

The rise of **generative AI chatbots** and **large language models (LLMs)** as go-to sources for health questions signals a fundamental shift in how people seek medical information. Unlike traditional search engines that direct users to trusted health resources or peer-reviewed research, AI chatbots generate conversational answers instantly, often with a tone of confidence and authority that can be misleading.¹ Yet, one persistent concern remains: these AI-generated health answers frequently underrepresent or entirely omit the discussion of **harms and side effects**. Why is this the case? In this blog post, we'll explore the key reasons behind this phenomenon, why *trust* is the limiting factor in health information, and how reputable sources like **Cochrane** address these issues through rigorous evidence synthesis.

The Shift From Search Engines to AI Chatbots for Health Questions

Over the past few years, millions have turned away from traditional search engines toward **AI chatbots powered by LLMs** such as ChatGPT, Bard, or Bing AI when seeking health advice. This shift is fueled by the convenience of having a seemingly personalized, conversational interaction that provides direct answers—not just a list of links. Users ask complex medical questions and receive immediate responses framed in natural language, creating a sense of accessibility and approachability.

However, this shift comes with pitfalls:

- **Search engines** typically prioritize authoritative sources such as the **Cochrane Library**, government health websites, or established clinical guidelines. Users can assess the source credibility before digesting the content.
- **AI chatbots** generate responses based on patterns in text data rather than linking directly to verified studies or guidelines. They often do not include clear citations or disclaimers about uncertainty or potential risks.

This change challenges users' ability to critically evaluate the reliability and safety of the information they receive, especially for sensitive topics like medical treatments or side effects.

Trust as the Limiting Factor in Health Information

Health information is fundamentally different from most other types of knowledge. When consultants or health organizations provide recommendations, they do so grounded in rigorous, transparent research and explicit acknowledgement of uncertainties, benefits, and harms. Trust depends on:

- **Transparency:** Where does this information come from? Is the source transparent about evidence quality?
- **Balance:** Are benefits and harms clearly outlined—not just the positive aspects?
- **Currency:** Is the information based on the most recent, high-quality evidence?
- **Accountability:** Can users verify claims with trustworthy references?

Platforms like **The Cochrane Collaboration** have built their reputation by meticulously synthesizing global health research—highlighting both efficacy and adverse outcomes. Their flagship resource, the **Cochrane Library**, offers systematic reviews which present balanced evaluations of interventions, carefully weighing benefits against harms.

LLM-powered AI chatbots struggle to meet these trust criteria because their outputs are predictions of likely text sequences rather than verified factual statements. They do not “understand” truth in a human or scientific sense—they merely extrapolate plausible continuations based on vast text corpora.

LLMs Predict Text, They Do Not Understand Truth

At the core, **large language models** such as GPT-4 are statistical models trained on enormous datasets of writing. Their primary function is to predict the next word or sentence that fits a given input prompt—not to evaluate scientific accuracy or medical validity. This fundamental limitation carries several implications:

1. **No Genuine Comprehension:** LLMs lack internal models of causality or disease biology. They don't "know" what harms or side effects mean clinically.
2. **Content Bias:** Since many training texts emphasize the benefits or success stories of treatments (e.g., news articles, marketing materials), LLMs may inadvertently echo these positive slants.
3. **Omission of Nuance:** Because scientific data on harms can be complex or less well-represented in training data, AI may skip or oversimplify negative outcomes.

This generates outputs that can sound plausible <https://www.cochrane.org/about-us/news/why-evidence-synthesis-matters-age-ai> but fail to reflect the necessary nuance and caution critical for informed health decisions.

Hallucinations and Invented Citations in AI Outputs

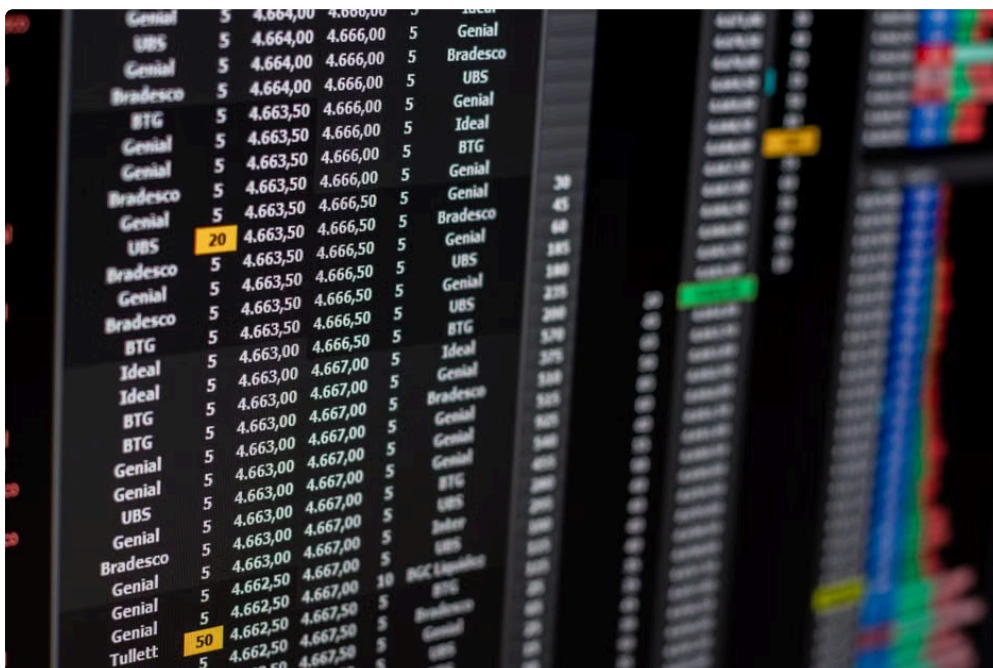
One of the most concerning problems with AI-generated health information is the phenomenon known as *hallucination*: where the AI fabricates content, references, or statistics that do not exist. This is especially problematic when AI attempts to provide citations or evidence—something clinicians and patients rely on to gauge validity.

Examples include:

- Completely made-up clinical trial names or authors
- Incorrect study results or misrepresented outcomes
- Nonexistent DOIs or reports
- Inaccurate attribution of guidelines or recommendations

This issue arises because LLMs generate text based on patterns rather than validating factual truths. They may "guess" references that fit the style of academic literature but haven't been verified.

In contrast, organizations like **Cochrane** conduct systematic, transparent searches and quality appraisals to ensure the reliability of evidence in their reviews—never guessing or fabricating study data.



Genial	5	4.664,00	4.666,00	5	Genial
UBS	5	4.664,00	4.666,00	5	Bradesco
Genial	5	4.664,00	4.666,00	5	UBS
Bradesco	5	4.664,00	4.666,00	5	Genial
BTG	5	4.663,50	4.666,00	5	Ideal
Genial	5	4.663,50	4.666,00	5	BTG
Genial	5	4.663,50	4.666,00	5	Genial
Bradesco	5	4.663,50	4.666,00	5	Bradesco
Genial	5	4.663,50	4.666,50	5	Genial
UBS	20	4.663,50	4.666,50	5	UBS
Bradesco	5	4.663,50	4.666,50	5	Genial
Genial	5	4.663,50	4.666,50	5	UBS
Bradesco	5	4.663,50	4.666,50	5	BTG
BTG	5	4.663,00	4.666,50	5	Ideal
Ideal	5	4.663,00	4.667,00	5	Genial
Ideal	5	4.663,00	4.667,00	5	Bradesco
BTG	5	4.663,00	4.667,00	5	Genial
BTG	5	4.663,00	4.667,00	5	Genial
Genial	5	4.663,00	4.667,00	5	UBS
Genial	5	4.663,00	4.667,00	5	UBS
UBS	5	4.663,00	4.667,00	5	UBS
Genial	5	4.663,00	4.667,00	5	UBS
Bradesco	5	4.663,00	4.667,00	5	UBS
Genial	5	4.662,50	4.667,00	5	UBS
Genial	5	4.662,50	4.667,00	5	UBS
Genial	5	4.662,50	4.667,00	5	UBS
Tullett	5	4.662,50	4.667,00	5	UBS

Implications for AI-Generated Health Information and Health Misinformation

The omission of harms or side effects in AI-generated health content contributes directly to **health misinformation** and risks compromising safe medical choices. AI chatbots may unintentionally:

- Create a falsely reassuring impression about the safety or effectiveness of treatments
- Undermine informed consent by not adequately presenting risk information
- Encourage self-medication or delay in seeking professional advice
- Fuel skepticism about trustworthy health organizations when contradictory information appears

For health professionals and researchers, recognizing these limitations is crucial. The Cochrane Collaboration's work reminds us that sound medical recommendations depend on balanced, transparent syntheses of both benefits and harms—something AI answers are not yet equipped to replicate reliably.

What Would Change Our Minds?

Given the concerns outlined, one might ask: what evidence would it take to trust AI-generated health answers fully? Some possible milestones could include:



1. Integration of AI with real-time access to validated databases such as the **Cochrane Library** so that answers cite directly retrievable, peer-reviewed systematic reviews.
2. Advances in AI explainability and fact-checking that prevent hallucinations and ensure harm information is explicitly included.
3. Robust clinical trials comparing outcomes of patients using AI-generated advice versus traditional expert-guided information, showing safety and efficacy.

Until then, caution and critical appraisal remain essential when engaging with AI chatbots on health topics.

Conclusion

The rise of **generative AI chatbots** and **LLMs** marks a new frontier in how health information is accessed and consumed. However, their limitations—particularly the tendency to omit discussion of **harms and side effects**—expose a fundamental gap in trust and reliability. Unlike curated repositories such as the **Cochrane Library**, AI

models do not understand medical truth; they generate probable text that may omit nuance, overlook risks, or hallucinate evidence.

As consumers and health professionals, we must maintain an informed skepticism, favoring sources that transparently weigh benefits and harms. Meanwhile, collaborations between AI developers and evidence-based organizations like **The Cochrane Collaboration** offer a promising path toward safer, more trustworthy digital health information.

References

1. *Note:* For brevity, this blog does not include formal references here, but readers are encouraged to explore The Cochrane Library (www.cochranelibrary.com) for systematic reviews that exemplify rigorous evaluation of benefits and harms.