1–19 Fluggastrechte-Verordnung, in: Gsell/Krüger/Lorenz/Reymann (Hrsg.), Beck'scher Online-Großkommentar (BeckOGK) Zivilrecht, Stand: 1.11.2022.

TAVAKOLI, ANUSCH ALEXANDER, Automatische Fluggast-Entschädigung durch smart contracts, Zeitschrift für Rechtspolitik (ZRP) 2020, S. 46–49.