



**DIGITALISIERUNG
ÖGD BW**

White Paper

Proof-of-Concept (POC): KI im Begutachtungswesen

calliora GmbH

Luise-Ullrich-Straße 20
80636 München, Deutschland

Vertreten durch: Dr. Frederik Träuble, Jonas Wildberger
Registergericht: Amtsgericht München, HRB 308903

Telefon: +49 8924 4154380
E-Mail: contact@calliora.com
Internet: www.calliora.com

I Einführung

Amtsärztinnen und Amtsärzte in baden-württembergischen Gesundheitsämtern erstellen täglich Gutachten von erheblicher Tragweite, von Dienstfähigkeitsbeurteilungen über Beihilfeentscheidungen bis hin zu Gutachten, auf deren Basis ggf. freiheitsentziehende Maßnahmen angeordnet werden. Ein einzelnes Gutachten beansprucht typischerweise mehrere Stunden, wobei ein großer Teil der Zeit nicht auf die eigentliche medizinische Beurteilung entfällt, sondern auf das Sichten, Zusammenfassen und Abgleichen von Befunden. Der hohe Bedarf an Gutachten und der zunehmende Fachkräftemangel erfordern neue Lösungen, damit auch zukünftig genügend Raum für die eigentliche ärztliche Kernaufgabe bleibt.

Deshalb hat die calliora GmbH im Auftrag des Ministeriums für Soziales, Gesundheit und Integration Baden-Württemberg einen KI-gestützten Assistenten für ein Proof of Concept (PoC) entwickelt, der gezielt diese zeitintensiven, repetitiven Teilaufgaben übernimmt. Die Erprobung erfolgte durchgehend mit synthetischen Daten, die in Zusammenarbeit mit Ärztinnen und Ärzten des Gesundheitsamts Reutlingen erstellt wurden, um Datenschutzkonformität bei gleichzeitig realitätsnaher Evaluation sicherzustellen.

Dabei gilt eine klare Abgrenzung: Das System ist ein Arbeitsbeschleuniger und eine Formulierungshilfe, kein automatisierter Entscheider. Die fachliche und rechtliche Verantwortung für jedes Gutachten verbleibt vollständig bei der unterzeichnenden ärztlichen Fachkraft.

2 Funktionsweise und Prinzipien

Amtsärztliche Gutachten entfalten in nachgelagerten Entscheidungsinstanzen eine rechtliche Wirkung für die betroffenen Personen. Ein System, das diesen Prozess unterstützt, muss daher Anforderungen genügen, die über rein funktionale Leistungsfähigkeit hinausgehen. Die Architektur des Assistenzsystems wurde auf Basis dieser Anforderungen entworfen.

Der Verarbeitungsprozess gliedert sich in vier Schritte: Dokumentenverarbeitung, strukturierte Zusammenfassung, Kontextbereitstellung und Gutachtenerstellung. Jeder Schritt erzeugt ein prüfbares Zwischenergebnis, das die Gutachterin bzw. der Gutachter einsehen, korrigieren oder verwerfen kann, bevor der nächste Schritt ausgelöst wird. Erkennt eine Ärztin oder ein Arzt etwa in der Dokumentenzusammenfassung eine Ungenauigkeit, kann diese unmittelbar korrigiert werden. Alle nachfolgenden Schritte arbeiten dann mit dem korrigierten Stand. So wird unkontrollierte Fehlerübertragung verhindert, und die Gutachterin bzw. der Gutachter behält jederzeit die volle Kontrolle über den Prozess. Die Originaldokumente bleiben dabei stets unmittelbar zugänglich und dienen als Referenz für die Detailprüfung.



[Abbildung 1: Architektur und Workflow des Assistenzsystems]

Gutachterinnen und Gutachtern arbeiten zusätzlich zu dem in der Untersuchung erhobenen Eigenbefund mit einer heterogenen Informationsgrundlage. Die Wichtung und Bewertung dieser Informationen ist ein essentieller Bestandteil der gutachterlichen Tätigkeit. Das System bildet dies explizit ab: Amtsärztliche Befunde und objektive Diagnostik werden priorisiert, subjektive Angaben als solche gekennzeichnet. Widersprüche zwischen Quellen werden nicht stillschweigend aufgelöst, sondern nebeneinander dargestellt. Diese Anforderung entstand direkt aus dem Feedback der Amtsärztinnen und Amtsärzte im Pilotbetrieb.

Dabei folgt das System keiner generischen Textgeneration, sondern expliziten Entscheidungslogiken, die gemeinsam mit den Fachärztinnen und Fachärzten erarbeitet wurden. Für Beihilfegutachten prüft es systematisch: Liegt eine wissenschaftlich allgemein anerkannte Indikation vor? Erfüllt die Maßnahme nach den einschlägigen Verordnungen die zur prüfenden Kriterien? Fehlen erforderliche Nachweise? Ein separater Reasoning-Schritt dokumentiert diese Prüfung und macht die Argumentation für die Gutachterin bzw. den Gutachter nachprüfbar.

Die zentrale Gutachtenentscheidung, etwa die Einstufung, ob die medizinischen Voraussetzungen der Dienstfähigkeit oder der Beihilfefähigkeit einer Aufwendung vorliegen, trifft ausschließlich die Ärztin oder der Arzt. Das System erwartet diese Vorgabe als explizite Eingabe, bevor es den Gutachtentext generiert. Es argumentiert und formuliert; der Mensch entscheidet und verantwortet. Diese Architektur adressiert gezielt das Risiko des sogenannten Automation Bias, also der Tendenz, maschinell erzeugte Empfehlungen unkritisch zu übernehmen. Das System begegnet diesem Risiko,

indem es eine aktive ärztliche Entscheidung als zwingenden Prozessschritt verankert, nicht als optionale Bestätigung.

Schließlich ist das System auf Erweiterbarkeit ausgelegt. Neue Gutachtentypen oder Gesetzesänderungen lassen sich integrieren, indem die relevanten Kontextdokumente hinterlegt oder erneuert werden; die zugrundeliegende Verarbeitungslogik bleibt dabei unverändert. Perspektivisch sollen Gutachterinnen bzw. Gutachter eigene geprüfte Beispielgutachten hinterlegen können, die das System als Referenz für Stil und Struktur verwendet.

Das System deckt bisher drei Gutachtentypen ab: Dienstfähigkeit, Beihilfe und Verfahren nach dem Psychisch Kranken-Hilfe-Gesetz (PsychKHG). Diese Auswahl ist bewusst gewählt, weil die drei Typen strukturell sehr unterschiedliche Anforderungen stellen, von der Art der auszuwertenden Informationen über den relevanten Rechtsrahmen bis zur Entscheidungslogik. Wenn ein System diese Bandbreite abbilden kann, ist die Erweiterung auf weitere Gutachtentypen keine grundsätzliche, sondern eine inkrementelle Aufgabe.

Diese Architektur lässt sich durch die vier Leitprinzipien *Nachvollziehbarkeit*, *Intervenierbarkeit*, *menschliche Entscheidungshoheit* und *Anpassungsfähigkeit* zusammenfassen, siehe Abbildung 1.

3 Was der Pilotbetrieb verändert hat

Die Entwicklung eines KI-gestützten Assistenzsystems für das amtsärztliche Begutachtungswesen ist in Deutschland Neuland. Es gibt kein vergleichbares Produkt, keine etablierten Praktiken und kaum übertragbare Erfahrungswerte, weder technisch noch methodisch.

Gerade deshalb ist die kontinuierliche Einbindung der späteren Anwenderinnen und Anwender eine methodische Notwendigkeit. Nur Gutachterinnen bzw. Gutachter selbst verfügen über das implizite Prozesswissen, das bestimmt, welche Funktionalitäten tatsächlich Entlastung schaffen und welche an der Praxis vorbeigehen. Hierbei stellen sich Fragen wie: Was sind nur eine - möglicherweise auch durch Fachkreise geäußerte – Meinungen? Welche Beurteilung ist durch objektivierende Befunde hinreichend belegt? Wie ordnet man subjektive Beschwerdeschilderung, die mit den objektiven Befunden nicht unmittelbar erklärbar sind, ein? Mit welcher Unsicherheit ist eine Prognose überhaupt noch möglich?

Diese Fragen haben keine technische Standardantwort. Sie erfordern ein Verständnis für die Denkweise bei Begutachtungen, das sich nur im fortlaufenden Austausch aufbauen lässt. Die calliora GmbH entwickelte den KI-gestützten Assistenten daher in enger Abstimmung mit Amtsärztinnen und Amtsärzten des Gesundheitsamts Reutlingen. In regelmäßigen Feedbackrunden wurden Anforderungen präzisiert und Funktionalitäten iterativ angepasst.

Dieser Prozess führte zu wesentlichen Verbesserungen. In den Feedbackrunden zeigte sich die strukturierte Zusammenfassung als besonders wertvoll für den Arbeitsalltag der Gutachtenden, weshalb dieser Schritt entsprechend verfeinert wurde. Gerade bei umfangreichen Fällen mit Dokumenten verschiedener Behandlerinnen und Behandler spart die kompakte Aufbereitung Zeit, da Amtsärztinnen und Amtsärzten oft widersprüchliche Unterlagen und Informationen vorliegen. Diese werden nun kompakt nebeneinander dargestellt, was den Abgleich erheblich erleichtert, siehe Abbildung 2. Die potenziellen Widersprüche zwischen den Befunden werden anschließend im Reasoning-Schritt unter Bezugnahme auf die ärztliche Entscheidungsvorgabe systematisch aufgelöst. Die grundlegendste Erkenntnis betrifft jedoch die Entscheidungsvorgabe selbst: Ursprünglich sah der Entwurf vor, dass das System eine eigene Einschätzung vorschlägt. Im Austausch mit den Ärztinnen und Ärzten zeigt sich, dass dies weder der fachlichen Praxis entspricht noch die gewünschte Akzeptanz und Qualität erzeugt. Die Gutachterinnen und Gutachter wollen und sollen selbst entscheiden. Die Aufgabe der KI beschränkt sich darauf, diese Entscheidung wägend und gleichzeitig argumentativ zu untermauern und auszuformulieren. Dieser konsequente Human-in-the-Loop-Ansatz adressiert die berechtigte Sorge vor Automation Bias. Die Erkenntnis kam nicht aus einer theoretischen Analyse, sondern aus der praktischen Erprobung.

Schritte
3/4 freigegeben
Vorverarbeitung
Freigegeben
Zusammenfassung
Freigegeben
Recherche
Freigegeben
Gutachtenerstellung
Ausstehend

Zusammenfassung
Zusammenfassung der Eingabedaten und Dokumenteninhalte
Freigegeben: 10.03.2026 09:58

Konfiguration & Eingaben

Medizinische Dokumente

Modellgröße: Groß

Klein ist schneller, Groß liefert tiefere Analysen.

Wählen Sie pro Dokument den Detailgrad. "Nicht verwenden" schließt ein Dokument aus. Standardmaessig wird "Mittel" verwendet.

amtsgutachten (Mittel)
arztbrief (Kurz)
klinikbericht (Mittel)

Dokumentenzusammenfassungen

amtsgutachten (Mittel)
arztbrief (Kurz)
klinikbericht (Mittel)

Eingabedokumente

arztbrief
Arztbrief - Fall 25
Dr. med. [Name]
Fachärztin für Orthopädie und Unfallchirurgie
Orthopädische Praxis Dr. [Name]
06151- [Name]
06151- [Name]
praxis@dr- [Name].de
13.06.2024
Sehr geehrte Kolleginnen und Kollegen,
Markus [Name]
05.03. [Name]
AOK65432 [Name]

Anamnese
Herr [Name] stellt sich erneut vor wegen seit etwa drei Jahren bestehenden Nackenschmerzen, die in die rechte Schulter und gelegentlich in den rechten Arm ausstrahlen. Die Beschwerden verschlechtern sich bei längerer Bildschirmarbeit und statischer Kopfhaltung. Seit einigen Monaten berichtet er zusätzlich über morgendliche Steifigkeit der Finger beider Hände, die nach etwa 20-30 Minuten nachlässt. Gelegentlich Kribbeln im rechten Daumen und Zeigefinger, keine ausgeprägten Taubheitsgefühle. Subjektiv fühlt er sich beim Heben schwerer Gegenstände kraftloser als früher. Keine Blasen- oder Mastdarmstörungen. Keine bekannten Allergien. Keine Hinweise auf Suchtmittelkonsum. Psychische Belastung durch die chronischen Schmerzen, aber keine psychiatrische Vorerkrankung.

Körperliche Untersuchung
Normgewicht (192 cm, 85 kg), aufmerksam und kooperativer Eindruck. Beweglichkeit der HWS in Extension und Rotation nach rechts eingeschränkt, dabei Schmerzprovokation. Druckschmerz über den

[Abbildung 2: Screenshot aus dem PoC am Beispiel eines Dienstfähigkeitfalls]

4 Ergebnisse und Ausblick

Der Proof-of-Concept hat gezeigt, dass KI-gestützte Assistenzsysteme im amtsärztlichen Begutachtungswesen praktikabel und sinnvoll einsetzbar sind, wenn die Rollenteilung zwischen Mensch und Maschine klar definiert ist.

Ein Beispiel: Ein Beihilfeantrag zur Erstattung einer Vitamin-D-Therapie. Die Gutachterin bzw. der Gutachter erhält den Untersuchungsauftrag mit Gutachtenfragen und mitunter ärztliche Atteste sowie radiologische und Laborbefunde. Bisher bedeutet das: Unterlagen sichten, die einschlägigen Passagen der Beihilfeverordnung berücksichtigen, ggf. operationalisierende Urteile recherchieren, Kriterien prüfen, Argumentation strukturieren, Gutachtentext schreiben. Das System übernimmt die Vorarbeit: Es fasst die Dokumente zusammen, stellt die relevanten Rechtsgrundlagen bereit und kennzeichnet die offenen Prüfpunkte. Die Ärztin oder der Arzt prüft die Aufbereitung, trifft die Entscheidung, und das System generiert den Gutachtentext. Die Gutachterin bzw. der Gutachter arbeitet so mit einer strukturierten Vorlage statt mit einem leeren Blatt.

Das Feedback der beteiligten Fachleute unterstreicht das Potenzial des Ansatzes. In Befragungen bewertete ein Großteil der Teilnehmenden den PoC als vielversprechende Grundlage für eine Umsetzung in ÖGDigital. Die am häufigsten genannten Eindrücke waren u. a. „innovativ“ und „vielversprechend“. Aus einem Fokusinterview mit einer weiteren Gutachtenstelle, die das System unabhängig vom Entwicklungsprozess testete, ergab sich ein ähnliches Bild: Die strukturierte Zusammenfassung spare bereits heute spürbar Zeit, und auch bei der Gutachtenerstellung selbst sehe man klares Potenzial, sobald die Wissensbasis weiter ausgebaut werde. Alle Beteiligten können sich eine regelmäßige Nutzung im Arbeitsalltag vorstellen. Für das Folgeprojekt wurden insbesondere der Ausbau der Wissensbasis (zur strukturierten Sammlung, Aufbereitung und Bereitstellung von Fachwissen) sowie Schulungskonzepte für die Einführung in den Ämtern als prioritäre Themen benannt.

Die öffentliche Debatte um generative KI bewegt sich häufig zwischen zwei Extremen: vollständige Automatisierung oder pauschale Ablehnung. Die Erfahrungen aus diesem Projekt zeigen einen differenzierteren Weg. KI-Systeme in sensiblen Fachdomänen sind keine Standardlösungen. Ihr Nutzen entsteht durch die Abstimmung auf den Anwendungskontext, durch Architekturen, die menschliche Kontrolle ermöglichen, und durch die iterative Zusammenarbeit mit den Fachleuten, die das System letztlich einsetzen. Die vier Leitprinzipien des Systems sind eine direkte Antwort auf diese Anforderungen.

Der Ansatz ist nicht auf das amtsärztliche Begutachtungswesen beschränkt. Er ist überall dort übertragbar, wo drei Bedingungen zusammenkommen: eine heterogene Dokumentenlage, die verlässlich und effizient ausgewertet werden muss; ein rechtlicher oder fachlicher Rahmen, an dem sich die Bewertung orientiert; und eine finale Entscheidung, die beim Menschen verbleibt. Innerhalb

des Öffentlichen Gesundheitsdienstes lassen sich etwa sozialmedizinische Stellungnahmen, Infektionsschuttermittlungen oder die Dokumentation von Betriebsbegehungen nach dem gleichen Muster unterstützen.

Konkret stehen zwei nächste Schritte an. Erstens der Aufbau einer kuratierten Wissensbasis: Die im PoC zusammengestellten Rechtsgrundlagen, Leitlinien und Entscheidungslogiken sollen in eine strukturierte, durchsuchbare Bibliothek überführt werden, die regelmäßig zentral unter Einbeziehung von ärztlicher und juristischer Expertise gepflegt wird. Zweitens soll das System aus den Korrekturen der Gutachterinnen und Gutachter lernen und sich an deren Arbeitsweise anpassen, etwa indem geprüfte Beispielgutachten als Referenz hinterlegt werden. Die technische Grundlage dafür ist gelegt; es gilt nun, diese systematisch in größerem Rahmen und gemäß der Anforderungen der DSGVO und der EU KI-Verordnung zu validieren.

Aktuell wird im Auftrag des Ministeriums für Soziales, Gesundheit und Integration Baden-Württemberg in einem Folgeprojekt **KI@ÖGDigital** geprüft, wie der Proof of Concept als Teil des Begutachtungsmoduls der neuen, einheitlichen, cloudbasierten Fachanwendung ÖGDigital für die Gesundheitsämter in Baden-Württemberg in eine Produktionsumgebung überführt werden kann. Die Workflow-Architektur ist generisch genug, um perspektivisch auch andere Fachmodule innerhalb von ÖGDigital mit KI zu unterstützen.