

ALLERGY, SOFTWARE

A USELESS STUDY OF AI ACCURACY

31.03.2025

Journal Pre-proof

ChatGPT performance on 120 interdisciplinary allergology questions – systematic evaluation with clinical error impact assessment for critical erroneous AI-guided chatbot-advice

Sonja Mathes, Dr. med., MD, Sebastian Seurig, MD, Friederike Bluhme, MD, Kirsten Beyer, Dr. med., MD, Prof., Fabian Heizmann, Dr. med., MD, H., Manfred Wagner, Dr. med., MD, Ina Neugärtner, Dr. med., MD, Tilo Biedermann, Dr. med., MD, Prof., Ulf Darsow, Dr. med., MD, Prof.

PII: S2213-2198(25)00280-6

DOI: <https://doi.org/10.1016/j.jaip.2025.03.030>

Reference: JAIP 5921

To appear in: *The Journal of Allergy and Clinical Immunology: In Practice*

Received Date: 4 January 2025

Revised Date: 28 February 2025

Accepted Date: 19 March 2025



[https://www.jaci-inpractice.org/article/S2213-2198\(25\)00280-6/pdf](https://www.jaci-inpractice.org/article/S2213-2198(25)00280-6/pdf)

Although promised, the supplement with the questions is missing and there is no information in the methods how the authors prompted – so basically useless as methods cannot be replicated.

And at the end, chatGPT is only a mirror of how good or bad these experts informed the public...

