



## CHAT GPT 4.0 DAN CHAT GPT 3.5 DALAM MENJAWAB PERTANYAAN MEDIS

Joko Tri Atmojo<sup>1\*</sup>, Amallia Wijiwinarsih<sup>2</sup>, Ika Yulianti<sup>3</sup>, Livana PH<sup>4</sup>, Aris Widiyanto<sup>1</sup>

<sup>1</sup>Sekolah Tinggi Ilmu Kesehatan Mamba'ul Ulum Surakarta, Jl Ring Road Utara, Km 0.3, Tawang Sari, Mojosongo, Jebres, Kota Surakarta, Jawa Tengah, Indonesia 57127

<sup>2</sup>Program Magister Ilmu Kesehatan Masyarakat, Sekolah Pascasarjana, Universitas Sebelas Maret Surakarta, Jl Ir. Sutami 36 A, Surakarta, 57126, Indonesia

<sup>3</sup>Universitas Borneo Tarakan, Jl. Amal Lama Pantai Amal, Tarakan, Kalimantan Utara, 77115, Indonesia

<sup>4</sup>Sekolah Tinggi Ilmu Kesehatan Kendal, Jl Laut 31 Kendal, Jawa Tengah 51311, Indonesia

\*[jokotriatmojo1@gmail.com](mailto:jokotriatmojo1@gmail.com)

### ABSTRAK

Kecerdasan Buatan (AI) merupakan bidang ilmu komputer yang dikhususkan untuk memecahkan masalah kognitif yang umumnya terkait dengan kecerdasan manusia. Chat Generative Pre-Trained Transformer (Chat GPT) adalah model pemrosesan bahasa alami produk kecerdasan buatan yang dikembangkan oleh OpenAI. Chat GPT berpotensi membantu menjawab pertanyaan medis dengan tingkat kesesuaian yang akurat. Penelitian ini bertujuan untuk mengetahui perbandingan akurasi kecerdasan buatan Chat GPT-4 dan Chat GPT-3 dalam menjawab pertanyaan medis. Penelitian ini merupakan penelitian metaanalisis dan telaah sistematis dengan menggunakan diagram PRISMA. Pencarian studi primer melalui beberapa indexing database diantaranya: PubMed, Google Scholar dan BASE. Kata kunci yang digunakan untuk mempermudah pencarian artikel yaitu; "Chat GPT AND Medical Question", atau "Chat GPT AND Accuracy", atau "Chat GPT AND Medical Question AND Accuracy". Kriteria inklusi penelitian ini adalah artikel yang terpublikasi menggunakan desain studi cross-sectional dari bulan Januari 2023-Agustus 2024. Analisis statistik yang digunakan pada penelitian ini menggunakan program metaanalisis RevMan 5.4.1 dengan pendekatan fixed effect dan Random effect serta menyajikan data funnel plot dan forest plot. Hasil penelitian menunjukkan bahwa Chat GPT-4 memiliki tingkat akurasi yang lebih tinggi dalam menjawab pertanyaan medis. Tingkat akurasi Chat GPT-4 dalam menjawab pertanyaan medis menunjukkan 3.07 kali lebih tinggi dibandingkan dengan Chat GPT-3 (OR: 3.07; 95% CI: 2.20-4.30; p<0.0001) dan signifikan secara statistik. Forest plot tersebut juga menunjukkan heterogenitas estimasi efek antar studi yang tinggi (I = 81%). Funnel plot menunjukkan terdapat bias publikasi yang cenderung melebih-lebihkan efek yang sesungguhnya (overestimate). Metaanalisis dari 9 studi menunjukkan bahwa Chat GPT-4 lebih akurat dibandingkan Chat GPT-3 dalam menjawab pertanyaan medis.

Kata kunci: chat gpt; kecerdasan buatan; pertanyaan medis

### CHAT GPT 4.0 AND CHAT GPT 3.5 IN ANSWERING MEDICAL QUESTIONS

#### ABSTRACT

Artificial Intelligence (AI) is a field of computer science devoted to solving cognitive problems commonly associated with human intelligence. Chat Generative Pre-Trained Transformer (Chat GPT) is an artificial intelligence product natural language processing model developed by OpenAI. Chat GPT has the potential to help answer medical questions with an accurate match rate. This study aims to compare the accuracy of Chat GPT-4 and Chat GPT-3 artificial intelligence in answering medical questions. This research is a meta-analysis and systematic review using the PRISMA diagram. The search for primary studies through several indexing databases including: PubMed, Google Scholar and BASE. The keywords used to facilitate the search for articles are; "Chat GPT AND Medical Question", or "Chat GPT AND Accuracy", or "Chat GPT AND Medical Question AND Accuracy". The inclusion criteria for this study were articles published using a cross-sectional study design from January 2023-August 2024. The statistical analysis used in this study used the RevMan 5.4.1 meta-analysis program with a fixed effect and random effect approach and presented funnel plot and forest

*plot data. The results showed that Chat GPT-4 has a higher level of accuracy in answering medical questions. The accuracy of Chat GPT-4 in answering medical questions was 3.07 times higher than Chat GPT-3 (OR: 3.07; 95%CI: 2.20-4.30;  $p < 0.0001$ ) and statistically significant. The forest plot also showed high heterogeneity of effect estimates between studies ( $I = 81\%$ ). The funnel plot showed a publication bias that tended to overestimate the true effect. Meta analysis of 9 studies showed that Chat GPT-4 was more accurate than Chat GPT-3 in answering medical questions.*

*Keywords: artificial intelligence; chat gpt; medical questions*

## **PENDAHULUAN**

Kecerdasan Buatan (AI) merupakan bidang ilmu komputer yang dikhususkan untuk memecahkan masalah kognitif yang umumnya terkait dengan kecerdasan manusia, seperti pembelajaran, penciptaan, dan pengenalan gambar (Xu et al., 2021). Istilah kecerdasan buatan dicetuskan oleh John McCarthy pada tahun 1956 (Mintz & Brodie, 2019). Pada tahun 1980-an kecerdasan buatan mulai dikembangkan untuk membantu menjawab pertanyaan klinis di bidang kesehatan (Davenport & Kalakota, 2019). Chat Generative Pre-Trained Transformer (ChatGPT) adalah model pemrosesan bahasa alami produk kecerdasan buatan yang dikembangkan oleh OpenAI dan dirilis pada November 2022 (Bagde et al., 2023). Sejak diluncurkan, Chat GPT telah mendapatkan popularitas yang luar biasa dalam sepuluh hari pertama bulan Januari 2023, Chat GPT diunduh sebanyak 3.771 kali oleh pengguna global (Thorp, 2023). Chat GPT telah memperoleh pengakuan signifikan di bidang perawatan kesehatan, karena telah terlisensi oleh the United States Medical Licensing Exam (USMLE). Chat GPT berpotensi membantu menjawab pertanyaan medis dengan tingkat kesesuaian yang akurat (Kung et al., 2023).

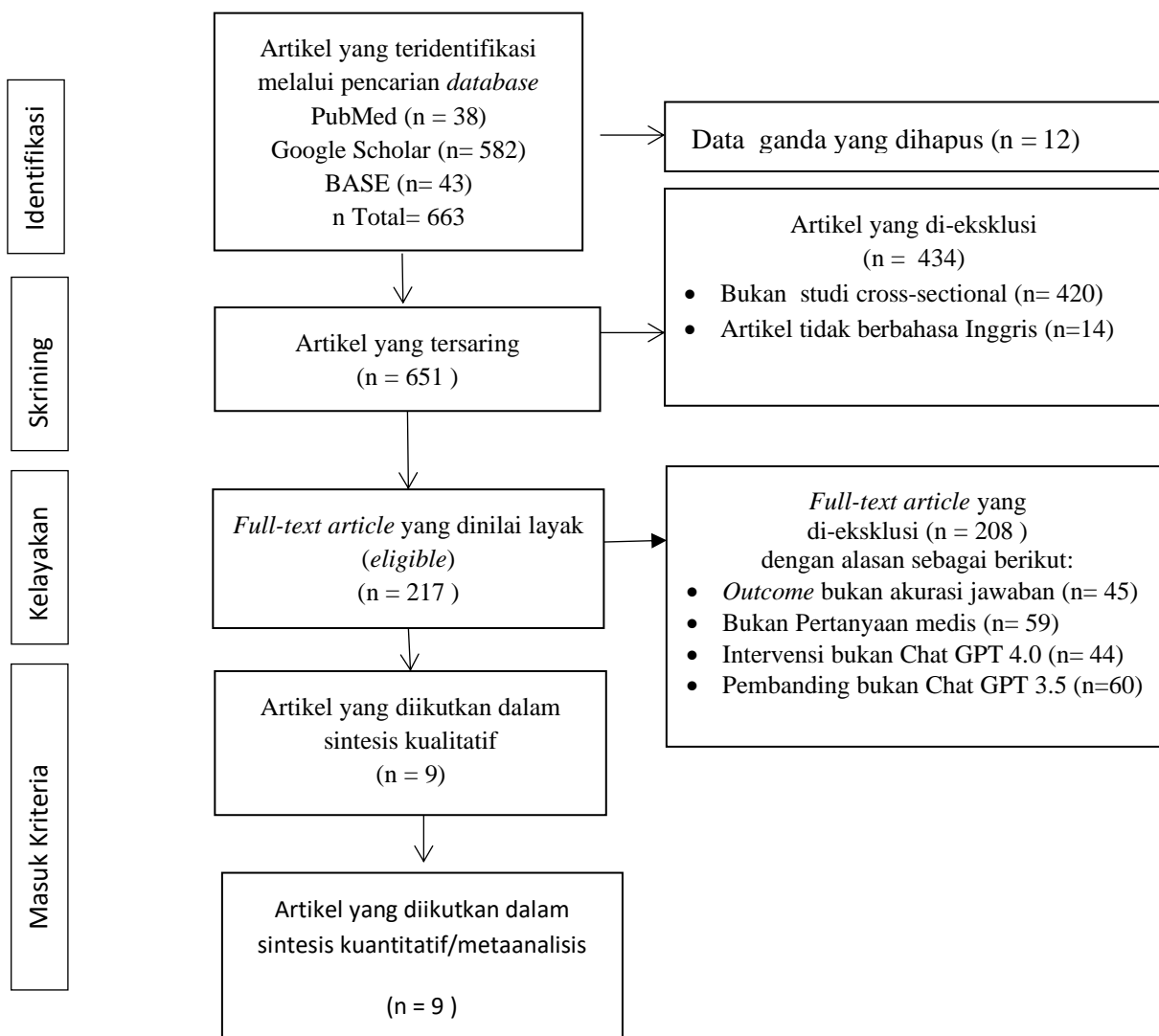
OpenAI telah mengembangkan dua model Chat GPT yaitu; GPT-3.5 dan GPT-4. Versi terbaru, Chat GPT-4 menawarkan peningkatan efisiensi dan akurasi dibandingkan pendahulunya berupa pemahaman kontekstual yang lebih baik, peningkatan kefasihan berbahasa, dan basis pengetahuan yang lebih luas (Wójcik et al., 2023), tujuan penelitian ini adalah apakah GPT-4 memiliki kemampuan yang lebih baik dibandingkan dengan Chat GPT-3 dalam menjawab pertanyaan medis?.

## **METODE**

Penelitian ini merupakan penelitian metaanalisis dan telaah sistematis dengan menggunakan diagram PRISMA. Pencarian studi primer melalui beberapa indexing database diantaranya: PubMed, Google Scholar dan BASE. Kata kunci yang digunakan untuk mempermudah pencarian artikel yaitu; "Chat GPT AND Medical Question", atau "Chat GPT AND Accuracy", atau "Chat GPT AND Medical Question AND Accuracy". Kriteria inklusi penelitian ini adalah artikel yang terpublikasi menggunakan desain studi cross-sectional dari bulan Januari 2023 - Agustus 2024, artikel yang terpublikasi menggunakan Bahasa Inggris, dan free full text atau artikel yang dapat diunduh secara gratis (open access). Artikel primer akan tereliminasi sesuai dengan kriteria eksklusi yaitu; artikel dengan desain studi RCT, kasus kontrol, cohort, quasi eksperimental, study protocol atau pilot study. Artikel yang dipublikasikan dengan Bahasa selain Bahasa Inggris juga akan memenuhi kriteria eksklusi. Analisis statistik yang digunakan pada penelitian ini menggunakan program metaanalisis RevMan 5.4.1 dengan pendekatan fixed effect dan Random effect serta menyajikan data funnel plot dan forest plot. Dengan langkah Tentukan pertanyaan penelitian. Mendefinisikan pertanyaan penelitian adalah langkah pertama dalam melakukan analisis meta, Cari studi yang relevan, Saring studi untuk dimasukkan, Ekstrak data, Analisis data, dan Menafsirkan hasil.

**HASIL**

Hasil akhir artikel yang diikutkan dalam sintesis metaanalisis pada PRISMA Flow Diagram berjumlah 9 dari 663 artikel. Berdasarkan Gambar 1. artikel yang teridentifikasi melalui pencarian *database* PubMed, Google Scholar dan BASE sebanyak 663 artikel, setelah data ganda sebanyak 12 artikel dihapus tersisa 651 artikel. Kemudian artikel di eksklusikan sebanyak 434 artikel dengan alasan bukan studi cross sectional, tidak menggunakan bahasa Inggris dan *Full-text article* yang dinilai layak (*eligible*) sebanyak 217 artikel. *Full-text article* yang dinilai layak kemudian di telaah secara mendalam untuk mendapatkan artikel yang akan diikutkan dalam sintesis kualitatif dan kuantitatif. 9 dari 217 artikel akhirnya dinilai layak diikutkan dalam sintesis kualitatif dan kuantitatif. Hasil akhir artikel yang didapat kemudian dinilai kualitasnya dengan menggunakan *critical appraisal* dari The Joanna Briggs Institute (JBI) dengan hasil skor *check list* semua layak digunakan untuk *systematic review* dan metaanalisis.



Gambar 1. PRISMA Flow Diagram

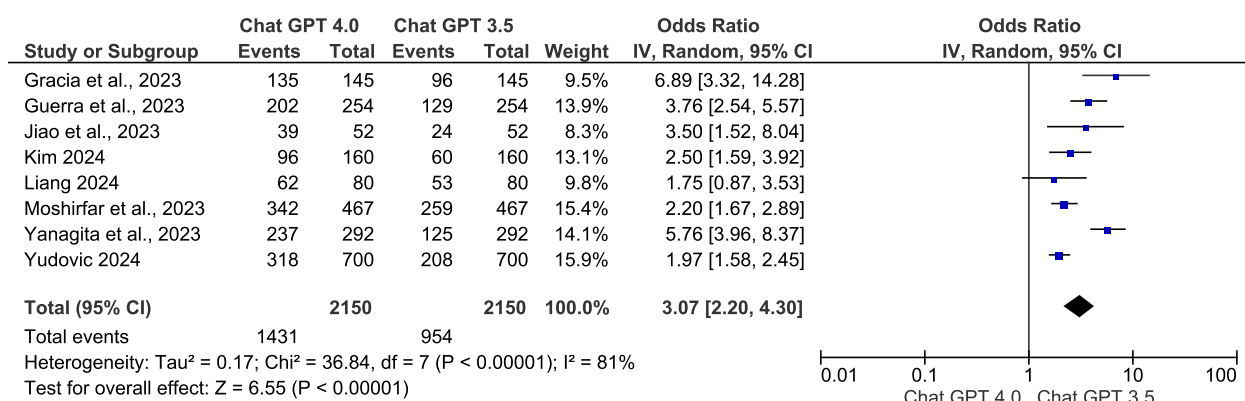
Tabel 1.

## Ringkasan Artikel yang Layak di masukan dalam Sintetis Kuantitatif tentang Perbandingan Akurasi Chat GPT-4 dan Chat GPT-3 dalam Menjawab Pertanyaan Medis.

No	Penulis	Tahun	Lokasi	Sampel	Hasil
1.	García et al., 2023	2023	Spanyol	145	Dari 143 pertanyaan, chat GPT-4 menjawab 134 pertanyaan dengan benar (93,71%), menunjukkan tingkat akurasi yang tinggi. Di sisi lain, Chat GPT-3 menjawab 95 pertanyaan dengan benar (66,43%), menunjukkan tingkat kinerja yang agak lebih rendah dibandingkan dengan GPT-4.
2.	Guerra et al., 2023	2023	USA	254	Chat GPT-4 berhasil menjawab 91% pertanyaan dari Congress of Neurological Surgeons dengan akurasi jawaban 76.6%. Hasil ini mengungguli Chat GPT-3 dan mahasiswa kedokteran.
3.	Jiao et al., 2023	2023	USA	52	Chat GPT-4 secara signifikan mengungguli chat GPT-3 (75% versus 46%, $p < 0,01$ ) dalam menjawab pertanyaan medis tentang neuro-oftalmologi.
4.	Moshirfar et al., 2023	2023	USA	467	GPT-4 memiliki kinerja yang jauh lebih baik daripada GPT-3.5 (73,2% vs. 55,46%; $p < 0,001$ ) dan manusia (73,23% vs. 58,31%, $p < 0,001$ ) pada 467 pertanyaan StatPearls oftalmologi.
5.	Takagi et al., 2023	2023	Jepang	254	Chat GPT-4 mengungguli GPT-3.5 secara signifikan sebesar 29,1% ( $P < .001$ ).
6.	Yanagita et al., 2023	2023	Jepang	292	Tingkat respons yang benar untuk GPT-4 adalah 81,5% (237/292), yang secara signifikan lebih tinggi daripada tingkat untuk GPT-3,5, 42,8% (125/292). Selain itu, GPT-4 melampaui standar kelulusan (>72%) untuk the National Medical Licensing Examination (NMLE), yang menunjukkan potensinya sebagai alat bantu pengambilan keputusan diagnostik dan terapeutik bagi dokter.
7.	Liang et al., 2024	2024	China	80	Tingkat akurasi rata-rata jawaban yang diberikan oleh ChatGPT versi 3.5 dan 4.0 masing-masing adalah 67,08% dan 77,50%. Chat GPT 4.0 mengungguli Chat GPT 3.5, dengan tingkat akurasi yang lebih tinggi dalam respons ( $p < 0,05$ ).
8.	Yudovich et al., 2024	2024	AS	700	GPT-4 menjawab 44,4% item dengan benar dibandingkan dengan 30,9% untuk GPT-3,5 ( $P < 0,00001$ ). GPT-4 memiliki akurasi yang lebih tinggi dibandingkan Chat GPT-3 dalam menjawab pertanyaan tentang urologi.

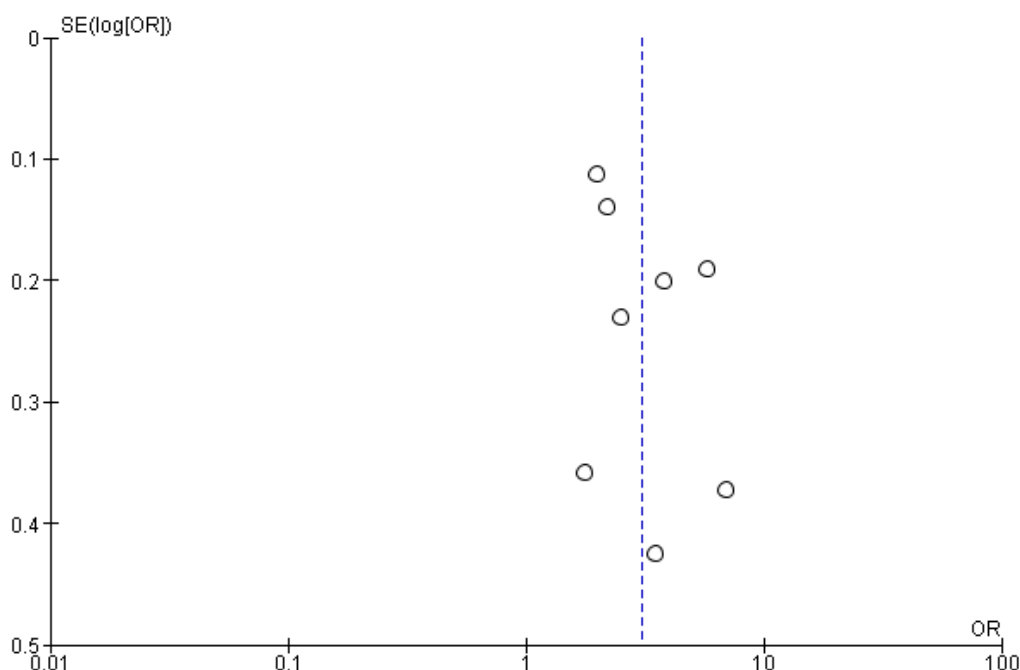
9.	Kim et al., 2024	2024	Korea Selatan	160	ChatGPT 3.5 menjawab pertanyaan dengan benar, sementara ChatGPT 4 menunjukkan tingkat akurasi sebesar 60.0% ( $p < 0,001$ ). ChatGPT 4 menunjukkan kinerja yang unggul di sebagian besar subkategori, kecuali untuk pertanyaan terkait tumor. Tingkat ketidakkonsistenan jawaban adalah 47.5% untuk ChatGPT 3.5 dan 9.4% untuk ChatGPT 4.
----	------------------	------	---------------	-----	---

*Forrest Plot* Pada Gambar 2. Menunjukkan bahwa Chat GPT-4 memiliki tingkat akurasi yang lebih tinggi dalam menjawab pertanyaan medis. Tingkat akurasi Chat GPT-4 dalam menjawab pertanyaan medis menunjukkan 3.07 kali lebih tinggi dibandingkan dengan Chat GPT-3 (OR: 3.07; 95%CI: 2.20-4.30;  $p < 0.0001$ ) dan signifikan secara statistik. Hal ini membuktikan bahwa Chat GPT-4 memiliki kemampuan yang lebih baik dibandingkan dengan Chat GPT-3 dalam menjawab pertanyaan medis.



Gambar 2. *Forrest Plot* Perbandingan Akurasi Kecerdasan Buatan Chat GPT 4.0 dan Chat GPT 3.5 dalam Menjawab Pertanyaan Medis

Funnel plot pada Gambar 3. Menunjukkan sebaran estimasi efek antar studi yang tidak seimbang di sebelah kanan dan kiri garis vertikal rata-rata estimasi. Gambar diatas menunjukkan adanya bias publikasi (overestimate). Plot sebelah kiri berjumlah 4 plot dengan standar error antara 0 dan 0.4, plot sebelah kanan berjumlah 4 plot dengan standard error antara 0 dan 0.5.



Gambar 3. *Funnel Plot* Perbandingan Akurasi Kecerdasan Buatan Chat GPT 4.0 dan Chat GPT 3.5 dalam Menjawab Pertanyaan Medis

## PEMBAHASAN

Kemajuan teknologi kecerdasan buatan telah memberikan dampak yang mendalam pada banyak industri, termasuk bidang kesehatan. Salah satu penerapan kecerdasan buatan dalam dunia medis yaitu penggunaan model percakapan kecerdasan buatan seperti Chat GPT yang dapat membantu dalam menjawab pertanyaan medis secara akurat dan tepat waktu (Moshirfar et al., 2023). Chat GPT dapat berfungsi sebagai alat yang berharga untuk pendidikan kedokteran, membantu para profesional kesehatan tetap mengikuti perkembangan terbaru dan meningkatkan praktik klinis mereka secara keseluruhan (Kung et al., 2023). Studi ini memberikan gambaran secara menyeluruh tentang penggunaan produk kecerdasan buatan khususnya Chat GPT dalam menjawab pertanyaan-pertanyaan medis yang dapat berdampak di masa depan untuk pengembangan bidang kesehatan maupun kedokteran (Bagde et al., 2023; Kung et al., 2023). Studi ini menunjukkan bahwa Chat GPT dapat memberikan respon yang akurat dalam menjawab pertanyaan medis dalam berbagai sub bidang kedokteran seperti urologi, neuro-oftamologi, syaraf dan lain sebagainya (Mago & Sharma, 2023; Moshirfar et al., 2023; Takagi et al., 2023). Temuan ini menunjukkan bahwa Chat GPT berpotensi menjadi alat yang berharga bagi penyedia layanan kesehatan dan peneliti medis dalam memfasilitasi diagnosis, pengobatan, dan pencegahan berbagai penyakit (Wójcik et al., 2023).

Temuan dalam studi ini menunjukkan bahwa Chat GPT-4 memiliki tingkat akurasi yang lebih baik dibandingkan dengan Chat GPT-3, hal ini menunjukkan kemajuan yang pesat perkembangan kecerdasan buatan dalam menjawab pertanyaan medis yang substansial. Temuan ini didukung oleh penelitian sebelumnya bahwa Chat GPT-4 memiliki tingkat akurasi yang lebih tinggi dibandingkan dengan Chat GPT-3 dalam menjawab pertanyaan pertanyaan medis (Guerra et al., 2023; Jiao et al., 2023; Kim et al., 2024; Liang et al., 2024; Madrid-García et al., 2023b; Moshirfar et al., 2023; Yudovich et al., 2024).

## SIMPULAN

Penelitian metaanalisis ini dilakukan dengan menggunakan 9 artikel yang berasal dari negara USA, AS, Jepang, Spanyol dan China. Seluruh penelitian tersebut diambil dengan desain cross-sectional. Total sampel sebesar 2150 pertanyaan medis. Forest plot menunjukkan Chat GPT-4 memiliki tingkat akurasi yang lebih tinggi dalam menjawab pertanyaan medis. Tingkat akurasi Chat GPT-4 dalam menjawab pertanyaan medis menunjukkan 3.07 kali lebih tinggi dibandingkan dengan Chat GPT-3 (OR: 3.07; 95%CI: 2.20-4.30;  $p < 0.0001$ ) dan signifikan secara statistik. Forest plot tersebut juga menunjukkan heterogenitas estimasi efek antar studi yang tinggi ( $I^2 = 81\%$ ). Funnel plot menunjukkan terdapat bias publikasi yang cenderung melebih-lebihkan efek yang sesungguhnya (overestimate).

## DAFTAR PUSTAKA

- Bagde, H., Dhopte, A., Alam, M. K., & Basri, R. (2023). A systematic review and meta-analysis on ChatGPT and its utilization in medical and dental research. *Heliyon*, 9(12). <https://doi.org/10.1016/j.heliyon.2023.e23050>
- Davenport, T., & Kalakota, R. (2019). DIGITAL TECHNOLOGY The potential for artificial intelligence in healthcare. In *Future Healthcare Journal* (Vol. 6, Issue 2).
- Guerra, G. A., Hofmann, H., Sobhani, S., Hofmann, G., Gomez, D., Soroudi, D., Hopkins, B. S., Dallas, J., Pangal, D. J., Cheok, S., Nguyen, V. N., Mack, W. J., & Zada, G. (2023). GPT-4 Artificial Intelligence Model Outperforms ChatGPT, Medical Students, and Neurosurgery Residents on Neurosurgery Written Board-Like Questions. *World Neurosurgery*, 179, e160–e165. <https://doi.org/10.1016/J.WNEU.2023.08.042>
- Jiao, C., Edupuganti, N. R., Patel, P. A., Bui, T., & Sheth, V. (2023). Evaluating the Artificial Intelligence Performance Growth in Ophthalmic Knowledge. *Cureus*. <https://doi.org/10.7759/cureus.45700>
- Kim, S. E., Lee, J. H., Choi, B. S., Han, H.-S., Lee, M. C., & Ro, D. H. (2024). Performance of ChatGPT on Solving Orthopedic Board-Style Questions: A Comparative Analysis of ChatGPT 3.5 and ChatGPT 4. *Clinics in Orthopedic Surgery*, 16(4), 669. <https://doi.org/10.4055/cios23179>
- Kung, T. H., Cheatham, M., Medenilla, A., Sillos, C., De Leon, L., Elepaño, C., Madriaga, M., Aggabao, R., Diaz-Candido, G., Maningo, J., & Tseng, V. (2023). Performance of ChatGPT on USMLE: Potential for AI-assisted medical education using large language models. *PLOS Digital Health*, 2(2), e0000198. <https://doi.org/10.1371/journal.pdig.0000198>
- Liang, R., Zhao, A., Peng, L., Xu, X., Zhong, J., Wu, F., Yi, F., Zhang, S., Wu, S., & Hou, J. (2024). Enhanced Artificial Intelligence Strategies in Renal Oncology: Iterative Optimization and Comparative Analysis of GPT 3.5 Versus 4.0. *Annals of Surgical Oncology*, 31(6), 3887–3893. <https://doi.org/10.1245/S10434-024-15107-0/METRICS>
- Madrid-García, A., Rosales-Rosado, Z., Freites-Nuñez, D., Pérez-Sancristóbal, I., Pato-Cour, E., Plasencia-Rodríguez, C., Cabeza-Osorio, L., Abasolo-Alcázar, L., León-Mateos, L., Fernández-Gutiérrez, B., & Rodríguez-Rodríguez, L. (2023a). Harnessing ChatGPT and GPT-4 for evaluating the rheumatology questions of the Spanish access exam to specialized medical training. *Scientific Reports*, 13(1). <https://doi.org/10.1038/s41598-023-49483-6>

- Madrid-García, A., Rosales-Rosado, Z., Freites-Nuñez, D., Pérez-Sancristóbal, I., Pato-Cour, E., Plasencia-Rodríguez, C., Cabeza-Osorio, L., Abasolo-Alcázar, L., León-Mateos, L., Fernández-Gutiérrez, B., & Rodríguez-Rodríguez, L. (2023b). Harnessing ChatGPT and GPT-4 for evaluating the rheumatology questions of the Spanish access exam to specialized medical training. *Scientific Reports*, 13(1). <https://doi.org/10.1038/s41598-023-49483-6>
- Mago, J., & Sharma, M. (2023). The Potential Usefulness of ChatGPT in Oral and Maxillofacial Radiology. *Cureus*, 15(7), e42133. <https://doi.org/10.7759/cureus.42133>
- Mintz, Y., & Brodie, R. (2019). Introduction to artificial intelligence in medicine. *Minimally Invasive Therapy & Allied Technologies: MITAT: Official Journal of the Society for Minimally Invasive Therapy*, 28(2), 73–81. <https://doi.org/10.1080/13645706.2019.1575882>
- Moshirfar, M., Altaf, A. W., Stoakes, I. M., Tuttle, J. J., & Hoopes, P. C. (2023). Artificial Intelligence in Ophthalmology: A Comparative Analysis of GPT-3.5, GPT-4, and Human Expertise in Answering StatPearls Questions. *Cureus*. <https://doi.org/10.7759/cureus.40822>
- Takagi, S., Watari, T., Erabi, A., & Sakaguchi, K. (2023). Performance of GPT-3.5 and GPT-4 on the Japanese Medical Licensing Examination: Comparison Study. *JMIR Medical Education*, 9. <https://doi.org/10.2196/48002>
- Thorp, H. H. (2023). ChatGPT is fun, but not an author. In *Science* (Vol. 379, Issue 6630, p. 313). American Association for the Advancement of Science.
- Wójcik, S., Rulkiewicz, A., Pruszczyk, P., Lisik, W., Poboży, M., & Domienik-Karłowicz, J. (2023). Reshaping medical education: Performance of ChatGPT on a PES medical examination. *Cardiology Journal*. <https://doi.org/10.5603/cj.97517>
- Xu, Y., Liu, X., Cao, X., Huang, C., Liu, E., Qian, S., Liu, X., Wu, Y., Dong, F., Qiu, C. W., Qiu, J., Hua, K., Su, W., Wu, J., Xu, H., Han, Y., Fu, C., Yin, Z., Liu, M., ... Zhang, J. (2021). Artificial intelligence: A powerful paradigm for scientific research. In *Innovation* (Vol. 2, Issue 4). Cell Press. <https://doi.org/10.1016/j.xinn.2021.100179>
- Yanagita, Y., Yokokawa, D., Uchida, S., Tawara, J., & Ikusaka, M. (2023). Can ChatGPT Answer Medical Questions of the National Medical Licensing Examination in Japan: Evaluation of Accuracy of ChatGPT (Preprint). *JMIR Formative Research*. <https://doi.org/10.2196/48023>
- Yudovich, M. S., Makarova, E., Hague, C. M., & Raman, J. D. (2024). Performance of GPT-3.5 and GPT-4 on standardized urology knowledge assessment items in the United States: a descriptive study. *Journal of Educational Evaluation for Health Professions*, 21, 17. <https://doi.org/10.3352/jeehp.2024.21.17>