

KI in der Arztpraxis: Datenschutz, lokale Verarbeitung und Programmauswahl

Erweiterte Zusammenfassung mit aktuellen Beispielen für lokale und cloudbasierte Lösungen

1. Die drei Kernfragen

- Sind die verwendeten Informationen noch personenbezogen?
- Wo und durch wen werden sie verarbeitet?
- Welche technische Lösung ist für den Praxisalltag sinnvoll?

Die Diskussion wird klarer, wenn Datenschutz, Anonymisierung und technische Architektur getrennt betrachtet werden.

2. Anonymisierung: „Name entfernt“ reicht nicht zwingend

Das Entfernen von Namen und Adressen aus einem Arztbrief bedeutet nicht automatisch, dass die verbleibenden Informationen anonymisiert sind. Gerade bei seltenen Erkrankungen können Alter, Diagnose, Eingriffe, Behandlungsort, Fachgebiet und zeitlicher Zusammenhang zusammen eine Person identifizierbar machen.

Entscheidend ist nicht, ob die KI selbst Zugriff auf ein Melderegister hat, sondern ob eine Identifizierung mit Mitteln möglich ist, deren Einsatz vernünftigerweise in Betracht kommt. Echte Anonymisierung ist von Pseudonymisierung zu unterscheiden; pseudonymisierte Daten bleiben personenbezogene Daten.

Beispiel: „5-jähriger Patient mit Fieber und Husten“ ist typischerweise wenig spezifisch. Eine Kombination aus seltenem Syndrom, Alter, mehreren Operationen und einer spezialisierten Einrichtung kann dagegen wesentlich identifizierbarer sein.

3. Technische Anonymisierung von Dokumenten

Eine sichtbare Schwärzung in einem PDF ist nicht automatisch eine technische Entfernung des darunterliegenden Textes. Je nach Erzeugung des PDFs kann der Text weiterhin per Copy-and-Paste, Textextraktion oder OCR auslesbar sein. Vor einer externen Verarbeitung sollte deshalb geprüft werden, was technisch tatsächlich noch im Dokument steckt.

Zu berücksichtigen sind außerdem Dateinamen, Metadaten, E-Mail-Adressen, eingebettete Bilder, OCR-Text sowie Namen oder Adressen, die auf einzelnen Seiten übersehen wurden.

4. Deutschland versus USA: nicht allein entscheidend

Bei personenbezogenen Gesundheitsdaten sind unter anderem Verantwortlicher, Auftragsverarbeiter, Rechtsgrundlage, Art. 9 DSGVO, AVV, Unterauftragsverarbeiter, Speicher- und Verarbeitungsorte, Drittlandtransfers, TOMs, Löschfristen und die Nutzung für Training oder eigene Zwecke des Anbieters zu prüfen.

Eine Verarbeitung in den USA ist nicht allein aufgrund des Standortes automatisch DSGVO-widrig. Umgekehrt macht ein deutscher Serverstandort eine Verarbeitung nicht automatisch DSGVO-konform. Bei Cloud-Diensten muss die gesamte Verarbeitungskette betrachtet werden.

5. Lokale Verarbeitung: das technisch einfachste Datenschutzmodell

Mikrofon → Praxis-PC → lokales Spracherkennungsmodell → Text

Wenn Audio und Transkript tatsächlich vollständig auf der eigenen Hardware verarbeitet werden und keine Cloud-, Synchronisations- oder Telemetriedienste beteiligt sind, werden die Patientendaten für

diesen Verarbeitungsschritt nicht an einen externen KI-Anbieter übermittelt.

Das ist eine andere Ausgangslage als bei einer Cloud-KI. Allerdings muss geprüft werden, ob die konkrete Anwendung tatsächlich vollständig lokal arbeitet; „lokales Modell“ allein garantiert nicht, dass die gesamte Anwendung offline bleibt.

6. Konkrete Programme und Ansätze

OpenWhispr

Typ: lokales Diktierprogramm. Stärke: Open Source, macOS/Windows/Linux, lokale Modelle wie Whisper und Parakeet. Im lokalen Modus wird das Audio laut Anbieter vollständig auf dem Gerät verarbeitet; keine Internetverbindung ist erforderlich. Cloud- und eigene-API-Modi sind separat und müssen in einer Praxis bewusst vermieden werden, wenn Patientendaten das Gerät nicht verlassen sollen. Transkriptionen können lokal bleiben; Cloud-Synchronisierung ist optional.

Einordnung: Geeignet als lokaler Ersatz für cloudbasierte Diktier-Tools; besonders interessant, wenn unterschiedliche lokale Modelle getestet werden sollen.

Quelle: <https://openwhispr.com/de/security>

Asklaion Typer

Typ: systemweites lokales Diktat. Per Tastenkombination wird Sprache direkt in das aktuell fokussierte Textfeld geschrieben, z. B. PVS, Browser oder Word. Laut Anbieter läuft Whisper vollständig lokal; Audio- und Textdaten werden nicht an einen Server übertragen. Eine GPU wird benötigt; eine einfache bzw. integrierte GPU kann laut Anbieter ausreichen.

Einordnung: Sehr interessant für den klassischen Diktat-Workflow ohne Cloud. Open Source und MIT-lizenziert.

Quelle: <https://asklaion.de/forge/asklaion-typer/>

Asklaion Medical Scribe

Typ: lokaler Medical Scribe. Arzt-Patienten-Gespräche werden lokal mit Whisper transkribiert; anschließend kann aus dem Transkript per konfigurierbarem Prompt eine Zusammenfassung erzeugt und ins PVS übernommen werden. Laut Anbieter läuft die Verarbeitung vollständig lokal. Für den Realbetrieb wird eine dedizierte GPU mit möglichst viel VRAM empfohlen; genannt werden mindestens 8 GB VRAM und CUDA-Unterstützung.

Einordnung: Interessant, wenn nicht nur diktiert, sondern ein komplettes Gespräch automatisch dokumentiert und strukturiert werden soll. Software ist frei/Open Source; Hardware kann der wesentliche Kostenfaktor sein.

Quelle: <https://asklaion.de/forge/medical-scribe/>

CuraSpeech

Typ: medizinisches Diktat mit lokaler KI. Der Anbieter beschreibt lokale Transkription ohne Cloud, medizinische Terminologie, individuelles Wörterbuch und lokale KI für Text- und Formularverarbeitung. Zusätzlich wird eine Chatfunktion mit lokaler KI und Dokumentanalyse angeboten; eine FHIR-Anbindung an T2med wird genannt.

Einordnung: Besonders interessant, wenn medizinische Terminologie und PVS-/FHIR-Integration wichtiger sind als ein möglichst allgemeines Diktierwerkzeug.

Quelle: <https://www.cura-sync.de/curaspeech.html>

ScaleWise Local

Typ: umfassendere lokale Praxis-KI. Laut Anbieter läuft das System auf eigener Hardware im Praxisnetz und auch ohne Internet. Vorgesehen sind u. a. Diktat, Dokumentensuche, Antworten mit Quellenangabe und Arbeit mit Patientenakten/Befunden.

Einordnung: Eher eine lokale KI-Infrastruktur bzw. Wissenszentrale als nur ein Diktierprogramm. Interessant für Praxen, die neben Diktat auch lokale Dokumentanalyse und einen KI-Assistenten benötigen.

Quelle: <https://www.scalewise-ai.de/lokale-ki-arztpraxis>

Whisper / whisper.cpp direkt lokal

Typ: technischer Baukasten statt fertiges Praxisprodukt. Whisper-Modelle können mit whisper.cpp lokal betrieben werden. Das bietet maximale Kontrolle über Datenfluss und Modellwahl, erfordert aber deutlich mehr technischen Eigenaufwand für Oberfläche, Diktat-Integration, Wörterbuch, Updates und PVS-Anbindung.

Einordnung: Interessant für technisch versierte Praxen oder IT-Betreiber, die eine vollständig kontrollierte Eigenlösung aufbauen wollen.

Quelle: <https://github.com/ggerganov/whisper.cpp>

VoicePad AI

Typ: lokales bzw. gerätebasiertes Diktat-Konzept. Vor einem produktiven Praxiseinsatz sollten konkrete aktuelle Version, lokale Modellverarbeitung, Telemetrie und Datenfluss geprüft werden.

Einordnung: Als Kandidat für einen Vergleichstest interessant; für medizinische Patientendaten sollte die lokale Verarbeitung vor Einsatz technisch verifiziert werden.

Quelle: Aktuelle Anbieter-/Projektunterlagen vor Praxiseinsatz prüfen.

Dragon Medical One

Typ: professionelle medizinische Spracherkennung als Cloud-Dienst. Nicht mit einer vollständig lokalen Whisper-Lösung gleichzusetzen. Für Praxen kann die medizinische Spezialisierung attraktiv sein, aber der Datenschutz ist über die Cloud-Architektur, Verträge und konkrete Verarbeitung zu bewerten.

Einordnung: Interessanter Vergleichsmaßstab für Erkennungsqualität und medizinische Terminologie, aber kein Offline-Modell im engeren Sinn.

Quelle: Anbieterunterlagen und aktuelle Vertrags-/Datenschutzbedingungen prüfen.

7. Sonderfall Wispr Flow: sehr komfortabel, aber nicht lokal

Wispr Flow ist für den Praxisvergleich besonders interessant, weil es genau den komfortablen Workflow bietet, den viele Nutzer von einem modernen Diktierprogramm erwarten: Tastenkürzel bzw. Sprachaufnahme, anschließend wird der Text direkt in die aktuelle Anwendung eingefügt. Es läuft auf Mac, Windows, iOS und Android und unterstützt viele Sprachen.

Der entscheidende Unterschied zu den lokalen Lösungen: Wispr Flow verarbeitet die Diktierdaten derzeit in der Cloud. Laut aktueller Unternehmensdokumentation gibt es keine On-Premises-, Private-Tenant- oder Offline-Variante. Kundendaten werden in den USA verarbeitet und gespeichert; als primäre Region wird Northern Virginia, mit Failover in Northern California, genannt. Eine EU-Datenresidenz wird derzeit nicht angeboten.

Für EU-Kunden stellt Wispr einen DPA bereit, der unter anderem EU-Standardvertragsklauseln und eine Liste der Unterauftragsverarbeiter enthält. Wispr weist außerdem darauf hin, dass der Verantwortliche im Rahmen des SCC-Modells die eigene Transfer Impact Assessment durchführen muss. Das ist also nicht dasselbe wie eine vollständig lokale Verarbeitung.

Positiv aus Unternehmenssicht: Wispr bietet Datenschutz- und Administrationsfunktionen, DPA/SCCs sowie die Möglichkeit, Modelltraining abzulehnen. Für Diktat gibt es außerdem eine HIPAA-geeignete Konfiguration mit unterschriebenem BAA. Diese Punkte können für eine rechtliche Prüfung relevant sein, ändern aber nichts daran, dass die Verarbeitung des Diktats cloudbasiert und derzeit in den USA erfolgt.

Für eine deutsche Arztpraxis lautet die entscheidende Frage daher nicht, ob Wispr technisch gut funktioniert, sondern ob die konkrete Verarbeitung personenbezogener Gesundheitsdaten über die US-Cloud im eigenen Datenschutzkonzept akzeptabel und rechtlich dokumentierbar ist. Wer die Vorgabe

„Patientendaten verlassen die Praxis nicht“ hat, kann Wispr Flow dafür nicht einsetzen.

Kosten: Die aktuelle Preisstruktur nennt für Pro 15 US-Dollar pro Nutzer und Monat bzw. 12 US-Dollar bei jährlicher Abrechnung. Für Teams beginnt Growth bei 23 US-Dollar pro Nutzer und Monat bzw. 18 US-Dollar bei jährlicher Abrechnung. Die Preise können sich ändern.

Quellen: <https://wisprflow.ai/enterprise-privacy> · <https://docs.wisprflow.ai/articles/4712082205-Requesting-a-Data-Processing-Addendum-%28DPA%29> · <https://wisprflow.ai/pricing>

8. Warum Wispr trotzdem ein Sonderfall ist

Wispr Flow ist kein schlechtes Beispiel für einen Diktier-Workflow, sondern ein gutes Beispiel dafür, dass Benutzerfreundlichkeit und Datenarchitektur zwei verschiedene Fragen sind. Der Dienst kann beim Diktieren sehr komfortabel sein, weil die eigentliche Spracherkennung serverseitig erfolgt und dadurch leistungsfähige Modelle eingesetzt werden können, ohne dass die Praxis dafür eine eigene GPU bereitstellen muss.

Bei einer lokalen Lösung verschiebt sich der Aufwand dagegen von der Cloud auf die eigene Hardware: GPU, VRAM, Installation, Updates, lokale Modellwahl und gegebenenfalls PVS-Integration müssen selbst organisiert werden. Gerade bei einem Medical Scribe kann dieser Unterschied erheblich sein.

Damit ergibt sich ein klarer Zielkonflikt: Cloud = weniger lokale Hardware und häufig sehr komfortabler Workflow; lokal = maximale Kontrolle über den Datenfluss, aber eigener Hardware- und Betriebsaufwand.

9. Kosten: Software ist nicht gleich Gesamtkosten

Bei lokalen Open-Source-Lösungen kann die Software selbst kostenlos sein. Die Gesamtkosten können trotzdem aus Hardware, Installation, Wartung, Updates, Datensicherung und gegebenenfalls IT-Betreuung bestehen.

Ein anschauliches Beispiel ist ein lokaler Medical Scribe: Für die Software fällt keine klassische monatliche Cloud-Gebühr an, für einen produktiven Betrieb mit größeren Modellen kann jedoch eine leistungsfähige GPU erforderlich sein. Es gibt Erfahrungsberichte über KI-Server in der Größenordnung von etwa 5.000 €, die bei längeren Gesprächen bzw. mehreren gleichzeitig arbeitenden Räumen bereits an Leistungsgrenzen kamen. Das ist kein allgemeingültiger Hardwarebedarf, verdeutlicht aber die Größenordnung, die bei lokaler KI entstehen kann.

10. Praktischer Vergleich

Lösung	Lokal?	Primärer Zweck	Besonderheit
OpenWhispr	Ja, optional Cloud	Diktat	Open Source; mehrere lokale Modelle
Asklaion Typer	Ja	Diktat	Systemweit; direkt ins aktive Textfeld
Asklaion Medical Scribe	Ja	Gespräch → Dokumentation	Lokale Transkription + lokales LLM; GPU empfohlen
CuraSpeech	Ja, laut Anbieter	Medizinisches Diktat	Terminologie, lokale KI, FHIR/T2med
ScaleWise Local	Ja	Praxis-KI	Dokumente, Akten, Diktat, lokaler Assistent
Whisper/whisper.cpp	Ja	Spracherkennung	Maximale technische Kontrolle, mehr Eigenaufwand
Wispr Flow	Nein	Diktat	Sehr komfortabel, aber Cloud/USA; keine Offline-Variante
Dragon Medical One	Nein	medizinisches Diktat	Medizinische Spezialisierung, Cloud-Architektur

11. Welche Lösung für welchen Zweck?

- Nur Diktieren: OpenWhisper, Asklaiion Typer oder eine direkte lokale Whisper-/whisper.cpp-Lösung.
- Diktieren mit stärkerem medizinischem Fokus: CuraSpeech als Kandidat für einen Praxistest.
- Gespräch automatisch dokumentieren: Asklaiion Medical Scribe; hier ist die Hardwarefrage deutlich wichtiger.
- Lokale KI über Diktat hinaus: ScaleWise Local als umfassenderer Ansatz für Dokumente, Wissen und Praxisassistent.
- Maximal komfortables Cloud-Diktat: Wispr Flow; dafür muss die US-Cloud-Verarbeitung bewusst akzeptiert und datenschutzrechtlich geprüft werden.
- Medizinische Cloud-Spracherkennung: Dragon Medical One als klassischer professioneller Vergleichskandidat, aber nicht als Offline-Lösung.

12. Auswahlkriterien für eine Praxis

1. Verlässt Audio den Praxisrechner tatsächlich nicht?
2. Bleibt auch das Transkript lokal?
3. Gibt es Cloud-Synchronisierung oder Telemetrie?
4. Welche lokalen Modelle stehen zur Verfügung?
5. Wie gut ist die Erkennung von deutschem medizinischem Fachvokabular?
6. Welche GPU und wie viel VRAM werden benötigt?
7. Funktioniert das System ohne Internet?
8. Wie werden Updates eingespielt?
9. Wie funktioniert die Integration in das PVS?
10. Werden Daten für Training oder andere Zwecke des Anbieters verwendet?
11. Falls Cloud: Wo werden Daten verarbeitet und gespeichert?
12. Welche Subprozessoren sind beteiligt?
13. Gibt es AVV/DPA und gegebenenfalls Regelungen für Drittlandtransfers?
14. Wie lange werden Audio und Transkripte gespeichert?

13. Fazit

Für die Auswahl einer KI in der Arztpraxis ist die einfache Gegenüberstellung „Deutschland gut, USA schlecht“ zu grob. Die wichtigere technische Trennlinie ist häufig vollständig lokale Verarbeitung versus externe Cloud-Verarbeitung.

Lokale Lösungen können den Datenfluss auf die Praxis beschränken, verlangen dafür aber eigene Rechenleistung und mehr technischen Betrieb. Cloud-Lösungen reduzieren den Hardwareaufwand und können sehr komfortabel sein, erfordern bei Patientendaten jedoch eine genaue Prüfung der gesamten Verarbeitungskette.

Besonders wichtig bleibt die Unterscheidung zwischen echter Anonymisierung und bloßem Entfernen von Namen. Gerade bei seltenen medizinischen Konstellationen kann die Kombination scheinbar harmloser Merkmale bereits einen Personenbezug herstellen.

Hinweis zur Einordnung

Die Angaben zu den einzelnen Produkten entsprechen dem bei der Erstellung dieser Fassung recherchierten Stand im September 2026 und beruhen überwiegend auf den Angaben der jeweiligen Anbieter. „Lokal“ bzw. „DSGVO-konform“ sind hier keine unabhängigen rechtlichen Zertifizierungen. Vor einem produktiven Einsatz mit Patientendaten sollten Datenfluss, konkrete Konfiguration, Verträge und die eigene datenschutzrechtliche Bewertung geprüft werden.