KI IN DER ARZTPRAXIS · MODUL 2 VON 5

KI richtig einsetzen: Prompting, Tools und Arbeitsabläufe

Teilnehmer-Handout zu Modul 2 der CME-Webinarreihe „KI in der Arztpraxis“. Vorabversion zur Vorbereitung auf das Webinar am 16. September 2026 und auf die Lernerfolgskontrolle.

Referent: Dr. med. Christian Mauch, Facharzt für Orthopädie und Unfallchirurgie, Chirotherapie, Sportmedizin. Geschäftsführer und Ärztlicher Leiter des Maybach Medical MVZ Stuttgart. Interessenkonflikt: Der Referent ist Geschäftsführer der SummitGO GmbH & Co. KG, die eine integrierte KI-Lösung für Arztpraxen entwickelt, und hat daran ein wirtschaftliches Interesse. Weitere Beziehungen zu Herstellern bestehen nicht. Anbieter werden in diesem Handout nur als Beispiele für Kategorien genannt, eine Kaufempfehlung ist damit nicht verbunden.

Wie Sie dieses Handout nutzen

Das Handout enthält den Stoff des zweiten Moduls in Textform. Es ersetzt den Vortrag nicht, es bereitet ihn vor. Wer die zwölf Abschnitte vor dem Webinar liest, kennt die Begriffe, die Studien und die Prüfschritte, auf denen die Lernerfolgskontrolle aufbaut. Die Kontrolle besteht aus zehn Fragen mit je fünf Antworten, von denen genau eine richtig ist. Sieben richtige Antworten genügen für die zwei CME-Punkte. Die Fragen verlangen kein Auswendiglernen von Zahlen, sie verlangen, dass Sie eine Situation aus der Praxis richtig einordnen: Welcher Fehler steckt in diesem Arztbrief, welche Antwort eines Anbieters ist ein Warnsignal, was verlangt die KI-Verordnung (EU) 2024/1689 von einem Telefonassistenten. Solche Einordnungen üben die Abschnitte dieses Handouts.

Am Ende finden Sie die Prompt-Bibliothek des Moduls mit fünf Musterprompts, die Sie in der Praxis direkt verwenden können, ein Verzeichnis der zitierten Studien und die Hausaufgabe bis Modul 3. Ein eigenes Glossar mit allen Begriffen der Reihe erhalten Sie als separates Dokument.

1. Wo die Praxen stehen

Datenschutz ist die Sorge Nummer eins, und trotzdem arbeitet die Hälfte der Ärzte mit ungeprüften Werkzeugen. Der Widerspruch erklärt, warum das Modul mit dem Handwerk beginnt und nicht mit dem Verbot.

Je nach Befragung nennen 54 bis 72 Prozent der Ärzte Datenschutz und Datensicherheit als größte Sorge beim Einsatz von KI. Die Zahlen stammen aus dem Doctolib Digital Health Report 2026 mit 414 befragten Ärzten und aus der Erhebung von Bitkom und Hartmannbund 2025 mit 616 Ärzten. Dieselbe Bitkom-Befragung zeigt, dass 50 Prozent der Befragten dennoch mit nicht geprüften Werkzeugen wie ChatGPT arbeiten. Die Sorge verändert das Verhalten nicht, solange die Praxis nichts Geprüftes bereitstellt. Für die Praxisleitung folgt daraus eine Aufgabe, die im dritten Abschnitt zu den Werkzeugen wieder auftaucht: ein freigegebenes Werkzeug anbieten, sonst arbeitet das Team mit privaten Konten weiter.

Zwei weitere Zahlen ordnen das Umfeld ein. Im Bitkom Cloud Report vom Juni 2026 halten 85 Prozent der befragten Unternehmen Deutschland für zu abhängig von US-Anbietern, im Vorjahr waren es 78 Prozent. 37 Prozent nähmen Nachteile in Kauf, wenn die Verarbeitung dafür ausschließlich in Deutschland läuft, im Vorjahr 27 Prozent. Beide Werte stammen aus einer Befragung von 603 Unternehmen aller Branchen, nicht aus Arztpraxen. Für die ambulante Medizin zeigen sie eine Richtung, gemessen ist dort nichts.

Ob Ärzte die Verarbeitung lieber in der Cloud oder auf einem Rechner in der Praxis hätten, hat bisher keine Erhebung direkt gefragt. Das Bild entsteht aus Nachbarbefunden. In den Kliniken bieten CGM, Corti, Tiplu und ARGO am Universitätsklinikum Hamburg-Eppendorf seit 2026 beide Betriebsarten an, ARGO wirbt ausdrücklich damit, dass die Daten das Kliniknetz nicht verlassen. Für Praxen laufen die Diktatdienste bisher als Cloud in der EU. tomedo teilt den Weg bereits auf: Die Spracherkennung läuft auf dem Praxis-Mac, die Textarbeit über einen

eigenen Sprachdienst. Aus den USA liegt eine KLAS-Erhebung vom Juli 2025 vor, nach der knapp 40 Prozent der befragten Einrichtungen mindestens 90 Prozent ihrer IT im eigenen Haus halten. Mit 36 Einrichtungen ist die Stichprobe klein und nicht übertragbar.

2. Die Begriffe, die Sie brauchen

Fünfzehn Begriffe reichen, um jedes Anbietergespräch und jede Diskussion im Team zu führen. Der wichtigste Satz des Moduls steht gleich am Anfang.

Ein Sprachmodell, englisch Large Language Model oder LLM, hat aus Milliarden Textbeispielen gelernt, welches Wort als nächstes am wahrscheinlichsten passt. Es rechnet, es versteht nicht. Alles Weitere in diesem Modul folgt aus diesem Satz. Ein Prompt ist alles, was Sie eintippen oder diktieren. Er ist ein Auftrag an eine Assistenz und keine Suchanfrage. Prompting ist die Technik, den Auftrag so zu formulieren, dass die Antwort brauchbar wird. Die Reihe arbeitet dafür mit fünf Bausteinen, die im vierten Abschnitt ausführlich vorkommen: Rolle, Kontext, Aufgabe, Format und Leitplanken.

Das Kontextfenster ist das Kurzzeitgedächtnis des Modells. Es umfasst so viel Text, wie das Modell gleichzeitig im Blick behält. Läuft es in einem langen Gespräch über, fällt der Anfang unbemerkt heraus. Deshalb wiederholen erfahrene Nutzer wichtige Vorgaben in langen Sitzungen. Gemessen wird das Fenster in Token, Textbausteinen von wenigen Zeichen. Token bestimmen zugleich, was der Anbieter in Rechnung stellt. Eine Halluzination ist der typische Fehler eines Sprachmodells: Es liefert Plausibles statt Belegtes, flüssig formuliert und frei erfunden. Der Begriff klingt dramatischer als der Umgang damit. Die Plausibilität einer Antwort prüfen Sie in zehn Sekunden, eine Quellenangabe in ein bis zwei Minuten.

Pseudonymisierung ersetzt identifizierende Daten durch Platzhalter, bevor ein Text die Praxis verlässt. Der Rückschluss auf die Person bleibt in der Praxis möglich, und genau deshalb bleiben pseudonymisierte Daten personenbezogene Daten im Sinne der Datenschutz-Grundverordnung. Bei der Anonymisierung ist der Rückschluss für niemanden mehr möglich. Wer die beiden Begriffe verwechselt, unterschätzt seine Pflichten. Modul 3 vertieft das. Bias bezeichnet die Verzerrung, die ein Modell aus seinen Trainingstexten übernimmt: überrepräsentierte Gruppen, veraltete Leitlinien, Versorgung nach angelsächsischem Muster. Eine Antwort zur Therapie in Deutschland kann auf amerikanischer Leitlinienlage beruhen, ohne dass der Text es verrät.

Reasoning-Modelle rechnen vor der Antwort Zwischenschritte durch. Sie sind bei mehrstufigen Fragen genauer, dafür langsamer und teurer. Die ausgegebenen Zwischenschritte sind eine Darstellung und kein Nachweis der Richtigkeit, sie gehören nicht als Begründung in die Akte. Multimodale Modelle verarbeiten Text, Bild, Ton und Tabellen gemeinsam, etwa ein abfotografiertes Laborblatt. Ein allgemeines Modell bleibt dabei ohne CE-Kennzeichnung kein Medizinprodukt, eine Befundung damit ist nicht vorgesehen. Schatten-KI meint die Nutzung privater KI-Konten durch Mitarbeitende ohne Wissen der Leitung. Ein Verbot beendet sie erfahrungsgemäß nicht, eine schriftliche Praxisregelung schon.

Vier Schlagworte begegnen Ihnen in Anbietergesprächen. Ein KI-Agent führt Schritte selbst aus, legt Termine an, versendet Briefe, trägt Daten ein. Die Frage an jeden Agenten lautet, wo der Vorschlag endet und Ihre Freigabe beginnt. RAG, ausgeschrieben Retrieval-Augmented Generation, bedeutet, dass das Modell zuerst Ihre eigenen Leitlinien und Vorlagen durchsucht und daraus antwortet. So entsteht eine belegte Fundstelle statt einer erfundenen Quelle, und so bringen Sie einem Modell Ihre Praxisstandards bei, ohne es zu trainieren. Ein Systemprompt hinterlegt Rolle, Format und Leitplanken dauerhaft, etwa in einem Projekt oder einem eigenen Assistenten, danach tippen Sie nur noch Fall und Aufgabe ein. Vibe-Coding schließlich beschreibt, kleine Helfer per Anweisung bauen zu lassen, ohne zu programmieren. Drei Grenzen gelten: keine Patientendaten, kein Medizinprodukt, und niemand wartet das Ergebnis.

> **WICHTIG**
> Ein Sprachmodell rechnet Wahrscheinlichkeiten, es weiß nichts. Pseudonymisierte Daten bleiben personenbezogene Daten. Die beiden Sätze tragen den größten Teil des Moduls.

3. Spracherkennung ist nicht KI

Die häufigste Verwechslung im Praxisalltag: Diktiersoftware wird für KI gehalten. Der Unterschied zeigt sich an der Art des Fehlers.

Spracherkennung, englisch Automatic Speech Recognition oder ASR, hört zu und schreibt mit. Sie ordnet Laute Wörtern zu und nicht Inhalte Sätzen. Sie erfindet nichts, sie verhört sich, und der Fehler fällt beim Lesen auf. Whisper ist ein Beispiel für diese Kategorie. Ein Sprachmodell liest Text und formuliert ihn um, es strukturiert, kürzt und kodiert. Dabei ordnet es Inhalte neu und ergänzt, was üblich klingt. Sein Fehler liest sich fehlerfrei. ChatGPT, Claude und Gemini sind Beispiele für diese Kategorie.

Ambient Documentation schaltet beide Techniken in Reihe. Das Mikrofon erzeugt das Rohprotokoll des Gesprächs, das Sprachmodell macht daraus den Brief. Zwei Techniken hintereinander bedeuten zwei Fehlerquellen und deshalb zwei Prüfschritte: das Transkript gegen das Gespräch, den Brief gegen das Transkript. Zu einer solchen Lösung gehören zwei Freigaben und nicht eine.

Im Webinar zeigt der Referent als offengelegtes Fallbeispiel an seiner eigenen Lösung, wie sich eine solche Kette aus Spracherkennung, Pseudonymisierung, Sprachmodell und ärztlicher Freigabe aufbauen lässt. Das Beispiel dient der Anschauung und ist keine Empfehlung. Die Fragen, die es beantworten muss, sind dieselben wie für jedes andere Produkt und stehen im sechsten Abschnitt: Wo wird transkribiert, was verlässt die Praxis, wird mit den Daten trainiert, wo stehen die Server, und wie kommt der Text mit welcher Freigabe zurück.

4. Prompting: fünf Bausteine und eine Reihenfolge

Wer einem Assistenzarzt sagt „schreib mal was zum Knie“, bekommt auch nicht, was er will. Ein guter Prompt hat fünf Bausteine.

Die Rolle beantwortet, wer die KI sein soll. Ein Satz genügt: Du bist ein erfahrener Orthopäde mit Schwerpunkt Sportmedizin. Ausgedachte Biografien verschlechtern das Ergebnis. Der Kontext beschreibt die Situation: Patient, 52 Jahre, chronische Knieschmerzen rechts seit sechs Monaten, BMI 28. Die Aufgabe sagt, was das Modell tun soll: Erstelle eine strukturierte Differentialdiagnose mit den drei wahrscheinlichsten Diagnosen. Das Format legt fest, wie das Ergebnis aussieht: tabellarisch mit Diagnose, Wahrscheinlichkeit und empfohlener Diagnostik. Die Leitplanken bestimmen, was das Modell nicht darf: Erfinde keine Befunde, die ich nicht genannt habe. Kennzeichne, was du nicht weißt. Dieser letzte Baustein zieht sich durch das ganze Modul, durch die Demos, die Checkliste und die Hausaufgabe.

Die meisten Kollegen lassen Rolle und Format weg und wundern sich über generische Ergebnisse. Zugleich hat sich die Dosierung verschoben. Bei aktuellen Modellen bringen ausgeschmückte Rollenbeschreibungen kaum noch Gewinn und können als Rauschen wirken, Anthropic nennt in seiner Leitlinie für 2026 starkes Rollen-Prompting ausdrücklich als weniger notwendig. Die fünf Bausteine bleiben der Standard der Reihe, geändert hat sich die Dosis der Rolle. Länger ist nicht besser.

PICO als Prompt-Baukasten

Für klinische Fragen bietet sich eine Struktur an, die dreißig Jahre älter ist als ChatGPT. Richardson und Wilson beschrieben 1995 im ACP Journal Club, wie eine klinische Frage gebaut sein muss, damit die Literatur sie beantworten kann. Entstanden ist die Schule an der McMaster University um Sackett und Guyatt, heute ist sie Standard in Leitlinienarbeit und bei Cochrane-Reviews. PICO steht für Population oder Patient, Intervention, Comparison und Outcome. Ein Beispiel: 58-jährige Patientin mit Gonarthrose Grad III und einer Varusachse von acht Grad, Umstellungsosteotomie gegen unikondylären Gelenkersatz, verglichen mit konservativer Therapie, gemessen an Schmerz, Funktion und Standzeit nach zehn Jahren.

PICO ersetzt den Baustein Kontext und die Frage. Rolle, Aufgabe, Format und Leitplanken kommen hinzu, erst dann wird aus der Fragestruktur ein Prompt. Das Format verlangt eine Antwort in drei Sätzen, dann je Studie Autor, Jahr, Studientyp, Fallzahl und Ergebnis, dann den Leitlinienbezug mit Fassung und Jahr, dann die Punkte, in denen sich die Studien widersprechen, zuletzt das, was für diese Patientin offen bleibt. Die Leitplanken

verlangen zu jeder Aussage die Quelle mit Jahr, verbieten erfundene Studien, DOI und Fallzahlen, schließen eine Therapieempfehlung aus und verlangen, dünne Evidenz ausdrücklich zu benennen. Der vollständige Prompt steht in der Prompt-Bibliothek am Ende.

Was die Struktur messbar bringt

Wu und Kollegen testeten 2023 in einem Preprint auf arXiv (2307.08922), ob Sprachmodelle klinische Fälle besser lösen, wenn man sie nach ihrem Gedankengang fragt. Die Trefferquote stieg um 15 Prozent, bei Fächern, von denen das Modell wenig gesehen hatte, um 18 Prozent. Drei Einschränkungen gehören zu dieser Zahl: Getestet wurde an Textfällen und nicht an Patienten, die Zahl stammt von den Autoren selbst, und die Modelle von 2023 sind nicht die von 2026. Neuere Modelle rechnen von sich aus in Schritten, der Zusatz „denk Schritt für Schritt“ bringt dort nichts mehr und kann die Ausgabe verschlechtern. Was bleibt, ist der Kern der Beobachtung: Sagen Sie dem Modell, in welcher Reihenfolge Sie die Antwort wollen, etwa Verdachtsdiagnosen, dann Begründung, dann nächster Schritt. Eine klare Ansage schlägt die gut gemeinte Frage.

Auch die Form der KI-Antwort verändert die ärztliche Entscheidung. In einer Arbeit in npj Digital Medicine 2026 bewerteten Radiologen 2.020 Fälle. Wie die KI ihre Begründung aufschrieb, veränderte das Urteil der Ärzte. Das ist ein Argument für das Format als Baustein und zugleich eine Warnung, die im achten Abschnitt unter dem Stichwort Automation Bias wiederkehrt.

Kontext verbessert die Sprache, nicht die Wahrheit

Dieselbe Aufgabe ergibt in drei Vorbereitungsstufen drei verschiedene Ergebnisse. Ohne Kontext, mit dem Auftrag „Schreib einen Arztbrief zum Knie“, entsteht allgemeine Sprache in beliebiger Struktur, und Sie schreiben den Entwurf weitgehend um. Mit Kontext, also Patientin, 54, Kniegelenk links, Zustand nach Arthroskopie, Streckdefizit zehn Grad, Brief an den Hausarzt, verengt das Modell seinen Wortschatz auf die Orthopädie, schreibt Streckdefizit statt eingeschränkter Beugung und ergibt einen brauchbaren Entwurf. Mit Wissensbasis, also hinterlegten eigenen Vorlagen und Nachbehandlungsschemata, antwortet das Modell aus Ihrem Haus-Standard mit belegter Fundstelle. Der Unterschied zwischen der zweiten und der dritten Stufe ist grundsätzlich: Kontext gibt dem Modell keine neuen Informationen, die Wissensbasis schon.

Ein fachlich präziserer Text macht falsche Seitenangaben und erfundene Dosierungen plausibler und damit schwerer erkennbar. Deshalb bleibt die Prüfliste des fünften Abschnitts auch beim besten Prompt in Kraft.

Looping: der Dialog als Fallbesprechung

Statt alles in einen einzigen Prompt zu packen, können Sie in Schleifen nachfragen. Sie geben den Fall ein, Patient 45 Jahre, akute Schulterschmerzen rechts, kein Trauma, nachts schlechter. Das Modell nennt Differentialdiagnosen. Sie ergänzen den klinischen Befund, Painful Arc positiv, Jobe negativ, Neer positiv, im Röntgen ein kraniales Kalkdepot. Das Modell verengt auf die Tendinosis calcarea. Sie fragen nach den Therapieoptionen nach aktueller Leitlinie. Das Vorgehen ist näher an der klinischen Realität als der Riesenprompt, weil Sie Zwischenergebnisse sehen, nachhaken und korrigieren können. Eine Regel gehört dazu: Klebt das Modell an einer Hypothese, führen Sie den Dialog nicht weiter. Öffnen Sie eine neue Sitzung und geben Sie den Fall neutral neu ein, denn im alten Gespräch liegt die Hypothese im Kontextfenster und färbt jede weitere Antwort.

> **WICHTIG**
> Fünf Bausteine in dieser Reihenfolge: Rolle, Kontext, Aufgabe, Format, Leitplanken. Die Leitplanke der Reihe lautet: Erfinde keine Befunde, die ich nicht genannt habe. Kennzeichne, was du nicht weißt.

5. Der KI-Arztbrief: vier Fehlerstellen

Wo ein KI-erzeugter Arztbrief typischerweise falsch ist, in der Reihenfolge, in der Sie ihn lesen sollten.

Erstens die Seite. Aus rechts wird links. Das ist der häufigste und teuerste Fehler in der Orthopädie, und er entsteht, weil ein Sprachmodell Seitenangaben als austauschbare Wörter behandelt. Prüfen Sie jede Seitenangabe im Text gegen Ihren Befund. Zweitens Zahlen und Grade. Aus einem Streckdefizit von zehn Grad werden zwanzig, Bewegungsausmaße, Gradangaben und Zeiträume stimmen oft nur ungefähr. Drittens die Medikation. Das Modell schreibt einen Wirkstoff ohne Dosis oder eine Dosis, die Sie nie genannt haben, weil es ergänzt, was üblich wäre. Viertens die Vorgeschichte. Alles, was nach Anamnese klingt und nicht im Prompt stand, kann erfunden sein. Ein plausibler Verlauf ist noch kein dokumentierter Verlauf.

Die vier Fehlerstellen sind die Gegenprobe zu den Leitplanken. Wer im Prompt verlangt, dass nur übergebene Daten verwendet werden und fehlende Angaben als Lücke gekennzeichnet werden, senkt die Fehlerrate. Er hebt sie nicht auf. Die Prüfung vor der Unterschrift bleibt.

6. Werkzeuge und Anbieter

Drei Kategorien, eine Grenze aus der Medizinprodukteverordnung, fünf Fragen an jeden Anbieter und die Evidenz zur Ambient-Dokumentation.

Wer nutzt was

Vier Zahlen werden oft zitiert. 2025 sagte jede siebte Praxis, sie nutze KI. Gefragt hatten Bitkom und Hartmannbund, geantwortet hatte, wer wollte, für alle Praxen steht die Zahl also nicht. 2026 nannte eine Erhebung 37,6 Prozent. Gefragt hatte diesmal ein Anbieter, geantwortet hatten 194 Praxen, das zeigt eine Richtung und mehr nicht. In den Krankenhäusern nutzt etwa jede fünfte Abteilung KI, 18 Prozent, doppelt so viele wie 2022, und meist geht es um Bilder und nicht um Briefe. Bei Werkzeugen, die für die Praxis nicht geprüft sind, liegen die Ärzte vor dem Team: beim Recherchieren 50 zu 30 Prozent, beim Dokumentieren 28 zu 17 Prozent. Die Praxisleitung, die kein geprüftes Werkzeug bereitstellt, überlässt die Auswahl den privaten Konten ihrer Mitarbeiter.

Drei Kategorien

Allgemeine Sprachmodelle wie ChatGPT, Claude oder Gemini sind vielseitig, sofort verfügbar und kostenlos oder günstig. Sie sind kein Medizinprodukt, die Verantwortung für Datenschutz und Validierung liegt vollständig bei Ihnen, und sie halluzinieren. Medizinische KI-Assistenten wie Heidi Health, Abridge, Suki, Nuance DAX oder Open Evidence sind auf Medizin ausgerichtet, teils MDR-konform und bieten Schnittstellen. Sie kosten, binden an einen Anbieter, und der CE-Status ist je Produkt zu prüfen. Integrierte Praxislösungen von CGM, medatixx, tomedo oder Doctolib sind durchgängig im Praxisablauf, der Datenschutz ist im Vertrag geregelt, dafür sind Investition und Einführungsaufwand höher. Der Referent ist Gründer einer Lösung der dritten Kategorie, der Interessenkonflikt ist offengelegt. Die Nennungen sind Beispiele je Kategorie mit Marktstand September 2026, keine Empfehlung.

Auch die Preislogik unterscheidet sich je Kategorie. Reine Spracherkennung wie Dragon Medical, T-PRO, Ambivio oder Docviant kostet nach öffentlichen Angaben 40 bis 150 Euro je Arzt und Monat, Diktierlösungen 60 bis 90 Euro. Praxisverwaltungssysteme liegen je nach System und Praxisgröße zwischen 45 und 1.900 Euro im Monat, die Frage ist, ob das KI-Modul enthalten ist oder einzeln berechnet wird. Integrierte KI-Module kosten für Ambient-Dokumentation 99 bis 149 Euro, für Telefonassistenten 49 bis 200 Euro je Monat. Bei jeder Kategorie stellen Sie eine andere Prüffrage: Läuft die Erkennung auf dem Praxisrechner oder in der Cloud, ist das Modul enthalten, und wie steht es um CE-Status, Auftragsverarbeitungsvertrag und Rückweg in die Akte. Die Spannen stammen aus Anbietervergleichen von Medizinio 2026, dem Praxissoftwarevergleich 2026 auf Basis des Zi-PVS-Monitorings 2025 und aus Preislisten der Anbieter.

Ein Hinweis zur Marktlage: Die Anbieter kämpfen um Marktanteile, und viele Preise decken derzeit weder die Entwicklung noch den laufenden Verbrauch an Rechenleistung. Wer heute günstig einkauft, rechnet besser mit späteren Erhöhungen. Über die Einbindung in das Praxisverwaltungssystem entscheidet in der Praxis der Rückweg des fertigen Briefs in die Akte, fragen Sie nach diesem Weg und nicht nach dem Schnittstellenkürzel.

Eine Lösung, die den ganzen Weg vom Gespräch über den Brief bis zur Abrechnung in ärztlicher Qualität abdeckt, gibt es im September 2026 nicht.

Die MDR-Grenze: Assistenz ja, Diagnose nein

Als Assistenz erlaubt sind Recherche und Literaturzusammenfassung, der Entwurf und die Strukturierung von Arztbriefen, die Differentialdiagnose als Denkstütze, die Vorbereitung der Patientenaufklärung, Vorschläge zu Kodierung und Abrechnung und das Einordnen von Laborwerten. Die Grenze verläuft bei Diagnose und Therapie. Die Diagnosestellung bleibt ärztlich, die Therapieentscheidung bleibt ärztlich, KI-Ergebnisse werden nie ungeprüft übernommen. Beansprucht ein Werkzeug eine medizinische Zweckbestimmung, sind CE-Status und Konformität nach der Verordnung (EU) 2017/745, der MDR, zu prüfen. Allgemeine Sprachmodelle sind keine Medizinprodukte. Die Bundesärztekammer hat das in ihrer Stellungnahme vom Januar 2025 so beschrieben, die Zentrale Ethikkommission bereits 2021.

Den CE-Status eines Produkts prüfen Sie in fünf Minuten. Lesen Sie die Zweckbestimmung in der Gebrauchsanweisung. Suchen Sie das CE-Zeichen mit der vierstelligen Nummer der Benannten Stelle. Prüfen Sie den Eintrag in EUDAMED, der europäischen Datenbank für Medizinprodukte. Suchen Sie in den Geschäftsbedingungen den Satz, mit dem sich der Anbieter selbst als Nicht-Medizinprodukt bezeichnet. Ein Produkt, das laut Werbung Diagnosen stellt und laut AGB kein Medizinprodukt ist, hat ein Problem, das Sie nicht übernehmen sollten. Sprachlich genau: Ein Medizinprodukt durchläuft eine Konformitätsbewertung und trägt eine CE-Kennzeichnung, eine „Zulassung“ gibt es in diesem Verfahren nicht.

Ambient-Dokumentation: was die Studien übrig lassen

Zur Ambient-Dokumentation liegen inzwischen zwei randomisierte Studien und eine große Auswertung vor. Lukac und Kollegen randomisierten an der UCLA 238 Ärzte aus 14 Fächern mit rund 72.000 Patientenkontakten auf zwei Systeme und eine Kontrollgruppe (NEJM AI, Dezember 2025). Mit dem stärkeren der beiden Systeme sank die Schreibzeit je Notiz um etwa 41 Sekunden, das sind rund zehn Prozent. Das zweite System erreichte keinen signifikanten Effekt. Das Signal stammt aus der Praxissoftware und misst die Editierzeit in der Scribe-Plattform selbst nicht mit, eine Einschränkung, die auch für die nächste Studie gilt.

Rotenstein und Kollegen werteten an fünf Zentren 8.581 Ärzte aus, von denen 1.809 ein Ambient-System nutzten (JAMA, April 2026). Die Dokumentationszeit sank um 16 Minuten je acht Stunden Patientenarbeit, die Zeit in der Akte um 13 Minuten, und die Ärzte sahen 0,49 Patienten mehr je Woche. Die relativierende Zahl gehört dazu: Die Zeit außerhalb der Sprechstunde blieb unverändert. Nur ein Drittel der Nutzer setzte das System in mindestens jedem zweiten Termin ein, und genau dort war der Gewinn am größten. Die gewonnene Zeit verschwindet, wenn sofort neue Aufgaben nachrücken.

Zur Erschöpfung liegen zwei Befunde vor, die in dieselbe Richtung zeigen und sich an einem Punkt unterscheiden. Bei freiwilligen Anwendern ging die Erschöpfung um 21,2 Prozentpunkte zurück (You et al., 2025), eine Beobachtung mit Selbstselektion. Eine Crossover-Studie mit 160 Behandlern fand die Besserung ebenfalls, aber keine kürzere Arbeit nach Dienstschluss (Chowdhury et al., JAMIA, Mai 2026). Ein Satz für das Sprechzimmer gehört zu jedem dieser Systeme: „Ich nutze ein System, das unser Gespräch mitschreibt. Ich lese und unterschreibe jeden Text selbst. Ist Ihnen das recht?“

Fünf Fragen an jeden Anbieter

Wo wird transkribiert? Eine gute Antwort nennt den Rechner in der Praxis, ein Warnsignal ist „in der Cloud, aber sicher“ ohne Ortsangabe. Was verlässt die Praxis? Gut ist ein Beispieltext ohne Namen, am Bildschirm vorgeführt, ein Warnsignal ist die Zusicherung ohne Beispiel. Wird mit meinen Daten trainiert? Gut ist ein Nein im Vertrag, ein Warnsignal ist ein Nein in der Broschüre. Wo stehen die Server? Gut sind Standort, Betreiber und Auftragsverarbeitungsvertrag schriftlich, ein Warnsignal ist „EU-Rechenzentrum“ als einzige Auskunft. Wie kommt der Text zurück? Gut ist eine Schnittstelle ins Praxissystem mit Freigabe durch Sie, ein Warnsignal ist Kopieren und Einfügen. Die fünf Fragen gelten für jedes Produkt, auch für das des Referenten.

Das C5-Testat

C5 steht für Cloud Computing Compliance Criteria Catalogue, die deutsche Prüfliste für Cloud-Sicherheit. Das Bundesamt für Sicherheit in der Informationstechnik hat darin über 100 Kriterien festgelegt, von der Verschlüsselung über den Zutritt zum Rechenzentrum bis zum Umgang mit Sicherheitsvorfällen. Unabhängige Wirtschaftsprüfer kontrollieren die Einhaltung nach dem internationalen Prüfstandard ISAE 3000, ähnlich einer Bilanzprüfung. Typ 1 ist eine Momentaufnahme, Typ 2 eine Verlaufskontrolle über mindestens sechs Monate. Nach § 393 SGB V dürfen Gesundheitsdaten nur in eine Cloud, deren Dienst ein aktuelles C5-Testat vorweist. C5 ist kein Zertifikat, es ist ein Testat, eine Prüfbescheinigung mit Ablaufdatum. Für Ihr Anbietergespräch heißt das: Fragen Sie nach dem Testat des Cloud-Dienstes, nach dem Typ und nach dem Datum.

Die Grundausstattung und die vier Schritte davor

Eine Praxis, die morgen anfangen will, braucht vier Dinge. Erstens eine Bezahlversion eines allgemeinen Sprachmodells, Größenordnung 20 bis 30 Euro je Nutzer und Monat, denn kostenlose Varianten trainieren teils mit den Eingaben. Zweitens einen Dokumentationsweg, Diktat oder Ambient je nach Arbeitsweise, und erst hier stellt sich die Frage nach CE und Auftragsverarbeitungsvertrag. Drittens ein Recherchewerkzeug, das Antworten an Fundstellen bindet, weil Antworten mit Fundstelle seltener frei erfundene Angaben enthalten. Viertens eine Seite Regeln: die Praxisregelung und den Schulungsnachweis. Der Teil kostet nichts und ist der, den die meisten auslassen.

Zwischen Kaufentscheidung und erstem Patienten liegen vier Schritte und etwa zwei Wochen. Zuerst Zweckbestimmung und CE: Beansprucht das Produkt einen medizinischen Zweck, lassen Sie sich CE-Kennzeichnung und MDR-Konformität nachweisen, vorher kein Einsatz. Dann Auftragsverarbeitungsvertrag und Serverstandort, mit Standort und Betreiber im Vertrag und schriftlichem Trainingsausschluss. Dann eine Testphase ohne Patienten, zwei Wochen mit erfundenen Testfällen, bei Telefonassistenten zusätzlich die Testliste zur Notfallerkennung aus dem siebten Abschnitt. Zuletzt Schulung und Freigabe: Team geschult mit Datum, eine benannte Person gibt frei, Eintrag in der Praxisregelung. Erst dann der erste echte Fall.

> **WICHTIG**
> Vor jedem Einsatz eines KI-Werkzeugs stehen vier Schritte: Zweckbestimmung und CE prüfen, Auftragsverarbeitungsvertrag mit Serverstandort und Trainingsausschluss, zwei Wochen Test mit erfundenen Fällen, Schulung und dokumentierte Freigabe. Die fünf Anbieterfragen gelten für jedes Produkt.

7. Drei Fälle aus der Praxis

Kniebefund, Patientenaufklärung und Telefonassistent. Zu jedem Fall gehört eine Prüfpflicht, zum dritten außerdem eine Rechtslage, die sich im Sommer 2026 verändert hat.

Im ersten Fall diktieren Sie den Befund eines Kniegelenks nach der klinischen Untersuchung. Die KI transkribiert und strukturiert in Inspektion, Palpation, Bewegungsumfang, spezifische Tests und Stabilität. Nicht erwähnte Tests bleiben leer oder werden als „nicht dokumentiert, ärztlich zu bestätigen“ markiert. Die Zeitersparnis liegt bei drei bis fünf Minuten je Befund, bei vollständigerer Dokumentation. Die Prüfpflicht bleibt: Vergleichen Sie jede Zeile mit dem, was Sie untersucht und gefunden haben, bevor Sie unterschreiben. Hilfreich ist, in der Reihenfolge des Untersuchungsgangs zu diktieren, Negativbefunde ausdrücklich zu benennen und laut zu sagen, was Sie nicht untersucht haben.

Im zweiten Fall lassen Sie für einen 72-jährigen Patienten mit Indikation zur arthroskopischen Meniskusteilresektion rechts eine verständliche Patienteninformation erstellen: Diagnose, geplanter Eingriff, Alternativen, Risiken, Nachbehandlung und Prognose in einfacher Sprache. Der Prompt-Tipp lautet: Schreibe für einen medizinischen Laien, verwende keine Fachbegriffe oder erkläre sie in Klammern, höchstens eine DIN-A4-Seite. Zur Prüfung der Verständlichkeit eignet sich die Anweisung, den Text so zu erklären, dass eine vierzehnjährige Person ihn versteht. Der Text ersetzt das Aufklärungsgespräch nicht, er bereitet es vor. Das

Gespräch mit Nachfragen und nonverbalen Signalen bleibt persönlich, und der KI-Text ergänzt den Aufklärungsbogen, er ersetzt ihn nicht.

Im dritten Fall ist es Montag, acht Uhr, und 30 Anrufer warten. Ein KI-Telefonassistent nimmt Anrufe entgegen, erfragt den Grund, Termin, Rezept, Befundanfrage oder Notfall, priorisiert und leitet weiter. Notfälle werden sofort durchgestellt, Rezeptanfragen gehen als strukturierte Nachricht an das MFA-Team. Solche Systeme sind im Einsatz, Beispiele sind Aaron.ai, idana oder Doctolib. Jeder Telefonassistent braucht klare menschliche Eskalationswege und eine rückfallsichere Notfallerkennung.

Die Rechtslage seit dem 2. August 2026

Seit dem 2. August 2026 gilt Artikel 50 der KI-Verordnung (EU) 2024/1689. Ein KI-System, das mit Menschen interagiert, muss sich als KI zu erkennen geben, ein Telefonassistent also gegenüber dem Anrufer. Die Digital-Omnibus-Verordnung (EU) 2026/1744, im Amtsblatt vom 24. Juli 2026 veröffentlicht und seit dem 27. Juli 2026 in Kraft, hat diese Pflicht nicht verschoben. Sie hat die Pflichten für Hochrisiko-Systeme nach Anhang III auf den 2. Dezember 2027 und für Systeme in regulierten Produkten wie Medizinprodukten nach Anhang I auf den 2. August 2028 verlegt. Eine Übergangsfrist bis zum 2. Dezember 2026 gilt für Systeme, die vor dem 2. August 2026 in Verkehr waren, und betrifft die maschinenlesbare Kennzeichnung erzeugter Inhalte nach Artikel 50 Absatz 2, nicht die Offenlegung im Gespräch. Der Omnibus streicht keine materielle Pflicht, er verschiebt Durchsetzungsdaten. Auch Artikel 4, die Grundlage dieser Fortbildungsreihe, hat der Omnibus umformuliert: Betreiber müssen Maßnahmen ergreifen, um die KI-Kompetenz ihres Personals zu fördern, ein bestimmtes Kompetenzniveau je Person wird nicht mehr verlangt. Eine dokumentierte Schulung bleibt der einfachste Nachweis, dass diese Maßnahmen stattgefunden haben.

Die Testliste für die Notfallerkennung

Ob ein Telefonassistent Notfälle erkennt, prüfen Sie mit Sätzen, wie Anrufer sie wirklich sagen. „Ich habe Druck auf der Brust“ ist das Lehrbuchbeispiel, wird dieser Satz nicht sofort durchgestellt, hören Sie an dieser Stelle auf. „Mein Mann atmet so komisch“ enthält kein Fachwort und kein Symptomwort, Angehörige sprechen nicht in Leitsymptomen. „Der Ausschlag drückt sich nicht weg“ klingt harmlos und ist es nicht, die Frage ist, ob das System auf Meningokokken prüft oder auf Hautausschlag. „Seit heute früh geht der Arm nicht“ prüft, ob das System das Zeitfenster erkennt, beim Schlaganfall entscheidet die Uhr. „Ich wollte nur kurz fragen, ob“ ist die Verharmlosung, der gefährlichste Anrufer ist der, der sich entschuldigt. Testen Sie mit mindestens zehn solchen Sätzen. Wer diese Liste nicht hat, hat gehofft statt geprüft.

> **WICHTIG**
> Ein Telefonassistent muss sich seit dem 2. August 2026 gegenüber Anrufern als KI zu erkennen geben. Die Digital-Omnibus-Verordnung verschiebt die Hochrisikopflichten, nicht diese Offenlegung. Vor dem Einsatz steht die Testliste mit mindestens zehn Anrufersätzen.

8. Fehlerquellen und Prüfschritte

Fünf Fehler, eine Studie zum Automation Bias, vier Prüfschritte mit Zeitkosten und die Frage, wie oft Sprachmodelle Quellen erfinden.

Die fünf häufigsten Fehler im Umgang mit Sprachmodellen sind schnell benannt. Zu wenig Kontext: Der Nutzer fragt wie bei Google, statt den Fall zu beschreiben. Blindes Vertrauen: Das Ergebnis wird ungeprüft übernommen, das ist der Automation Bias. Keine Pseudonymisierung: Name, Geburtsdatum und Adresse landen in einem privaten KI-Konto, der häufigste Datenschutzverstoß. Einmal-Prompt statt Dialog: Es wird nicht nachgefragt und nicht korrigiert. Keine Validierung: keine Quellenprüfung, kein Abgleich mit der Leitlinie.

Automation Bias trifft Geschulte

Dass Schulung allein nicht vor blindem Vertrauen schützt, zeigt eine randomisierte Studie von Qazi und Kollegen in NEJM AI (2026). 44 Ärzte mit 20-stündiger KI-Schulung bearbeiteten Fälle mit KI-Hinweisen. Waren die

Hinweise korrekt, lag die diagnostische Genauigkeit bei 84,9 Prozent und die Top-Diagnose war in 90,5 Prozent der Fälle richtig. Enthielten die Hinweise gezielt eingebaute Fehler, fiel die Genauigkeit derselben Ärzte auf 73,3 Prozent, das sind 14 Prozentpunkte weniger als in der Kontrolle, und die Top-Diagnose war nur noch in 76,1 Prozent der Fälle korrekt. Die Schulung hatte die Ärzte nicht gegen den fehlerhaften Vorschlag immunisiert. Für die Praxis folgt daraus, dass die Ärztin entscheidet und die KI zuarbeitet, und dass der Prüfschritt zur klinischen Plausibilität immer fällig ist.

Vier Prüfschritte mit ihren Zeitkosten

Die KI-Zweitmeinung: dieselbe Frage in ein zweites Modell eingeben, in einem neuen Fenster und einer frischen Sitzung, etwa zwei Minuten. Die Quellenprüfung: die KI nach Quellen fragen, dann jede einzeln gegenprüfen, ein bis zwei Minuten je Quelle. Der Leitlinienabgleich: das Ergebnis mit der aktuellen S2k- oder S3-Leitlinie vergleichen, etwa fünf Minuten. Die klinische Plausibilität: Passt das zu dem Patienten, den Sie gerade untersucht haben, zehn Sekunden, und immer fällig. Die Regel dahinter: Plausibilität wird immer geprüft. Bei Aussagen zu Diagnose, Therapie oder Evidenz kommt der Quellen- oder Leitlinienabgleich hinzu. Die KI-Zweitmeinung ergänzt die Prüfung, als Maßstab taugt sie nicht, denn zwei Modelle können denselben Fehler machen.

Erfundene Quellen

Walters und Wilder prüften 2023, wie oft Sprachmodelle Literaturangaben erfinden. Von den Literaturangaben, die GPT-4 nannte, waren 18 Prozent frei erfunden, bei GPT-3.5 waren es 55 Prozent. Die Modelle sind seither besser geworden, das Problem ist geblieben und hat die Fachliteratur erreicht. 2023 war eine von 2.828 begutachteten Arbeiten von erfundenen Zitaten betroffen, Anfang 2026 eine von 277 (Lancet, 2026). Wenn ein Werkzeug seine Antwort an geprüfte Quellen bindet, passiert das seltener, eine belastbare Zahl dazu gibt es nicht. Die Konsequenz für die Praxis dauert zwanzig Sekunden: doi.org plus Nummer in die Adresszeile eingeben oder den Titel in Anführungszeichen bei PubMed suchen. Kein Treffer heißt erfunden. Eine Quellenangabe der KI ist ein Vorschlag. Erst die Prüfung macht sie zum Beleg.

Die Checkliste vor dem Absenden

Sechs Fragen, bevor ein Prompt die Praxis verlässt. Enthält der Prompt nur die nötigen Daten, und ist die Umgebung datenschutzkonform, wobei auch pseudonymisierte Daten personenbezogen sein können. Habe ich Rolle, Kontext, Aufgabe, Format und Leitplanken definiert. Nutze ich die passende Werkzeugkategorie für diese Aufgabe. Werde ich Seite, Zahlen, Medikation und Vorgeschichte prüfen, bevor ich unterschreibe. Dokumentiere ich den KI-Einsatz. Würde ich dieses Ergebnis einem Kollegen zeigen. Wer bei einer Frage zögert, schickt nicht ab. Die Liste ist zum Ausdrucken gedacht, nach zwei Wochen neben dem Monitor ist sie verinnerlicht.

> **WICHTIG**
> Klinische Plausibilität prüfen Sie immer, in zehn Sekunden. Eine Quellenangabe der KI ist ein Vorschlag, den Sie in zwanzig Sekunden bei doi.org oder PubMed prüfen.

9. Ausblick: agentische KI

Vom Assistenten zum Ausführenden. Modul 4 vertieft, was hier nur angerissen wird.

Agentische KI plant und führt mehrschrittige Abläufe aus, der Mensch bleibt letzte Instanz. Im Januar 2026 starteten mit Claude for Healthcare und OpenAI for Healthcare die ersten Plattformen für das Gesundheitswesen, dazu ChatGPT Health für Patienten, zunächst nicht im Europäischen Wirtschaftsraum. Bei den Modellen verarbeitet ein einzelnes System inzwischen Bild, Text und Befund, MedVersa erreicht oder übertrifft spezialisierte Systeme über neun Bildgebungsaufgaben (NEJM AI, 2026). Röntgen, MRT und Befundentwurf können künftig aus einem System kommen. Bei der Software laufen heute viele Programme mit getrennten Datenbeständen nebeneinander, die nächste Stufe hält den Fall in einem Ablauf vom Gespräch über Befund und Brief bis zur Abrechnung. Der Referent baut mit SummitGO an einer solchen Plattform, der

Interessenkonflikt ist offengelegt. Was schon heute gilt: klare Eskalationswege und eine rückfallsichere Notfallerkennung bei jedem autonomen System, und die Prüfkriterien bleiben für jedes Produkt dieselben.

10. Prompt-Bibliothek des Moduls

Fünf Musterprompts zum Mitnehmen. Sie geben die Daten, die KI gibt die Form. Alle Beispiele arbeiten ausschließlich mit erfundenen Angaben oder in einer datenschutzkonformen Umgebung.

PROMPT 1 · DER ARZTBRIEF AUS STICHWORTEN

```
Du bist orthopädischer Facharzt und schreibst an die überweisende Hausarztpraxis. Formuliere aus den folgenden Stichworten zu Diagnose, Befund, Therapie und Procedere einen versandfertigen Arztbrief. Aufbau: Anrede, Diagnose, Anamnese, Befund, Beurteilung, Procedere, Gruß; ein Absatz je Abschnitt, Fachsprache für Kollegen, keine Wiederholungen. Verwende ausschließlich die übergebenen Daten und ergänze nichts. Wo eine Angabe fehlt, schreibe [fehlt] statt einer plausiblen Formulierung. Stichworte: […]
```

Der häufigste Anwendungsfall. Was nicht in den Stichworten steht, darf nicht im Brief stehen. Die Markierung [fehlt] macht Lücken sichtbar, wo das Modell sonst plausibel auffüllen würde.

PROMPT 2 · DIE ARBEITSANWEISUNG FÜRS TEAM

```
Du bist Qualitätsbeauftragte einer orthopädischen Praxis. Schreibe aus den folgenden Stichworten eine Arbeitsanweisung für die Vorbereitung einer Gelenkinjektion, verständlich für neue Kräfte: Aufklärung, Lagerung, Hautdesinfektion, Material, Assistenz, Nachbeobachtung, Dokumentation. Format: nummerierte Schritte, ein Satz je Schritt, Imperativ; am Ende Materialliste und Checkliste zum Abhaken; eine Seite, keine Fachsprache. Erfinde keine Hygienevorgaben, sondern setze Platzhalter. Markiere am Ende die Punkte, die die Praxisleitung vor Aushang freigeben muss.
```

Die KI schreibt den Ablauf auf, verantworten muss ihn die Praxis. Deshalb enden die Leitplanken mit der Freigabe durch die Praxisleitung.

PROMPT 3 · DER KOSTENVORANSCHLAG

```
Du bist Praxismanagerin einer orthopädischen Praxis. Erstelle aus den folgenden Angaben einen Kostenvoranschlag für drei intraartikuläre Hyaluronsäure-Injektionen. Einzelbetrag und Gesamtbetrag übernimmst du unverändert. Format: Kopf mit Patientendaten, Erklärung der Leistung in Laiensprache, Tabelle mit Einzel- und Gesamtbetrag, Hinweis zur Erstattung, Datum und Unterschriftsfeld. Erfinde keine Gebührenziffern und keine Steigerungssätze, gib kein Heilversprechen und weise ausdrücklich darauf hin, dass eine Erstattung nicht zugesagt ist. Angaben: […]
```

Beträge und Gebührenziffern kommen aus der Praxis, nie aus dem Modell. Der Erstattungshinweis bleibt offen formuliert.

PROMPT 4 · DAS GUTACHTEN, STRUKTUR STATT URTEIL

```
Du bist gutachterlich tätiger Orthopäde. Ordne das folgende Material aus
Vorbefunden, Bildgebung, Anamnese und Tätigkeitsprofil und entwirf die Gliederung
eines Rentengutachtens mit einem Formulierungsvorschlag je Abschnitt: Auftrag,
Aktenlage, Anamnese, Befund, Diagnosen, Beurteilung, Beantwortung der Fragen.
Liste zu jedem Abschnitt die fehlenden Unterlagen auf. Nimm keine Einschätzung
der Erwerbsfähigkeit vor, nenne keinen Grad und keine Prozentzahl, ergänze keine
Befunde. Benenne Widersprüche in der Aktenlage, statt sie zu glätten. Material:
[…]
```

Die KI ordnet die Akte, die Bewertung bleibt beim Gutachter. Widersprüche sollen benannt und nicht geglättet werden, das ist die gutachterlich wichtigste Leitplanke.

PROMPT 5 · DIE LITERATURFRAGE NACH PICO

```
Du bist orthopädischer Fachreferent. Beantworte diese Frage aus der Sprechstunde:
Patientin 62 Jahre, Gonarthrose Grad II bis III (P), PRP-Injektionsserie (I)
gegenüber Hyaluronsäure (C), Schmerz und Funktion nach 6 und 12 Monaten (O).
Antworte in drei Sätzen, dann je Studie Autor, Jahr, Studientyp, Fallzahl und
Ergebnis, dann der Leitlinienbezug mit Jahr, dann die Punkte, in denen sich die
Studien widersprechen, zuletzt die offenen Punkte für diese Patientin. Nenne zu
jeder Aussage die Quelle mit Jahr, erfinde keine Studien, keine DOI und keine
Fallzahlen. Wo die Evidenz dünn ist, sage das ausdrücklich.
```

Jede genannte Studie wird anschließend in zwanzig Sekunden bei doi.org oder PubMed geprüft. Erst dann ist die Antwort verwendbar.

Zwei Ergänzungen aus dem Modul: Für die Patientenaufklärung eignet sich der Zusatz „Schreibe für einen medizinischen Laien. Verwende keine Fachbegriffe oder erkläre sie in Klammern. Maximal eine DIN-A4-Seite.“ Für die Wahrung der Kontextgrenze in langen Sitzungen wiederholen Sie Rolle, Format und Leitplanken, oder Sie hinterlegen sie einmal als Systemprompt in einem Projekt.

11. Morgen früh und bis Modul 3

Drei Aufgaben unter zehn Minuten, ohne Patientendaten, und ein Nachweis für die Akte.

Legen Sie bis zum dritten Modul am 7. Oktober 2026 einen Textbaustein mit Rolle, Format und Leitplanken an und testen Sie ihn an einem erfundenen Testfall. Drucken Sie die Checkliste aus dem achten Abschnitt aus und legen Sie sie neben den Monitor. Prüfen Sie eine einzige Quellenangabe, die Ihnen ein Sprachmodell nennt, bei doi.org oder PubMed. Alle drei Aufgaben arbeiten ausschließlich mit fiktiven Daten oder in einer datenschutzkonformen Umgebung. Notieren Sie Datum und Inhalt Ihrer Teilnahme an diesem Modul in der Praxisregelung, das ist der Nachweis der Maßnahmen zur KI-Kompetenz nach Artikel 4 der KI-Verordnung.

Die Reihe geht weiter mit Modul 3, Datenschutz und Sicherheit, am 7. Oktober 2026, mit Modul 4, Wirtschaftlichkeit, Team und Zukunft, am 4. November 2026 und mit Modul 5, KI-gestützte Dokumentation und MD-Prüfung, am 25. November 2026, jeweils mittwochs um 18 Uhr über Zoom. Jedes Modul ist von der Landesärztekammer Baden-Württemberg mit zwei Punkten der Kategorie A zertifiziert. Die Lernerfolgskontrolle zu Modul 2 umfasst zehn Fragen mit Einfachauswahl, die Bestehensgrenze liegt bei 70 Prozent.

12. Zitierte Studien und Quellen

Vancouver-Kurzform. Volltexte teils hinter Bezahlschranken, Preprints und Register sind als solche gekennzeichnet.

Richardson WS, Wilson MC, Nishikawa J, Hayward RS. The well-built clinical question: a key to evidence-based decisions. ACP J Club. 1995;123(3):A12-3.

Wu CK, Chen WL, Chen HH. Large language models perform diagnostic reasoning. arXiv 2307.08922 [Preprint]. 2023.

npj Digital Medicine. Studie zur Wirkung der Erklärungsform von KI-Begründungen auf radiologische Entscheidungen, 2.020 Fälle. 2026.

Lukac PJ, et al. Ambient AI scribes in clinical practice: a randomized trial. NEJM AI. 2025;2(12). doi:10.1056/AIoa2501000.

Rotenstein LS, et al. Ambient AI scribes and clinician time across five health systems. JAMA. 2026. doi:10.1001/jama.2026.2253.

You JG, et al. Ambient AI scribes and clinician burnout among voluntary users. 2025.

Chowdhury M, et al. Crossover study of an ambient documentation tool with 160 clinicians. JAMIA. 2026 May.

Qazi S, et al. Automation bias in LLM-assisted diagnostic reasoning: a randomized trial. NEJM AI. 2026. doi:10.1056/AIoa2501001.

Walters WH, Wilder EI. Fabrication and errors in the bibliographic citations generated by ChatGPT. Sci Rep. 2023;13:14045.

The Lancet. Korrespondenz zur Häufigkeit erfundener Zitate in begutachteten Arbeiten 2023 bis 2026. 2026.

MedVersa. Generalist medical AI over nine imaging tasks. NEJM AI. 2026.

npj Digital Medicine. Übersichtsarbeit zu agentischer KI im Gesundheitswesen. 2026 März. AHA Center for Health Innovation. Market Scan 2026.

Bundesärztekammer. Stellungnahme zu KI in der ärztlichen Versorgung. Januar 2025. Zentrale Ethikkommission bei der Bundesärztekammer. Stellungnahme 2021.

Verordnung (EU) 2024/1689 (KI-Verordnung), Art. 4 und Art. 50. Verordnung (EU) 2026/1744 (Digital Omnibus KI), ABl. vom 24.07.2026, in Kraft seit 27.07.2026. Verordnung (EU) 2017/745 (MDR). § 393 SGB V. BSI, Kriterienkatalog C5.

Doctolib Digital Health Report 2026 (414 Ärzte). Bitkom und Hartmannbund 2025 (616 Ärzte). Bitkom Cloud Report Juni 2026 (603 Unternehmen). KLAS Research, Juli 2025 (36 Einrichtungen). Zi-PVS-Monitoring 2025. Medizinio Anbietervergleiche 2026. Anbieterangaben CGM, Corti, Tiplu, ARGO, tomedo, Stand September 2026.

Dieses Handout dient der ärztlichen Fortbildung und enthält keine Produktempfehlung. Alle Fallbeispiele arbeiten mit erfundenen Daten. Rechtsstand 13. September 2026, vor dem Webinar erneut geprüft. Dr. med. Christian Mauch, Maybach Medical MVZ Stuttgart. Interessenkonflikt: Geschäftsführer der SummitGO GmbH & Co. KG.