

The Clinical Question

How to ask AI so it reasons well

The models are now strong reasoners. The evidence shows that capability does not reach the patient automatically. It passes through a question, and the question sets the ceiling. This guide is about asking well.

FOR	Doctors, nurses and allied health professionals using clinical AI
INSIDE	The evidence · How your question travels · The SCOPE method · 12 worked rewrites · Question traps · Quick-reference card
TIME TO READ	20 minutes. Time to apply: seconds per question
EDITION	1.0 · September 2026

The machine can reason. The gap is in the asking.

Across the best recent trials, AI models working alone have outperformed the people using them. The capability was present. It did not survive the interaction.

In a 2024 randomised trial, 50 physicians worked through complex diagnostic cases. Those given GPT-4 alongside their usual resources scored a median of 76% against 74% for those without it, a difference that was not statistically significant. The model on its own scored 16 percentage points higher than the physicians using conventional resources.

A 2026 preregistered trial in *Nature Medicine* found the same gap at scale. Alone, the models identified the relevant condition in 94.9% of scenarios. The 1,298 people using those same models managed it in fewer than 34.5% of cases, no better than a control group. When researchers read the conversations, they found people often gave the model incomplete or incorrect information.

STUDY	FINDING	WHAT IT MEANS FOR YOU
Goh et al. JAMA Netw Open 2024 · 50 physicians	Model alone +16pp; physicians with the model 76% vs 74%	Access to a strong model is not the same as benefiting from it
Bean et al. Nature Medicine 2026 · 1,298 people	Model alone 94.9%; people using it under 34.5%	Incomplete information in produces poor reasoning out
Johri et al. (CRAFT-MD) Nature Medicine 2025 · 2,000 cases	Significant accuracy drops when exam questions became real conversations	Benchmarks overstate performance on messy, real exchanges
Chen et al. npj Digit Med 2025 · 5 models	Compliance with illogical drug requests reached 100%	Models tend to accept the premise of your question
Goh et al. Nature Medicine 2025 · 92 clinicians	Model assistance improved management reasoning by 6.5pp	Deliberate engagement closes the gap. Better performers spent longer

THE PRINCIPLE

A model can only reason toward the goal your question sets. **The quality of the answer is decided before the model begins.**

This is not a failing peculiar to AI. Clinicians have always known that the history is most of the diagnosis. The same holds here.

What happens between pressing enter and reading the answer

Clinical AI works through stages, and your wording steers every one of them.



STAGE	WHAT THE SYSTEM DOES	WHAT YOUR WORDING CONTROLS
Question	Interprets your intent from your words alone. It cannot see what you meant but did not say.	The goal, the scope and the standard of evidence
Retrieval	Decides which parts of the record and which sources to pull.	A narrow question retrieves narrowly. An unbounded one lets the system choose what matters
Reasoning	Connects what it retrieved. Compares, trends, checks for conflicts.	A buried assumption becomes the starting point of the reasoning
Draft	Writes an answer shaped to what it believes you asked.	Length, structure, and whether uncertainty is shown or smoothed over

Three mechanisms behind a poor answer

- 1

It answers what you asked, not what you meant

Every decision you leave unstated is made for you, silently. Most disappointing answers are faithful answers to vague questions.
- 2

It tends to agree with the premise

In one study, models wrote misinformation built on a false drug premise in up to 100% of cases, despite knowing the facts. Explicit permission to refuse sharply reduced this.
- 3

Small changes in wording change the answer

Prompt style alters accuracy, and the same question does not always produce the same answer. Precision reduces the variance.

WHY AGENTIC SYSTEMS RAISE THE STAKES

An agentic assistant retrieves, reasons, drafts, and should decline or escalate when evidence runs out. A flawed question can therefore produce an answer that is fully sourced and still addresses the wrong thing. **Sources protect the answer. Only you can protect the question.**

You are not a prompt engineer. You are the one who knows the question.

You should not need to write elaborate instructions. A well-built clinical system carries the context. What it cannot carry is your intent.

THE SYSTEM SHOULD CARRY	ONLY YOU CAN SUPPLY
Which patient, and their full record	The clinical question you actually need answered
Retrieval of the relevant documents and results	The period that matters
Allergies, medications and critical results, surfaced by default	The decision this answer will inform
A source behind every clinical statement	What you already suspect, and want tested
Declining when the record does not support an answer	Permission to be told you are wrong

THE RULE OF THUMB

Brief the AI the way you would brief a capable registrar who has never met the patient. **The difference is that the registrar would ask what you meant. The model will often guess.**

A note on safety-critical facts

No clinician should have to remember to ask whether a patient will be harmed. A trustworthy system surfaces allergies, interactions and critical results whatever the question. If a tool only mentions an allergy when you ask about allergies, it is relying on your memory, and that is a property worth testing before you rely on it.

SCOPE: the anatomy of a precise clinical question

A precise clinical question has five parts. Most weak questions are missing three of them. You rarely need all five in writing, but you should have considered each.

	ELEMENT	ASK YOURSELF	EXAMPLE
S	Subject	Which patient, and which part of the record?	Her renal results and current medication list
C	Clinical question	What single thing do I need to know?	Has her kidney function changed?
O	Over what period	Since when does it matter?	Over the last six months
P	Purpose	Which decision does this inform?	Before I review her metformin dose
E	Evidence standard	How should uncertainty be handled?	Cite each result and tell me what is missing

BEFORE

Tell me about her kidneys.

AFTER · SCOPE APPLIED

Over the last six months, how has her renal function changed, and which of her current medications are relevant to it? I am reviewing her metformin dose. Cite each result, and tell me if any expected results are missing from the record.

Why each element earns its place

ELEMENT	WHAT GOES WRONG WITHOUT IT
Subject	The system draws from the wrong part of the record, or everything at once
Clinical question	You receive a summary when you needed an answer
Over what period	Old data crowds out recent change, or recent data hides a long trend
Purpose	The answer is accurate but irrelevant to the decision in front of you
Evidence standard	Gaps are smoothed into confident prose instead of being flagged

Twelve everyday questions, asked well

Each rewrite applies SCOPE to a common clinical task. Adapt the pattern rather than copying the words. All patients are illustrative.

TASK	WEAK	PRECISE
Pre-consult preparation	Anything I should know?	Since his last visit in March, what has changed that bears on his heart failure management? Include results, medication changes and admissions, with sources.
Record summary	Summarise this patient.	Summarise her active problems and current medications for a colleague covering my clinic. Flag anything unresolved from the last three visits.
Allergies	Confirm no allergies.	List every documented allergy and adverse reaction with dates and sources. State explicitly if no allergy record exists.
Medication review	Are her meds OK?	Review her current medications for interactions and for doses needing adjustment given her latest eGFR. Show the evidence for each concern.
Result trends	Why did his HbA1c get worse?	How has his HbA1c trended over the past year, and what changed in his medication or documented adherence over the same period?
Referral	Write a referral.	Draft a cardiology referral focused on his recent chest pain episodes. Include relevant ECG and troponin results, and list what the cardiologist will need that is not in the record.
Discharge letter	Do the discharge summary.	Draft a discharge summary for her GP covering the admission diagnosis, key results, medication changes with reasons, and the follow-up plan. Mark anything I need to confirm.
Handover	Give me a handover.	Prepare a handover for a colleague who has never met her: active problems, current treatment, pending results and what to watch for overnight.
Treatment question	Is this the right treatment?	What in the record supports continuing this treatment, and what argues against it? What information is missing that would change the picture?
Guideline check	What does the guideline say?	Which current guideline recommendations apply to a patient with these characteristics, and where does her situation fall outside them? Cite the guideline sections.
Outstanding items	Anything outstanding?	List all investigations, screening and follow-ups that were planned or deferred in the past year and not completed.
Patient letter	Explain this to the patient.	Draft a plain-language letter explaining her new diagnosis of pre-diabetes, what it means, and the agreed next steps, at a reading level suitable for a general audience.

Five question traps

Each trap feels natural to type. Each quietly lowers the quality of the answer.

- 1 Presumption** "WHY DID X GET WORSE?"
The model reasons from an assumption it never checks. **Fix:** ask what the record shows before asking why.
- 2 Leading** "CONFIRM THERE IS NO..."
Invites agreement over accuracy, and blurs the difference between "none" and "none recorded". **Fix:** ask the open question and request sources.
- 3 Unbounded scope** "TELL ME EVERYTHING"
The system silently decides what matters. **Fix:** name the period and the purpose.
- 4 Compound** THREE QUESTIONS IN ONE
Later parts are answered thinly or dropped altogether. **Fix:** one clinical question per request, then follow up.
- 5 Demand for certainty** "WHAT IS THE DIAGNOSIS?"
Pushes the model past what the record supports, and invites a confident guess. **Fix:** ask for evidence for and against, and what is missing.

THE SUBTLE ONE: "NONE" VERSUS "NONE RECORDED"

"Does she have any allergies?" and "Is there an allergy record?" are different questions. A patient with no documented allergies may simply never have been asked. A precise question keeps that distinction visible, and a good answer states it.

Phrases that make uncertainty visible

The Evidence element of SCOPE is the one most often left out, and the one that most protects patients. These phrases can be added to almost any question.

ADD THIS	WHAT IT DOES
Cite the source for each clinical statement.	Makes every claim checkable against the record
Tell me if the record does not contain this.	Stops gaps being filled with plausible guesswork
List what information is missing.	Turns omission, the invisible error, into something you can see
If my assumption is wrong, say so.	Gives explicit permission to disagree, which counters agreement bias
What would change this conclusion?	Exposes how fragile the answer is
Separate what is documented from what is inferred.	Distinguishes record from reasoning

Follow-up questions are where the value is

The first answer is a draft. The most useful clinicians in the trials engaged longer, not shorter. Three follow-ups cover most needs.

- **Verify**
"Show me the source for the statement about her potassium."
- **Narrow**
"Focus only on changes since her admission in June."
- **Challenge**
"What is the strongest argument against that conclusion?"

A CAUTION ON PUSHING BACK

Challenge the answer to test it, not to win it. Research on model sycophancy shows that when a user objects, a model may retract a correct answer and endorse a false one. If an answer changes the moment you disagree, ask for the evidence behind both versions.

Asking well is half the job. Reviewing well is the other half.

You remain the author of the record. A structured pass takes less time than writing from scratch and catches the errors that matter.

One of the largest manual evaluations of AI clinical note generation, covering almost 13,000 annotated clinical note sentences, found a hallucination rate of 1.47% and an omission rate of 3.45%. Missing content was more than twice as common as invented content, and it is harder to spot because nothing on the page signals a gap.

CHECK	ASK OF EVERY CLINICALLY MATERIAL STATEMENT
Sourced	Can I trace this to the record or the encounter?
Faithful	Are the negations, doses, values, dates and laterality exactly right?
Complete	Is anything clinically relevant missing, especially allergies and current medications?
Current	Is this still true now, or has it been superseded?

What a well-built system should do with an imperfect question

WHEN THE QUESTION IS...	A TRUSTWORTHY SYSTEM SHOULD...
Ambiguous	Ask what you mean, rather than guess
Built on a false premise	Say so, rather than reason from it
Beyond what the record supports	Decline, and state what is missing
About a patient with safety-critical facts	Surface them, whether or not you asked

REFUSAL IS A FEATURE

In a clinical tool, declining to answer is not a limitation. **It is the highest expression of safety.**

A system that always has an answer is a system that sometimes invents one.

Quick-reference card

Print this page and keep it by your workstation.

The Clinical Question · SCOPE

Respocare Connect AI · How to ask clinical AI so it reasons well

S	Subject	Which patient, which part of the record
C	Clinical question	One specific thing you need to know
O	Over what period	The time window that matters
P	Purpose	The decision this answer informs
E	Evidence standard	Sources for every claim; say what is missing

AVOID

Presumption · "Why did X worsen?"
 Leading · "Confirm there is no..."
 Unbounded · "Tell me everything"
 Compound · three questions in one
 Certainty · "What is the diagnosis?"

ADD

Cite the source for each statement
 Tell me if the record lacks this
 List what is missing
 If my assumption is wrong, say so
 What would change this conclusion?

BEFORE YOU SIGN

Sourced · traceable to the record **Faithful** · doses, negations, sides, dates **Complete** · allergies and medications present **Current** · still true now

The AI drafts. The record verifies. The clinician decides.

REFERENCES

The evidence behind this guide

1. Goh E, Gallo R, Hom J, et al. Large language model influence on diagnostic reasoning: a randomized clinical trial. *JAMA Network Open*. 2024;7(10):e2440969.
2. Bean AM, Payne RE, Parsons G, et al. Reliability of LLMs as medical assistants for the general public: a randomized preregistered study. *Nature Medicine*. 2026.
3. Johri S, Jeong J, Tran BA, et al. An evaluation framework for clinical use of large language models in patient interaction tasks. *Nature Medicine*. 2025.
4. Chen S, Gao M, Sasse K, et al. When helpfulness backfires: LLMs and the risk of false medical information due to sycophantic behavior. *npj Digital Medicine*. 2025;8:605.
5. Goh E, Gallo RJ, Strong E, et al. GPT-4 assistance for improvement of physician performance on patient care tasks: a randomized controlled trial. *Nature Medicine*. 2025.
6. Asgari E, Montaña-Brown N, Dubois M, et al. A framework to assess clinical safety and hallucination rates of LLMs for medical text summarisation. *npj Digital Medicine*. 2025;8:274.
7. Wang L, et al. Prompt engineering in consistency and reliability with the evidence-based guideline for LLMs. *npj Digital Medicine*. 2024.

ABOUT

Respocare Connect AI

Respocare Connect AI is an agentic clinical assistant built by Respocare (PTY) Ltd, a licensed South African healthcare practice established in 2018. It is designed to carry patient context, retrieval and sourcing so that clinicians can focus on the question, and it is built to clarify, challenge or decline rather than guess. Clinical validation is ongoing.

Respocare Connect AI is a clinical assistant. It is not a medical device, is not diagnostic, and is not a substitute for clinical judgement. All output requires independent clinician review. This guide is educational and applies to clinical AI generally.

Respocare (PTY) Ltd · Reg. 2018/411829/07 · Practice No. 9990900010775614 · respocareconnectai.com · support@respocareconnectai.com