



Should we accept systematically the text provided by Chat GPT or Perplexity?

Faut-il accepter systématiquement le texte fourni par Chat GPT ou Perplexity ?

Chadli Dziri¹, Abe Fingerhut²

1. *Professor Emeritus – General Surgery - Tunis University El Manar – Medical School of Tunis Director of Honoris Medical Simulation Center – Tunis (Editorial Board Member of The American Journal of Surgery)*
2. *FRCPS (g), FRCS (Ed) (hon) GuangCi Laureate Professor - Shanghai Jiao Tong University School of Medicine, Shanghai China and Medical University of Graz, Division for Surgical Research, Graz, Austria (Co-Editor-in-Chief of Annals of Laparoscopic and Endoscopic Surgery (ALES))*

Artificial Intelligence (AI), defined in simple words by Google, is the simulation of human intelligence processes by machines, especially computer systems.

In Wikipedia, Artificial intelligence applications include advanced web search engines (e.g., Google Search), recommendation systems (used by YouTube, Amazon, and Netflix), understanding human speech (such as Siri and Alexa), self-driving cars (e.g., Waymo), generative or creative tools (Chat GPT and AI art), automated decision-making, and competing at the highest level in strategic game systems (such as chess and Go), among others.

Even though widely used today (A recent Twitter (@shedntcare_), published June 22, 2023 indicated more than 170 million users!!!), these AI based writing programs may be a source of misinformation. Extreme caution is warranted before embarking on any publication with their help. As an examples, we decided to perform TWO exercises on May 15th, 2023:

1. The first Question was addressed to perplexity: Give me a heartfelt tribute to a Tunisian surgeon who passed away a few years ago. Perplexity proposed the following text:

“We are here today to pay tribute to a great Tunisian surgeon who passed away a few years ago. He was a

man passionate about his profession, who devoted his life to saving the lives of others. He was renowned for his expertise and dedication to his patients. He trained many surgeons and left an indelible mark on the field of medicine in Tunisia. We will always remember his smile, his kindness, and his generosity. Today, we pay tribute to him and express our gratitude for all he did for medicine and for the patients he cared for.”

We mention in this paragraph several adjectives imagined by AI without having previous confirmation by other information concerning our dead colleague.

2. We introduced in Chat GPT and Perplexity platforms the question: Provide me the content for ten slides explaining heterogeneity in Meta-Analysis?

We obtained slides that look alike except for how to evaluate the heterogeneity:

- Chat GPT provided the following “Several statistical measures are commonly used to quantify heterogeneity:
 - a. Cochran’s Q statistic: It tests the null hypothesis that all studies are estimating the same effect size.
 - b. I-squared (I²) statistic: It quantifies the proportion of total variability in effect sizes that is due to heterogeneity rather than chance”.

Correspondance

Chadli Dziri

Professor Emeritus – General Surgery - Tunis University El Manar – Medical School of Tunis Director of Honoris Medical Simulation Center – Tunis
Editorial Board Member of The American Journal

Email: cdziri@tn.honoris.net

- Perplexity provided another statistical method “Heterogeneity may arise from methodological or clinical heterogeneity” and “Prediction intervals from random-effects meta-analyses are a useful device for presenting the extent of between-study variation.”

In practice, there is a fundamental difference between these two responses concerning well defined scientific concepts. Borenstein and Higgins (1) stated that: “ I^2 is not an absolute measure of heterogeneity”. However, predictive intervals are the appropriate statistical method to evaluate the heterogeneity and it is wrong to consider I^2 statistics (1-4). Therefore, Artificial intelligence is not able to make difference between the wheat and the chaff, on other words, the correct and the wrong concept.

A recent editorial (5) focused on Artificial intelligence and medical writing. The author asked “where are we going?”. He concluded that artificial intelligence could generate enticing texts, but was incapable of human reasoning and argumentation, so necessary for medical writing and proposing advice or decision-making policies in medicine, and, worse, can make false citations, or even invent references, at the same time...

Extreme caution for the time being!!!

REFERENCES

1. Borenstein M, Higgins JP, Hedges LV, Rothstein HR. Basics of Meta-Analysis: I^2 is not an absolute measure of heterogeneity. *Res Synth Methods*. 2017 Mar;8(1):5-18. doi: 10.1002/jrsm.1230. Epub 2017 Jan 6. PMID: 28058794.
2. Dziri C, Slim K. What the surgeons need to know about meta-analyses in 2023, beyond the simple Odds ratio. *J Visc Surg*. 2022 Dec;159(6):480-485. doi: 10.1016/j.jvisc-surg.2022.10.003. Epub 2022 Nov 1. PMID: 36333183.
3. Dziri C. Evaluation of heterogeneity in Meta-Analysis. *Colorectal Dis*. 2023 May;25(5):1037-1038. doi: 10.1111/codi.16473. Epub 2023 Jan 19. PMID: 36626069.
4. Dziri C. How to assess heterogeneity for a Meta-Analysis? *Tunis Med*. 2022 Mai;100(5):353. PMID: 36206082; PMCID: PMC9552239.
5. Fingerhut A, Winter DC. Artificial intelligence and medical writing: where are we going? *Br J Surg*. 2023 Jun 13: znad169. doi: 10.1093/bjs/znad169. Epub ahead of print. PMID: 37310139.