Correspondence

Reply to Dages et al: You AIn't using it right—artificial intelligence progress in allergy

To the Editor:

We (ChatGPT and I) read with interest the letter by Dages et al regarding the limitations of the artificial intelligence (AI) tool ChatGPT.[1] We concur that the tool did not accurately address their query on cefazolin administration in a patient with penicillin allergy. However, we believe that this outcome reflects the current limitations of the tool's use as a search engine rather than limitations of the AI model itself. We fully agree that the submission by Dages et al[1] is typical and acknowledge that results can often deviate convincingly from what is expected. Thus, the report by Dages et al[1] is an important cautionary example.

Because language models generate responses based on data likelihood and not factual accuracy, it is important that we understand their function so that we can use them effectively.

To demonstrate, we used ChatGPT v4 for a more complex approach to answer the same question asked by Dages et al.[1] We propose that one must (for now) explicitly pose queries such as, "Can we automate downloads from PubMed, analyze the content, and interpret the result concerning the safety of cefazolin for patients with penicillin allergy?"

We began by asking, "Which R package is able to query PubMed and return abstracts?" (R being our programming language of choice). ChatGPT automatically generated R code that could gather relevant abstracts from PubMed by using our key words. From here, we incrementally asked for additional code to process and interpret the information.

The full code set provided us with an R computer program, written by ChatGPT, that performs the automated PubMed queries relative to the question and saves all abstracts and references. It then performs term frequency and correlation analysis on the texts. Lastly, the text corpus is used as input for ChatGPT's interpretation of the original question to return a response based on the referenced summaries. The result intuitively demonstrates, to the human reader, what the text-based information consisted of before being interpreted by ChatGPT's model.

The final result consists of Fig 1, illustrating the summarized content from PubMed (Fig 1, *A-D*), a balanced text report based on those summaries (Fig 1, *E*), and the list of publications used in the analysis (Fig 1, *F*).

The supporting data were automatically sourced for the first 20 PubMed results, as well as the Uniform Resource Locators (URLs) and PubMed identifiers (PMIDs) printed and automatically imported into a citation manager. We limited the method to the default of 20 publication abstracts so that it would be easy to replicate; however, using the ChatGPT application programming interface (API) allows the system to be automated, with relative ease, to run on far larger data sets.

Our findings show that AI can provide valuable information for problem solving. We believe that the concerns expressed by

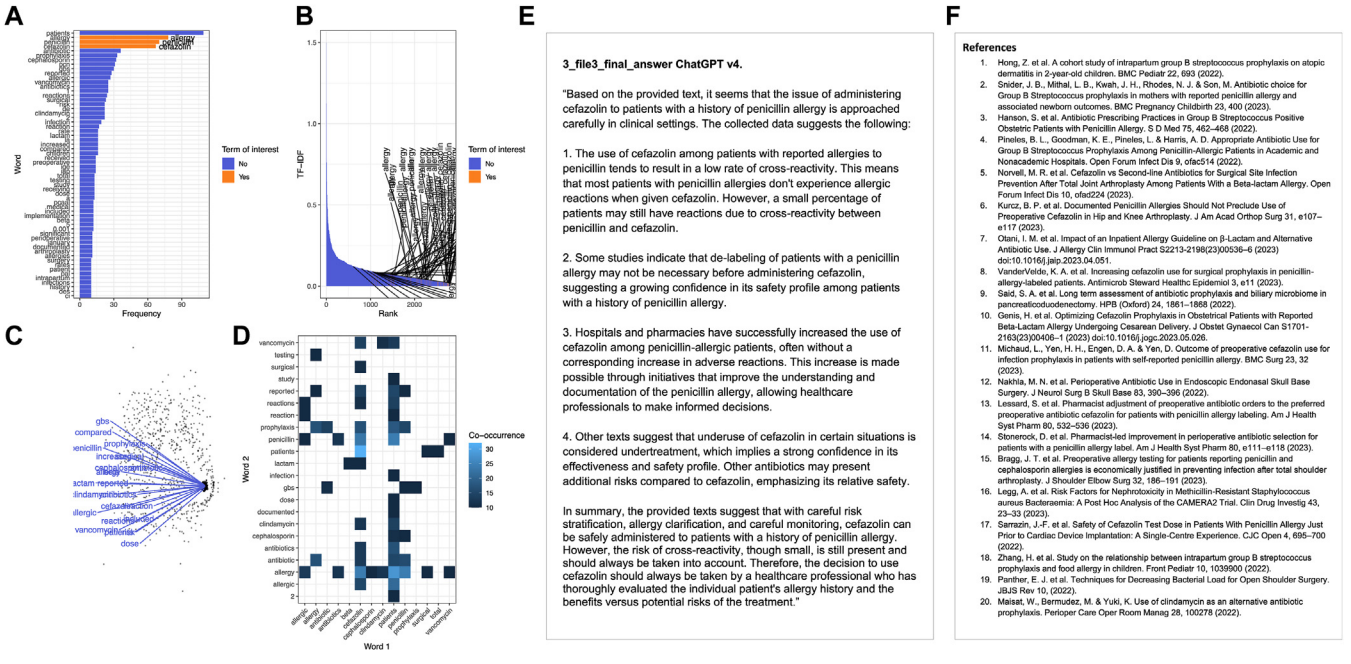

FIG 1. Term frequency in automated PubMed queries and final interpretation. **A,** The most common words in abstracts related to cefazolin and penicillin allergy (quantile 0.95 is shown). **B,** Term frequency–inverse document frequency (TF-IDF) for terms of interest in abstracts related to the search terms "cefazolin" and "penicillin" "allergy" (text labels highlight matches to the 3 terms). **C,** Network plot of edge weights based on the frequency of co-occurrence in abstracts related to cefazolin and penicillin allergy (quantile 0.98 is shown). **D,** Heatmap of term co-occurrence in abstracts (co-occurrence threshold of >9 is shown). This same data set was used throughout for subsequent interpretation by ChatGPT. **E,** Final response by ChatGPT based on summaries from abstracts. **F,** References used for ChatGPT's final response.

Dages et al[1] are justified because most users initially assume that responses to queries are supposed to be fact based. By adjusting the approach, we can use AI tools to improve how we find and analyze information.

However, we must highlight the fact that as it stands, the process is time-intensive—our session required 2 hours. It would be unrealistic to expect every user to devote such effort. We anticipate the advent of more user-friendly tools and interfaces, making data retrieval and analysis accessible and efficient.

DISCLOSURE STATEMENT

Disclosure of potential conflict of interest: The author declares no relevant conflicts of interest.

Dylan Lawless, PhD[a,b]

From [a]the Department of Intensive Care and Neonatology, University Children's Hospital Zurich, University of Zurich, Zurich, Switzerland; and [b]the Global Health Institute, School of Life Sciences, Ecole Polytechnique Fédérale de Lausanne, Lausanne, Switzerland. E-mail: Dylan.Lawless@epfl.ch.

REFERENCE

1. Dages KN, Ender EC, Park MA, Li JT. It AIn't perfect: artificial intelligence errors in allergy. J Allergy Clin Immunol 2023;152:817-8.

https://doi.org/10.1016/j.jaci.2023.09.023