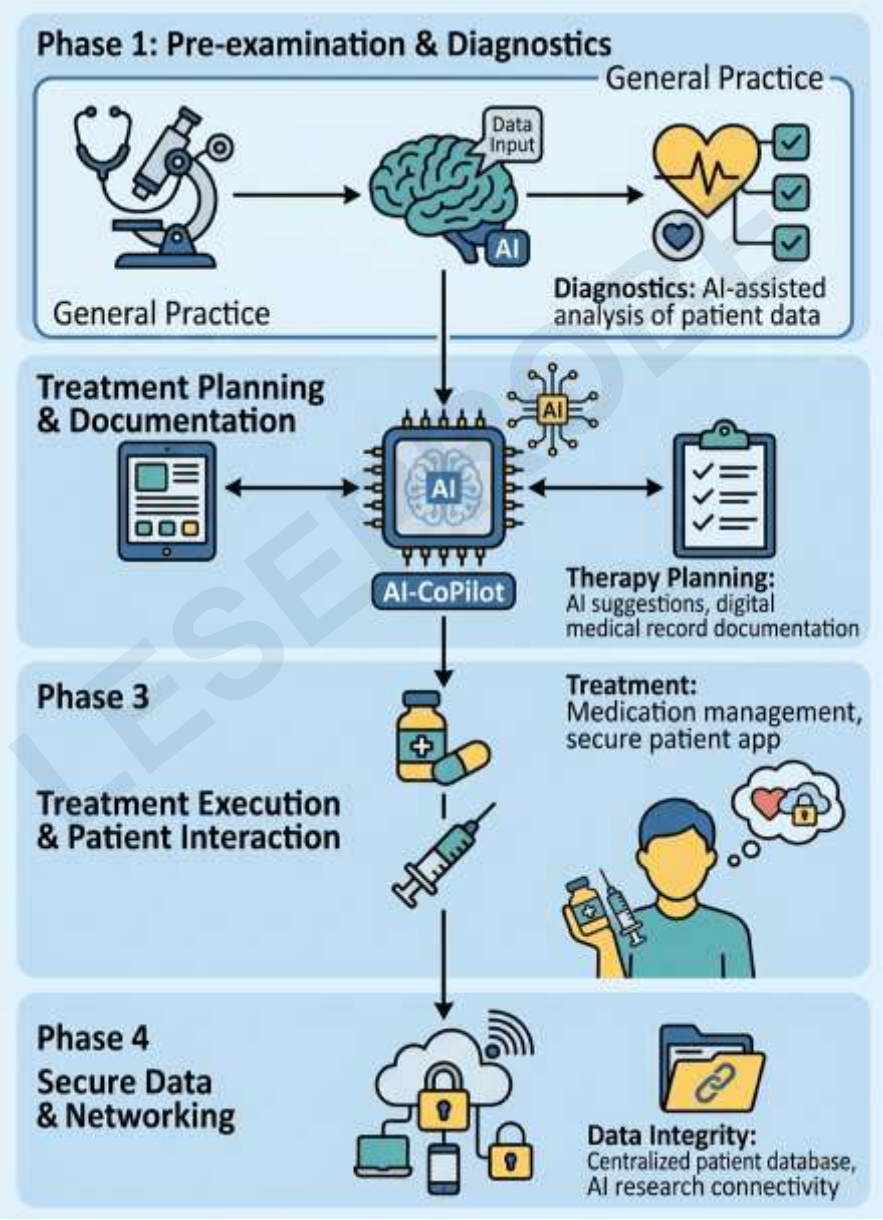
ClinicalOS – The clinical operating system for general practice

Volume 5

Digital Tools & AI in GP practice



Using AI – Medical Validation – Medical Accountability

From AI-assisted research to digital documentation – a practical guide to safe digital medicine

About this sample

Volume 5 · Digital Tools & AI

This curated sample contains 29 pages from a 193-page volume, including three complete chapters and selected introductory pages. Original page numbers are retained; gaps are intentional. The reproduced contents describe the full volume. PDF bookmarks lead only to included content.

- **Chapter 2 · How AI works**
- **Chapter 3 · Which tool for which purpose?**
- **Chapter 6 · AI-assisted documentation**

The original notices on use and responsibility also apply to this sample.

Copyright and Legal Notice

Dr Götz Huber, MD

The clinical operating system of general practice

Volume 5: Digital Tools & AI – Using AI – Medical Validation – Medical Accountability

Note: The information contained in this book has been carefully researched and verified. As AI tools and regulatory frameworks change rapidly, some details may already be out of date by the time you read this. You can find the latest updates at https://clinicalos.de/. The mention of products and manufacturers is without guarantee and does not constitute a recommendation to purchase or use them.

Note on Translation, Content, Use and Liability

This English edition is a translation of the original work written in German. The concept, structure, and core medical content of ClinicalOS were originally developed for German general practice and the German healthcare system, to provide practising GPs, trainee GPs, healthcare assistants (HCAs) and medical students with a structured, practice-oriented learning and reference resource. The content is based on current medical literature, guideline-based sources and practical experience in general practice.

During the linguistic adaptation, references to guidelines, clinical practices and regulatory frameworks in other English-speaking settings were taken into account and highlighted where relevant and feasible; in some areas, differences from German recommendations are also presented explicitly. However, the translation does not constitute a systematic adaptation of every medical, therapeutic or regulatory aspect to all English-speaking countries. Professional guidelines, marketing authorisations, approved dosages, prescribing procedures, reimbursement rules and regulatory requirements may vary between countries;

the guidelines, official prescribing information and regulatory provisions currently applicable in the respective country or healthcare system always take precedence. Dosage information applies to adults without significant renal impairment; individual adjustments are required for elderly patients, those with polypharmacy and during pregnancy. **Current national/regional guidelines (e.g. NICE, ESC, ADA) must be followed as a priority.**

Digital tools, including AI-assisted text generation, were used in the creation of this manual. All content was subsequently checked by the author for technical accuracy, consistency and potential errors. **No guarantee can be given for the correctness, currency or completeness of the information provided;** the complete absence of errors cannot be excluded, including with regard to guidelines or dosage recommendations that may have been updated in the meantime.

The contents of this manual do not replace an individual medical assessment and should not be construed as binding treatment recommendations. Diagnostic and therapeutic decisions must always be made on a case-by-case basis for each individual patient, taking into account the full clinical picture and the currently valid guidelines. **Doctors bear full and ultimate medical responsibility:** any use of content from this manual — whether as a text module, SOP or treatment recommendation — requires independent professional review by the treating doctor. **AI-supported or digitally generated content cannot and must not replace medical judgement.**

Any use of the information presented here is the sole responsibility of the reader. Liability claims relating to damage of a material or immaterial nature arising from the direct or indirect use of the information contained in this manual — even where such information is incomplete or contains errors — are generally excluded.

Contents

Section	Contents
	(Idana), video consultations (indications, billing, integration with healthcare assistants (HCA)), web-based therapies beyond DiGA (typology, evidence base, risks), data-driven logic rather than paper-based logic, care home/chronic patient coordination, DiGA in GP practice
Chapter 10: Web-based therapies (DiGA and generative AI for self-help)	DiGA in GP practice (5-step prescription routine, fact sheets, distinction from DMPs), web-based therapies and online programmes beyond DiGA (typology, evidence base, fact sheets), generative AI as a self-help tool for mental health conditions (suitability check, traffic light model, prompt library)
Chapter 11: Digital resources for GPs	Point-of-care systems, clinical scores as digital tools, interaction databases and medication tools, key URLs and resources, English-language CME platforms
Chapter 12: Telematics infrastructure — ePA, e-prescription and KIM	ePA for all (timeline and obligations), e-prescription and eML, KIM and TIM, overview of the practice software market 2025/26, TI readiness checklist, outlook: AI analyses in the ePA from 2026/27, TI friction in everyday practice (obligation/benefit/friction)
PART V – Security, Implementation, Workflows	
Chapter 13: Risks and Limitations of AI in Medicine	Identifying, avoiding and verifying hallucinations; automation bias; data protection (risk zones, 72-hour reporting obligation, team training); risk matrix; four-level confidence level model (RAG quality, prompt characteristics, personalisation)
Chapter 14: Regulatory Framework — What Medical Practices Need to Know	UK AI regulation (framework and obligations), MDR/IVDR distinction, KBV IT Security Guideline 2025, cloud and C5 certification (Section 393 of Book V of the Social Code), regulatory compliance checklist, AI competence within the practice team (three-tier responsibility, quick wins, training concept, tool checklist)
Chapter 15: Productive AI Workflows for Day-to-Day Practice	The 10 most common AI mistakes in practice, four workflow examples (guideline → patient information, hospital discharge summary → treatment plan, CME article → team training, practice meeting → task list), 12-month implementation plan, quick reference for the first day with AI, 12 productive AI micro-competencies, tool types and prompt logic, AI as a thinking partner

Table of Contents

Chapter 2: How AI works — the essentials for doctors

This chapter explains AI in the same way a clinical chemist would explain a blood gas analysis to you: enough to use it safely, without a lecture on molecular biology. You don't need to become an AI expert. You need to understand why AI is sometimes wrong — and how to recognise it.

2.1 Large Language Models (LLMs) in 5 Minutes

Almost all the AI tools relevant to GPs are based on so-called Large Language Models (LLMs). These include ChatGPT (OpenAI), Claude (Anthropic), Gemini (Google) and the models behind Perplexity, OpenEvidence and the Ambient AI Scribes.

What is an LLM?

An LLM is a statistical language model that has been trained on billions of text examples. It has learnt the probability with which one word follows another. This results in three characteristics that are crucial for clinical use:

Characteristic	Implications for clinical practice	Consistency
Pattern recognition	LLMs recognise text patterns and can organise, summarise and translate	Particularly useful for documentation, research and text generation
No world knowledge	LLMs do not 'understand' anything — they calculate probable word sequences	They may sound convincing but still be wrong
Depends on training data	Quality and timeliness depend on the training data	Medical information must always be verified

An analogy for everyday practice

Imagine an LLM as an extremely well-read but inexperienced junior doctor:

They've read everything — PubMed, textbooks, guidelines, even Reddit.

They are brilliant at summarising and articulating ideas.

But they've never worked in a practice. They lack clinical intuition.

He sometimes makes mistakes that he doesn't even realise he's made.

He needs supervision — just like a final-year medical student.

2.2 Hallucinations, bias, limitations of training data

AI systems have three sources of error that are particularly relevant in a clinical context:

Hallucinations

AI hallucinations are outputs that are convincingly worded but factually incorrect. The LLM 'invents' facts, studies or dosages that do not exist. The first randomised controlled trial on ambient AI scribes (UCLA RCT, n=238) showed that clinically relevant errors occur regularly — with a frequency score of 2.7–2.8 on a 5-point Likert scale (1 = never, 5 = very frequently; self-assessment by the participating doctors, not an objective measure of errors).

Bias (systematic distortion)

LLMs reproduce biases present in their training data. If US guidelines are over-represented, the responses tend to favour US medicine. If certain patient groups are under-represented in the training data, recommendations may be biased.

Limitations regarding timeliness

Every LLM has a knowledge cutoff — a cut-off date after which no new information has been incorporated into the training. For clinical questions, this means that the model may not be aware of recent changes to guidelines, new study results or amended regulatory approvals.

PROMPT EXAMPLE Source of error || Example in everyday practice || Countermeasure ||

Source of error Example in everyday practice Countermeasure Hallucination AI suggests a dosage not listed in the product information Always verify dosages against the product information or the ABDA-DB Bias AI recommends US screening intervals instead of German guidelines Cross-check research results against German guidelines (DEGAM, AWMF) Out-of-date information AI is unaware of a new NVL update For time-critical questions: Use tools with live search (Perplexity, OpenEvidence) Contextual errors AI ignores comorbidities in the summary Formulate the prompt precisely: 'Patient with stage 3 CKD and AFib' Convincingly incorrect AI cites a study that does not exist Check references manually (Pub

Verification: *Always check AI outputs against current guidelines and specialist information.*

Practice Pearl

Three verification rules: (1) Never adopt dosages from AI — always check the product information. (2) Check source references individually — does the study actually exist? (3) For critical decisions, use a second tool or the primary source.

2.3 Prompt engineering: basic principles for medical queries

The quality of the AI's response depends directly on the quality of your query. In the context of AI, this query is called a 'prompt'. Good prompt engineering is not a programming skill — it is structured questioning. As a doctor, you already have this under your belt: you don't ask your patients open-ended questions when you need a specific answer.

The KLAR formula for medical prompts

Letter	Principle	Explanation	Example
K	Provide context	Who are you, and what is your working environment?	"I'm a GP in a German NHS practice"
L	Setting boundaries	What should the AI do, and what should it not do?	"No treatment recommendations, just a summary of information."
A	Define the task	What exactly should the AI deliver?	"Create a structured summary of the current DEGAM guideline"
R	Define the output format	In what format should the response be?	"As a table with 3 columns: recommendation, level of evidence, clinical relevance"

Example prompts

PROMPT EXAMPLE: Guideline research

I am a GP working in a German NHS practice. Summarise the current recommendations from the DEGAM guideline on the management of patients with atrial fibrillation. Focus on: (1) indications for oral anticoagulation, (2) the CHA_2DS_2-VASc score — used solely as a calculation tool, not as a decision-making tool, (3) practical steps for GPs. No dosage recommendations. Format: Table with columns: Recommendation | Level of evidence | Practical advice.

Verification: *Always check AI outputs against current guidelines and specialist information.*

PROMPT EXAMPLE: Structuring a doctor's letter

Structure the following bullet points as a doctor's letter to the cardiologist. Format: Standard structure with salutation, reason for writing, history-taking, findings, diagnoses, course of action, closing. Style: factual, precise, no repetitions. Please do not add any medical information that is not included in my bullet points.

Verification: *Always check AI outputs against current guidelines and specialist information.*

PROMPT EXAMPLE: Patient information in plain language

Create a patient information leaflet on high blood pressure at language level B1 (simple German). Target audience: Patients with low health literacy. Content: What is high blood pressure, why is it dangerous, what can patients do themselves, and when should they see a doctor? Do not use technical terms without explanation. Format: Short paragraphs, maximum 300 words.

Verification: *Always check AI outputs against current guidelines and specialist information.*

2.4 Quick Reference: Identifying typical AI errors

PROMPT EXAMPLE: Hallucination test

I am testing this AI system for reliability. Answer the following question — I will then check the answer manually:
What does the current DEGAM guideline [topic] say about first-line treatment?
Please state: recommendation, level of evidence, year of publication, DOI or URL.
I will check your answer directly on degam.de and awmf.org.

Verification: *Always check the guideline version and year of publication against degam.de or awmf.org — AI systems often output outdated versions without warning.*

To conclude this chapter, here is a quick reference guide to the five most common warning signs in AI output:

No.	Warning sign	Example	What to do?
1	Too smooth, too safe	AI provides a dosage without restrictions	Be wary — check the prescribing information
2	Source cannot be verified	AI cites ‘study by Müller et al. 2023’	Search for the DOI — does the study exist?
3	US-centric recommendation	AI recommends an “annual physical exam”	Explicitly ask for German guidelines
4	Contradiction with your own knowledge	AI lists a medicine as first-line that you know differently	Check the primary source — your clinical knowledge counts
5	Excessive detail	AI produces 2 pages when 3 sentences would suffice	Limit the length in the prompt: ‘max. 200 words’

Practice Pearl
Rule of thumb: the more significant the clinical implications, the more rigorous the verification. For patient information, a plausibility and fact check by the treating practitioner before handing it over is usually sufficient. For a dosing query: always consult the primary source.

2.5 ClinicalOS Decision Architecture and Cognitive Operators — thinking with AI without leaving the decision to it

AI systems are happy to answer almost any medical question — but not always the right one. If you ask an AI for 'probable diagnoses' without guidance, you'll get a list. If you guide it through the ten cognitive operations in this section, you'll get an assessment that a GP can use. The difference lies not in the model, but in the structure of the query.

> **NOTE The ClinicalOS Decision Architecture is not an AI function. It is the GP's decision architecture from Volume 1, to which AI is applied — not the other way round.**

2.5.1 Two levels: thinking architecture and tool layer

The Probability-Safety-Time (P-S-T) model from Volume 1 remains the guiding framework for every GP decision: how likely is a diagnosis, how dangerous would it be to overlook it, and how much time is available? This section does not repeat the derivation of P-S-T — it is set out in full in Volume 1 (Decision-Making Architecture). It shows how P-S-T can be translated into ten repeatable cognitive operations that can be formulated as AI tasks, without the AI itself becoming the decision-making authority.

The separation is strict: Decision Architecture determines which questions must be asked. The AI assists in answering these questions more quickly and comprehensively. Which answer is correct in each individual case remains the responsibility of the clinician.

2.5.2 The ten ClinicalOS Cognitive Operators

A Cognitive Operator is a repeatable cognitive operation used to analyse, prioritise, validate and translate a GP-related problem into action. The following ten operators cover the entire spectrum of a GP's decision-making process, from problem definition to follow-up.

Operator	Key question	AI's Role	Failure Mode
1. Define the problem	What is the actual clinical question today?	Suggests appropriate question types (ruling out risk, symptom management, observation, treatment adjustment, explanation)	Answers a question other than the clinically relevant one
2. Prioritise hypotheses	What is likely, what is dangerous, what is easy to overlook?	Organises explanations into three categories rather than a long list	Rare diagnoses are given greater weight through detailed description
3. Identify gaps in information	What missing information could change the decision?	Identify categories of missing information	Invent unmentioned patient data rather than flagging gaps

Operator	Key question	AI's Role	Failure Mode
		(history-taking, findings, clinical course)	
4. Assess the diagnostic impact	Which test would change my management approach?	Structured question, consequences of a positive/negative result, alternatives	Generates comprehensive laboratory and imaging checklists without reference to the decision
5. Distinguish between severity and urgency	How serious is the situation, and how urgently must action be taken?	Keeps the two axes separate rather than mixing them	Automatically derives an emergency measure from a rare, serious diagnosis
6. Assess reversibility	What happens if I take action today, or if I don't take action yet?	Lists questions regarding reversibility in relation to treatment, diagnosis and labelling	Overlooks the fact that a test can trigger a cascade that is difficult to reverse
7. Calibrating uncertainty	What do I know, what do I suspect, what remains unclear?	Distinguishes confirmed findings from plausible and unconfirmed assumptions	Linguistic certainty is misunderstood as epistemic certainty
8. Integrating the patient's perspective	Which goals, values and preferences influence the decision?	Suggests relevant questions, presents options clearly	Takes on the task of prioritisation, which the doctor and patient must decide on together
9. Define the course of action and safety netting	What are we doing now, and what would change the plan?	Sets out the measures, timeframe, warning signs and escalation criteria	Presents 'watchful waiting' without safety netting as a complete plan
10. Recursively reassess the course of events	How does the course of events alter the previous assessment?	Compare the initial findings with the course of events, identify revised hypotheses	Treats a one-off AI response as a permanently valid plan

The operators are not a linear algorithm, but a recursive cycle: any new information — a laboratory result, a follow-up appointment, a change in the patient's details — can lead back to Operator 2 (prioritising hypotheses).

2.5.3 The ClinicalOS Decision Map — ten steps at a glance

For day-to-day practice, the ten operators can be condensed into a compact Decision Map. It is intended as a memory aid, not as a form that must be worked through step by step.

Step	Content
A · Define	What is the problem and what decision needs to be made today?
B · Prioritise	Likely – dangerous – easy to overlook
C · Complete	What information relevant to the decision is missing?
D · Test impact	Which analysis will lead to a change in management?
E · Weigh	Weigh up probability, severity, urgency and reversibility
F · Calibrate	What uncertainty remains, and is it acceptable?
G · Share	What are the patient's goals and preferences?
H · Act	What action will be taken now, by whom, and by when?
I · Safety net	Expected course of treatment, warning signs, timing of follow-up
J · Reassess	How does the course of events alter the previous assessment?

2.5.4 What AI can do — and what remains the domain of doctors

Not every clinician is equally suited to AI support. A realistic assessment protects against two mistakes at once: failing to utilise AI where it provides genuine relief, and assigning it tasks that must remain the preserve of medical professionals.

Assessment	Examples
AI provides good support	Identifying information gaps, structuring the pros and cons of investigations, distinguishing between categories of uncertainty, formulating a safety-netting framework, structuring progress comparisons
AI provides partial support	Ranking hypotheses (structure: yes; reliable figures without valid case data: no), classifying risk and urgency (systematic approach: yes; binding reference value: no)
Must remain the responsibility of a doctor	Physical examination and assessment of findings, weighing up patient preferences, final decision on whether to act or wait and see, responsibility for diagnosis and treatment

2.5.5 From operator to prompt: the ClinicalOS Master Template

Every operator can be translated into a standardised prompt template. To ensure a template works across different AI systems, it always follows the same structure: objective, suitable and unsuitable use, the actual prompt, expected output and a safety gate for medical review. The complete library of twelve master templates is provided as a separate module; the following two examples illustrate the format.

PROMPT EXAMPLE Operator 2 — Likely, dangerous, overlooked

Classify possible explanations into three separate groups: likely / less likely but dangerous / frequently overlooked and relevant in the specific context. Briefly justify each classification. Do not use numerical probabilities unless valid case data is available. State any information that would alter the ranking.

Pitfall: *Without this explicit restriction, models tend to produce made-up percentages that give a false sense of certainty where no case data is available.*

PROMPT EXAMPLE Operator 4 — Diagnostic Impact

Assess each test under consideration using the following questions: What specific question does it answer? What would a positive result change? What would a negative result change? Can the same information be obtained with less risk or more easily? What side effects or incidental findings are possible? Is it appropriate now, later, or not appropriate at present? Output: appropriate now / depends on clinical course / not necessary at present.

Pitfall: *Without the final question, the answer often ends up as a mere list of possible tests, without prioritisation based on actual relevance to decision-making.*

2.5.6 Scientific context: what is new, and what is not

The individual building blocks of decision architecture are, for the most part, not new. The rule ‘Every test must change a decision’ stems from the threshold approach developed by Pauker and Kassirer (1980), who derive test and treatment thresholds from the benefits and risks of a test. The operators ‘calibrating uncertainty’ and ‘testing counter-hypotheses’ build on Croskerry’s work on cognitive dispositions and debiasing strategies. Safety netting, shared decision-making and watchful waiting are also established concepts developed independently of AI.

It is not the individual building blocks that are original, but rather their systematic combination into a unified structure applicable across all specialist fields, which can simultaneously be formulated as an AI mandate. This combination has not yet been formally validated. The honest assessment of this is:

> **NOTE The ClinicalOS Decision Architecture is an evidence-informed, practice-based working model. It combines established principles of general practice and decision science within a structure that can be operationalised for digital and AI-supported work processes. Formal empirical validation is still pending.**

2.5.7 Regulatory framework

Prompt templates based on this model remain structural aids; they are not autonomous diagnostic or treatment recommendations — this distinction is relevant not only from an educational perspective but also, increasingly, from a regulatory one.

WARNING Medical device and AI regulation classification

Classification: AI-supported clinical decision support with a diagnostic or treatment recommendation function will typically meet the UK definition of Software as a Medical Device (SaMD) / AI as a Medical Device (AIaMD) — but, unlike in the EU, there is no binding AI-Act-style 'high-risk' conformity layer on top of the medical device rules.

Timeline: MHRA's guidance on AI/ML as a medical device is advisory rather than mandatory. The UK's draft Medical Devices (Amendment) Regulations 2026 introduce stricter AI/software oversight via Predetermined Change Control Plans (PCCPs) and enhanced post-market surveillance, but there is currently no UK equivalent to the EU AI Act's binding high-risk conformity deadline.

Classification for this chapter: A printed chapter containing prompt templates is not in itself a medical device or AI system. However, once operators and templates are transferred to a tool integrated into a patient record system (PMS), the medical device classification is likely to apply.

Reference: See Chapter 14.2 (medical device regulation) for details.

2.5.8 From everyday practice

FROM EVERYDAY PRACTICE Chest pain on a Friday afternoon

Situation: 58-year-old patient, experiencing atypical left-sided chest pain for the past two days; the surgery is fully occupied shortly before closing time.

AI application 1: Operator 2 (prioritising hypotheses): A structured query categorises possible explanations as ‘likely’, ‘dangerous’ and ‘easily overlooked’, rather than providing an unsorted differential diagnosis list.

AI application 2: Operator 4 (Diagnostic Impact): The AI outlines how an immediate troponin test and an ECG would alter the course of action depending on whether the results are positive or negative.

AI application 3: Operator 9 (safety netting): Whilst the doctor examines the patient and makes a decision, the AI prepares a draft of the patient’s warning signs information, which the healthcare assistant (HCA) then simply checks and hands over.

Time saved: ~4 minutes instead of ~12 minutes for structuring and drafting the safety-netting text — the time saved is spent on history-taking and examination, not on documentation.

Caution: The AI does not replace either the ECG interpretation or the decision regarding hospital admission. Both remain the responsibility of the doctor, even if the AI’s response is convincingly worded.

2.5.9 Conclusion

ClinicalOS does not dictate what answer an AI should provide. It defines which clinical questions should be asked, which uncertainties should be highlighted, and which decisions must continue to be the responsibility of the clinician. The ten Cognitive Operators and the Decision Map are the tools for this — no more, but also no less.

RESOURCES

Chapter 2.5 – Scientific Foundations and Regulatory Framework

Pauker & Kassirer (1980): The Threshold Approach to Clinical Decision Making – pioneering work on the test and treatment threshold model, NEJM (primary source)

Croskerry: Cognitive Debiasing, Parts 1 and 2 – Foundational work on cognitive biases and debiasing, AHRQ PSNet (free)

EU AI Act: High-Level Summary – Concise overview of high-risk classification and deadlines (free)

References for Part I

1. Bitkom press release 2025. AI in use in almost one in seven GP practices and many hospitals.
2. KBV PraxisWissen 2025. Guidance on the use of AI in medical practices. Available at: kbv.de
3. German Medical Association 2025. Position paper on AI in healthcare.
4. Huang J et al. AI and Primary Care: Scoping Review. J Med Internet Res. 2025;27:e65950.
5. Liyanage H et al. AI in primary care — a general practitioner's bucket list. EGPRN Keynote 2025. Eur J Gen Pract. 2025.
6. Smith C et al. Artificial intelligence in primary care: frameworks, challenges, and guardrails. Lancet Primary Care. 2025.
7. Chen A et al. A randomised clinical trial of two ambient AI scribes: measuring documentation efficiency and physician burnout. JAMA Netw Open. 2025.
8. Topol EJ. Deep Medicine: How Artificial Intelligence Can Make Healthcare Human Again. New York: Basic Books; 2019.
9. KBV IT Security Policy. Effective from 1 April 2025. Available at: kbv.de
10. Central Institute for Statutory Health Insurance Medical Care (Zi). Practice Software Survey 2025.
11. Thornton T. Tacit knowledge as the unifying factor in evidence-based medicine and clinical judgement. Philos Ethics Humanit Med. 2006;1:2.
12. Videtta G, Busilacchi S, Bartoccioni G, Cirella L, Barone Y, Delvecchio G. Effects of therapeutic alliance on patients with major depressive disorder: a literature review. Front Psychol. 2024;15:1465017.
13. Huang Y, Wei X, Wu T, Chen R, Guo A. Collaborative care for patients with depression and diabetes mellitus: a systematic review and meta-analysis. BMC Psychiatry. 2013;13:260.

14. Aramrat C, et al. Advancing the management of multimorbidity in primary care: a narrative review. Prim Health Care Res Dev. 2022;23:e36.

15. Haynes AB, Weiser TG, Berry WR, et al. A surgical safety checklist to reduce morbidity and mortality in a global population. N Engl J Med. 2009;360(5):491–499.

16. National Institute for Health and Care Excellence. Multimorbidity: clinical assessment and management. NICE guideline NG56. London: NICE; 2016.

17. Shanafelt TD, Dyrbye LN, Sinsky C, et al. Relationship between administrative burden and characteristics of the electronic environment with physician burnout and professional satisfaction. Mayo Clin Proc. 2016;91(7):836–848.

18. Arndt BG, Beasley JW, Watkinson MD, et al. Tethered to the EHR: primary care physicians' time spent on the electronic health record. Ann Fam Med. 2017;15(5):419–426.

19. Wehkamp K, Krawczak M, Schreiber S. Benefits and risks of artificial intelligence in medicine. Dtsch Arztebl Int. 2023;120:463–469.

20. Lakhani P, Sundaram B. Deep learning in chest radiography: automated classification of pulmonary tuberculosis using convolutional neural networks. Radiology. 2017;284(2):574–582.

21. 'Bis der Arzt kommt' — recruitment podcast by Ärztestellen, Deutscher Ärzteverlag. Episode 56: Telemedicine, AI and data security. aerztestellen.podigee.io (clinical context, no scientific evidence).

PART II – AI tools for clinical practice

Chapters 3–5: Overview of tools • Research • Further training

Chapter 3: Overview — Which tool for which purpose?

The number of AI tools is growing by the week. For GPs, the crucial question is not 'Which tool is the best?', but rather: 'Which tool will solve my specific problem within the next 5 minutes?' This chapter provides the overview you need to answer this question systematically.

3.1 A Comparison of LLM Platforms

A direct comparison of the four most relevant LLM platforms for everyday practice:

Criterion	ChatGPT (OpenAI)	Claude (Anthropic)	Gemini (Google)	Perplexity
Core strength	Versatile, large user base, plugins	Long texts, precise adherence to instructions, analysis	Google integration, multimodal, NotebookLM	Real-time web search with source citations
Suitability for medical use	Good for documentation and text generation	Strong at nuanced analysis and long documents	Good for image recognition and multimedia	Best for research using up-to-date sources
German	Very good	Very good	Good	Good
Data protection	Opt-out available, API GDPR-compliant	No use of training data with API/Pro	Google account required, privacy concerns	Anonymous use possible
Costs (approx.)	Free / Plus €20/month / Pro €200/month	Free / Pro €20/month / Max €100/month	Free / Advanced €22/month	Free / Pro €20/month
Live web access	Yes (with browsing)	No (tool use only)	Yes	Yes (core feature)

Criterion	ChatGPT (OpenAI)	Claude (Anthropic)	Gemini (Google)	Perplexity
File upload	Yes (PDFs, images, code)	Yes (PDFs, images, long texts)	Yes (including video)	Limited
Ideal for	Doctor's letters, SOPs, text generation	Study analysis, guideline summaries, complex questions	Image-based questions, YouTube summaries	Clinical research with references

Practice Pearl
To get started: Perplexity for research, Claude or ChatGPT for documentation. You don't need all four — start with two.

3.2 Decision tree: Which tool for my question?

The following decision tree helps you choose a tool quickly:

Your situation	Recommended tool	Why
I need an up-to-date guideline recommendation	Perplexity or OpenEvidence	Live web search with source references
I want to have a study summarised	Claude or Elicit	Powerful text analysis, precise summarisation
I need to dictate a doctor's letter	Dragon Medical One	Speech recognition >99%, GDPR-compliant in Germany
I want to structure a doctor's letter from bullet points	ChatGPT or Claude	Text generation and structuring
I need patient information in plain language	ChatGPT or Claude	Adapting language to the target audience
I need to translate a text	DeepL	Highest translation quality, suitable for medical use
I want to compare several papers	Elicit or Consensus	Structured data extraction from studies
I want to process a CME article	NotebookLM or Claude	PDF upload, summary, quiz questions
I'm looking for a clinical score	MDCalc or AMBOSS	Validated calculators, ready to use
I want to create a standard operating procedure (SOP) for the practice	ChatGPT or Claude	Structured template creation

3.3 Costs, data protection and availability in the DACH region

Before you start using a tool in practice, check three criteria:

Tool	Cost per month	GDPR status	Server location	German	Available in Germany, Austria and Switzerland
ChatGPT Plus	~€20	Opt-out available; API: DPA available	USA (API: EU available)	Yes	Yes
Claude Pro	~€20	No training data for Pro/API	USA/EU	Yes	Yes
Gemini Advanced	~€22	Google Privacy Policy	Global (Google Cloud)	Yes	Yes
Perplexity Pro	~€20	Anonymous use possible	USA	Yes	Yes
OpenEvidence	Free	US-based, NPI required	USA	English	Limited
DeepL Pro	from €8	GDPR-compliant, EU servers	Germany/EU	Primary language	Yes
Dragon Medical One	€129–189	ISO 27001, Azure DE	Germany (Berlin/Frankfurt)	Yes	Yes
AMBOSS	from €15 (doctor)	GDPR-compliant, Berlin	Germany	Yes	Yes
NotebookLM	Free	Google Privacy Policy	Google Cloud	Yes	Yes
Elicit	Free/from $10	US-based	USA	English	Yes
Consensus	Free/from $10	US-based	USA	English	Yes

A practical rule of thumb for data protection

- **NEVER enter patient data into an LLM that is not explicitly GDPR-compliant and authorised for medical data.**
- ***For research (excluding patient data): All tools may be used.***
- ***For patient-related documentation: Only use GDPR-compliant solutions with a DPA (Dragon Medical One, certified Ambient Scribes).***
- ***For text templates (without patient-related information): All tools may be used — always anonymise the data.***
- **If in doubt: anonymise all patient data before entering it into an AI tool.**

3.4 Practice Pearl: The question first, then the tool

PROMPT EXAMPLE Choose a tool for my question

I have the following clinical question: [enter your question]
Which AI tool is best suited to this? Justify your choice.
How do I phrase the prompt effectively?
How do I verify the result?
Criteria: GDPR-compliant, available in German, suitable for a GP practice.

Verification: Tool recommendations change regularly. Note the date of publication of the recommendation and review it again after 6 months.

Practice Pearl

The most common mistake: opening a tool and then searching for a question. The correct approach: first formulate the clinical question, then select the appropriate tool. The KLAR formula from Chapter 2.3 always applies.

PART III – Documentation and Communication

Chapters 6–8: Documentation • Patient Communication • Translation

Chapter 6: AI-assisted documentation

Documentation is the biggest time-sink in day-to-day practice. Studies show that doctors spend up to 50 per cent of their working hours on documentation. AI-assisted documentation promises the greatest time savings of all the tools described in this manual — provided one is aware of its strengths and limitations.

6.1 Ambient AI Scribes: The current state of the art

Ambient AI Scribes are AI systems that record the doctor–patient consultation in the background and automatically generate structured documentation. The first randomised controlled trial (UCLA RCT, November 2024 – January 2025, n=238 doctors, 14 specialities) provides robust evidence for the first time:

Criterion	DAX Copilot (Microsoft/Nuance)	Nabla	Abridge	Suki
Evidence (RCT)	UCLA RCT: 23-second reduction (n.s.)	UCLA RCT: 41-second reduction (significant)	Best in KLAS 2025 and 2026	72% reduction in note-taking time (manufacturer's figure)
Adoption rate (RCT)	~30% of contacts	~30% of contacts	N/A	N/A
Accuracy (RCT)	2.7/5 Likert (frequent errors)	2.8/5 Likert (frequent errors)	N/A	N/A
Languages	English (German in development)	40+ languages	English	English
EHR integration	Epic, Cerner	Open API	Epic	Epic, Cerner, athena
Price (approx.)	$$$+	$$	$$$	$$

Criterion	DAX Copilot (Microsoft/Nuance)	Nabla	Abridge	Suki
Available in Germany, Austria and Switzerland	Limited	Yes (API)	No	No
Burnout effect	~7% improvement	~7% improvement	N/A	N/A

Evidence assessment: Ambient AI Scribes

Key finding of the UCLA RCT:

Not all AI Scribes are the same — the choice of product makes a measurable difference.

The adoption rate was only around 30%, even though the tools were provided free of charge.

Clinically relevant errors occur regularly — all AI-generated documentation must be reviewed by a doctor.

Conclusion: Promising, but not yet a 'set-and-forget' tool. Medical review remains mandatory.

6.2 Dragon Medical One: Speech recognition for German medical practices

Dragon Medical One (DMO) from Nuance/Microsoft is the most established speech recognition solution for German-speaking countries and advertises GDPR compliance for practice documentation; before implementation, check whether this also applies to current alternatives:

Feature	Details
Recognition accuracy	According to the manufacturer > 99%, dialect-independent, no training required
Language	German (medical terminology)
Server location	Microsoft Azure Germany (Berlin, Frankfurt am Main)
Certification	ISO 27001, GDPR-compliant
Price	approx. €129–189/month (12–24-month licences)
Integration	Can be integrated into common PMS systems
Path to DAX Copilot	DMO is the entry-level option; DAX Copilot is planned as Ambient Scribe
Trial	14–30 days' free trial with various providers

Workflow: Dictate a doctor's letter in 3 minutes
Step 1: Complete the consultation; note down key points mentally or by hand
Step 2: Open Dragon Medical One, launch the medical report template in the practice management system
Step 3: Dictate free-form text: history-taking, findings, diagnoses, treatment plan (real-time speech recognition)
Step 4: Check the dictation: correct any typos, adjust the structure (30–60 seconds)
Step 5: Send the letter or add it to the patient's file

6.3 SOAP notes, medical reports and progress notes with AI

Even without Ambient Scribe or speech recognition, you can use AI tools for documentation — by organising bullet points into finished documents:

SOAP notes

PROMPT EXAMPLE: SOAP note from bullet points

Create a SOAP note for the patient's file using the following bullet points:
[Insert your bullet points here]
Format: S (Subjective) | O (Objective) | A (Assessment) | P (Plan).
Style: Factual, precise, medical terminology, no repetition.
IMPORTANT: DO NOT add any medical information that is not included in my bullet points. Do not speculate on diagnoses or suggest treatment.

Verification: Always check AI outputs against current guidelines and specialist information.

Letter to a specialist

PROMPT EXAMPLE: Generating a doctor's letter

Organise the following bullet points into a doctor's letter to the [specialist]. Format:
- Subject: Brief description
- Salutation
- Reason for referral (1–2 sentences)
- Relevant history-taking (bullet points)
- Current findings
- Previous medication (list)
- Query for the consultant
- Salutation
Style: Factual, collegial, no filler words. Do not add any further information.

[Key points here]

Verification: Always check AI outputs against current guidelines and specialist information.

Case history documentation

PROMPT EXAMPLE: Medical record entry

Create a structured progress note for the patient's medical record using these bullet points:
[Key points]
Format: Date | Reason | Findings | Action | Next step.
Maximum of 5 lines. Medical abbreviations permitted.

Verification: Always check AI outputs against current guidelines and specialist information.

6.4 Workflow: Documentation with AI support

CHECKLIST 10-point Note-QA checklist — Checking AI documentation

Check every AI-generated note before the doctor's sign-off:

1. Have allergies and relevant intolerances been recorded correctly?
2. Is the list of medications complete and accurate?
3. Have red flags or warning signs been mentioned where relevant?
4. Are confirmed diagnoses clearly distinguished from suspected diagnoses?
5. Has patient education or counselling been correctly documented?
6. Is the plan clear — including follow-up and re-evaluation?
7. Are codes or structured fields plausible?
8. Are there any risks of copy-and-paste errors or text fragments from other cases?
9. Does the note contain linguistic misunderstandings or unclear phrasing?
10. Has a deliberate clinical sign-off been carried out?

Verification: AI-generated documentation without a final medical review is not legally sufficient. A sign-off is mandatory, not optional.

Workflow: Documentation in 4 steps

Step 1: During the consultation: Note down key points (by hand, in the patient record system or mentally)

Step 2: After the consultation: Enter key points into an AI tool (ChatGPT, Claude) using the appropriate prompt (SOAP, doctor's letter, medical history)

Step 3: Medical review: Read the AI output critically — are all the facts correct? Has anything been made up? Is the structure appropriate?

Step 4: Transfer to the medical record: Copy the reviewed text and paste it into the practice management system or send it

Practice Pearl

Documentation using AI is a two-step process: (1) AI structures and formulates. (2) The doctor reviews and takes responsibility. Step 2 must never be omitted. The AI output is a draft, not a diagnosis.

6.5 Case vignette: Home visit to a patient with multiple comorbidities

> *Good documentation AI does not replace clinical judgement — it only becomes reliable through review, awareness of potential errors and consistent routines.*

Case vignette
Situation:
Mrs K., aged 82, type 2 diabetes, heart failure, early-stage dementia. Home visit due to deterioration in general condition. Daughter reports: reduced appetite, confused for the past 3 days, swollen legs.

AI-assisted documentation:

During the visit, you note the following key points:

- General condition reduced, confused for 3 days, eating little, oedema in both lower legs
- BP 100/60, pulse 88 irregular, SpO_2 94%, temperature 37.2
- Weight +3 kg in 1 week, basal moist rales in both lungs
- Primarily cardiac decompensation due to dehydration and reduced fluid intake
- Furosemide 40 mg increased to 60 mg; daily weight monitoring by daughter
- Lab tests tomorrow (creatinine, eGFR, Na, K, BNP), weight check in 2 days

Entered into Claude using the SOAP prompt, this generates a structured progress note in 30 seconds, which you can review and transfer to the medical record.

Time saved: approx. 5–8 minutes per home visit record.

The next step in development is not only improved speech recognition, but the integration of the AI note into PMS, the patient record and the team workflow.

6.6 PMS-integrated AI documentation

Ambient Scribes and AI dictations only provide lasting relief if the generated note is seamlessly transferred into the PMS and patient record — not as an additional copy alongside the actual documentation. This requires documented consent for AI-assisted recording, as well as a mandatory review process prior to approval.

Types of errors in AI notes	Clinical relevance
Omission (relevant information missing)	Red flags or medication may be overlooked
Invention (hallucination, see 13.1)	Findings that were never documented appear in the text
Misattribution (speaker/time)	Confusion between patient, relatives and team

Warning

Allergies, diagnoses, medication, red flags, patient education and follow-up must never be copied from an AI-generated note without verification.

Minimum SOP and 4-week audit

- Obtain and document consent for AI-assisted documentation before commencing.
- Each AI note must be checked by the treating clinician before being added to the patient's record.
- After 4 weeks: check a random sample of 10 notes for omissions and hallucinations.
- Discuss the error rate and types of errors within the team; adjust the SOP if necessary.

Cross-references: 9.7 Data logic (poor patient-reported data compromises AI documentation) · 14.6 AI competence (team and approval) · 13.1 Hallucinations.

ClinicalOS

LESEPROBE / PREVIEW

Explore the complete ClinicalOS edition

Volume 5 · Digital Tools & AI

For information about the complete edition and the ClinicalOS series, visit:

clinicalos.de

The original notices on use and responsibility also apply to this sample.