

Red Flags in AI-Generated Patient Content

Medical writer
Plain language consultant
GenAI trainer for patient advocacy groups
PIF TICK competent freelancer

AI-generated content often reads fluently and confidently — which is precisely when you should be most alert. These seven red flags are the most common failure modes in AI-drafted patient content. Pin this to your desk, share it with your team.

01 Confident claims with no source

AI presents statements as fact without citing where they came from. The tone is assertive — the evidence is absent.

EXAMPLE: *A paragraph stating "studies show" or "research suggests" with no study, journal, or year attached.*

02 Fabricated citations that sound real

AI generates plausible-looking references — correct journal, credible author name, feasible year — for studies that do not exist.

EXAMPLE: *"Williams et al., Journal of Clinical Oncology, 2021" for a paper that no-one has ever written.*

03 Complex clinical situations made too simple

Nuance, caveats, and contraindications get smoothed away in favour of clean, readable prose. What's lost is often what matters most to patients.

EXAMPLE: *A summary of treatment options that reads well but omits who should not have the treatment and why.*

04 Outdated guidelines presented as current

AI training data has a cut-off date. Recent changes in clinical practice, drug approvals, or national guidelines may be missing — but the output will not tell you that.

EXAMPLE: *Screening age recommendations from three years ago presented without qualification.*

05 Missing caveats and collapsed uncertainty

Genuine clinical uncertainty — "this is usually but not always", "evidence is mixed", "individual response varies" — disappears from the output in favour of confident summary.

EXAMPLE: *"This medication causes [side effect X]" instead of "Some people taking this medication experience [X]".*

06 Tone that misses the patient

Output swings between clinical jargon and unintended condescension. Neither works for patient audiences, and the default can shift unpredictably between prompts.

EXAMPLE: *"Your oncologist will orchestrate your treatment journey" — a sentence no real person would say to a newly diagnosed patient.*

07 Polished, fluent output — especially be alert then

The most dangerous AI output is the output that reads most smoothly. Fluent prose creates a false sense of authority. Every polished paragraph needs the same scrutiny as a rough one.

EXAMPLE: *A paragraph that flows beautifully, feels expert, and turns out to have two factual errors in it.*

THE UNDERLYING RULE

Every AI output is a first draft. Use it as a starting point — never as a finished product. Verification is not optional; it is the work.