

ANALISIS PERFORMA CHATGPT DALAM KASUS PROBLEM BASED LEARNING

Baharuddin^{1*}, Devitya Angielevi¹ dan Dita Sukmaya Prawitasari¹

¹ Fakultas Kedokteran, Universitas Surabaya, Surabaya, 60293
Jl. Raya Kalirungkut, Rungkut, Surabaya 60293.

*Email: baharuddin@staff.ubaya.ac.id

Abstrak

Kecerdasan Buatan (Artificial Intelligence) telah mengalami perkembangan pesat, terutama dalam bidang Natural Language Processing (NLP). Salah satu contoh aplikasi NLP yang menjanjikan adalah ChatGPT, yang dikembangkan oleh OpenAI. ChatGPT telah menunjukkan kemampuan yang signifikan dalam berbagai aspek, mulai dari penyelesaian matematika hingga menjawab pertanyaan umum. Dalam bidang pendidikan kedokteran, penggunaan metode Problem Based Learning (PBL) sangat penting dalam melatih kemampuan pemikiran kritis dan pemecahan masalah mahasiswa. Penelitian ini bertujuan untuk membuktikan akurasi kemampuan ChatGPT dalam memahami kasus PBL, menganalisis tingkat akurasi diagnosa dan diagnosa banding ChatGPT terhadap kasus PBL, serta menunjukkan tingkat variasi diagnosa banding. Metode penelitian yang digunakan adalah penelitian deskriptif kuantitatif. Metode penelitian yang digunakan dalam studi ini adalah penelitian eksperimental-deskriptif kuantitatif. Pada penelitian ini dilakukan pengujian akurasi terhadap aplikasi ChatGPT menggunakan kasus PBL. Hasil penelitian menunjukkan bahwa ChatGPT memiliki tingkat akurasi tinggi dalam memberikan respon diagnosa pada kasus PBL. Dalam 6 kasus PBL yang diuji, ChatGPT mencapai tingkat akurasi sekitar 92% dalam memberikan diagnosa. Namun, terdapat beberapa variasi dalam diagnosa banding yang diberikan oleh ChatGPT, yang lebih banyak dibandingkan dengan diagnosa banding faktual dari kasus PBL yang telah diujikan ke siswa sebelumnya. Hasil analisis menunjukkan bahwa ChatGPT memiliki tingkat akurasi tinggi dalam memberikan respon diagnosa pada kasus PBL. Tingkat akurasi diagnosa adalah sekitar 92%, sedangkan akurasi diagnosa banding adalah sekitar 88%. Terdapat variasi diagnosa banding yang lebih banyak dibandingkan dengan diagnosa banding faktual. Oleh karena itu, ChatGPT dapat digunakan sebagai sumber informasi tambahan untuk meningkatkan wawasan, namun tetap perlu melakukan validasi informasi.

Kata kunci: Performa, ChatGPT, Problem-Based Learning

1. PENDAHULUAN

Kecerdasan buatan (Artificial Intelligence) telah mengalami perkembangan pesat dalam beberapa tahun terakhir. Terutama dalam bidang Natural Language Processing (NLP) (Grünebaum et al., 2023). Salah satu contoh aplikasi NLP yang telah menunjukkan kemampuan menjanjikan adalah ChatGPT. Suatu aplikasi cerdas yang dikembangkan oleh perusahaan OpenAI (O'Connor & ChatGPT, 2023; OPEN AI, 2023). ChatGPT telah menunjukkan kemampuan yang sangat signifikan dalam berbagai aspek, mulai dari kemampuan penyelesaian matematik, sintesis teks hingga menjawab pertanyaan-pertanyaan umum. Bahkan aplikasi ini mencoba didorong untuk menyelesaikan berbagai pertanyaan kompleks dalam bidang pendidikan. Menjadi suatu lompatan dan sekaligus tantangan untuk para pendidik di bidang kesehatan ketika ia juga mampu melakukan hal luar biasa tersebut pada bidang pendidikan kedokteran seperti menjawab kasus PBL dengan akurasi yang tinggi.

Problem Based Learning (PBL) sejatinya adalah metode pembelajaran yang menekankan pada penggunaan kasus-kasus atau permasalahan nyata untuk mengembangkan pemikiran kritis, analisis (Servant-Miklos, 2019), dan kemampuan pemecahan masalah pada mahasiswa. Dalam konteks ilmu kedokteran, PBL merupakan metode yang diyakini efektif untuk melatih sikap dan pengetahuan kritis mahasiswa dalam mendiagnosa penyakit dan merancang rencana perawatan yang sesuai (Howard S.

Barrows & Robyn M. Tamblyn, 1980). Dalam pelaksanaannya saat ini fasilitator dan siswa saling memegang narasi kasus yang berisi informasi penting. Siswa diharapkan dapat melakukan diagnosa dengan tepat disertai dengan pengetahuan diagnosa banding.

Mengingat kemampuan dan kecanggihan ChatGPT dalam memahami dan menjawab pertanyaan dengan konteks yang kompleks, terdapat potensi untuk mengintegrasikan teknologi ini dalam proses PBL. Artinya aplikasi ini berpeluang digunakan oleh siswa dan pendidik di level preklinik. Khususnya dalam membantu mendiagnosa kasus-kasus yang diberikan. Namun sayangnya, belum ada penelitian di Indonesia yang spesifik menyelidiki kemampuan ChatGPT dalam mendiagnosa kasus pada PBL. Terutama dalam konteks ilmu kedokteran di fase preklinik. Untuk itu perlu dilakukan kajian mendalam tentang kemampuan dan akurasi ChatGPT untuk menjawab kasus PBL. Khususnya pada level diagnosa dan diagnosa banding.

2. METODE PENELITIAN

Penelitian ini mengadopsi teknik pada beberapa penelitian sebelumnya yang menggunakan atau menguji secara langsung kemampuan ChatGPT atau teknologi LLM (Grünebaum et al., 2023; Peng et al., 2023). Adapun tahapan lengkap tahapan dalam penelitian ini adalah sebagai berikut:

- Akses dan Persiapan Kasus PBL
- Akses Portal ChatGPT
- Menyalin dan Memasukkan prompt (perintah)
- Menyalin dan Tabulasi Output
- Preparasi dan Interpretasi Data

3. HASIL DAN PEMBAHASAN

Dalam penelitian ini dilakukan analisa kemampuan aplikasi yang dikembangkan oleh perusahaan OpenAi. Aplikasi tersebut adalah ChatGPT yang dalam hal ini diuji menggunakan soal kasus PBL. Uji kemampuan terhadap ChatGPT dalam dunia pendidikan dan kedokteran telah ada yang melaporkan. Bahkan Peng dan Timnya berhasil membuktikan kemampuan menjawab permasalahan klinis (Peng et al., 2023). Penelitian oleh Ramazan dkk, melaporkan adanya peningkatan pemahaman materi oleh siswa yang menggunakan ChatGPT (Yilmaz & Karaoglan Yilmaz, 2023). Meskipun demikian bahasan spesifik tentang ChatGPT dan PBL masih sangat kurang padahal dibutuhkan analisa mendalam tentang kemampuan ChatGPT dalam pembacaan kasus. Oleh karena itu, penelitian ini berupaya menganalisa performa ChatGPT yang sekaligus menjawab beberapa kekhawatiran oleh Chain dkk (Cain et al., 2023).

Jumlah kasus PBL yang digunakan dalam penelitian ini sebanyak 6 kasus. Dalam tiap kasus PBL tersebut umumnya akan disertakan 4 lembar soal dengan berbagai tingkatan kesulitan. Pada setiap lembarnya (Page 1-4) akan tersedia informasi dan dalam penelitian ini ChatGPT diminta untuk melakukan prediksi “menjawab” soal yang diberikan. Ada 2 target jawaban yang diminta: pertama adalah menentukan diagnosa dengan akurat dan kedua adalah diagnosa banding. Laporan lengkap respon ChatGPT terhadap kasus PBL dapat dilihat pada Tabel 1.

Dalam laporan respon ChatGPT di atas terlihat nilai akurasi diagnosa dan akurasi diagnosa banding. Selain itu, dalam laporan di atas telah disertakan juga nilai variasi diagnosa banding. Data-data ini penting karena bermakna keluasan dan keragaman informasi yang dimiliki oleh ChatGPT. Memberikan juga informasi tentang akurasi informasi yang disajikan kepada pengguna.

Tabel 1.Data Performa ChatGPT pada Kasus PBL

ID PBL	Item	L1	L2	L3	L4	Rerata
1	Akurasi Diagnosa	100%	100%	100%	100%	100%
1	Jumlah Diagnosa Banding	5	3	2	2	3
1	Akurasi Diagnosa Banding	100%	100%	100%	100%	100%
1	Variasi Diagnosa Banding	5,0	3,0	2,0	2,0	3,0
2	Akurasi Diagnosa	100%	100%	100%	100%	100%
2	Jumlah Diagnosa Banding	3	3	2	0	2
2	Akurasi Diagnosa Banding	100%	100%	100%	100%	100%
2	Variasi Diagnosa Banding	3,0	3,0	2,0	2,0	3
3	Akurasi Diagnosa	0%	100%	100%	100%	75%
3	Jumlah Diagnosa Banding	5	4	3	0	3
3	Akurasi Diagnosa Banding	0%	100%	100%	100%	75%
3	Variasi Diagnosa Banding	2,5	2,0	2,0	2,0	2,3
4	Akurasi Diagnosa	50%	50%	100%	100%	75%
4	Jumlah Diagnosa Banding	4	3	3	4	4
4	Akurasi Diagnosa Banding	0%	100%	100%	100%	75%
4	Variasi Diagnosa Banding	1,0	2,0	0,8	1,0	1,2
5	Akurasi Diagnosa	100%	100%	100%		100%
5	Jumlah Diagnosa Banding	4	4	4	-	4
5	Akurasi Diagnosa Banding	100%	100%	100%	-	100%
5	Variasi Diagnosa Banding	2,0	2,0	2,0	-	2,0
6	Akurasi Diagnosa	100%	100%	100%	100%	100%
6	Jumlah Diagnosa Banding	4	2	2	4	3
6	Akurasi Diagnosa Banding	100%	100%	100%	0%	75%
6	Variasi Diagnosa Banding	2,0	1,0	1,0	2,0	1,5

Sumber: Data Primer Pengujian Aplikasi.

Kasus PBL 1

Pada kasus PBL 1 diberikan score 100% pada diagnosa banding *page 1* karena diagnosa banding yang diberikan oleh ChatGPT telah sesuai dalam *domain* kasus. ChatGPT memberikan respon gangguan pernafasan pada diagnosa awal. Hal ini mampu dilakukan oleh ChatGPT karena ada beberapa kata kunci yang kemudian dikonstruksi dalam domain diagnosa. Kata kunci tersebut adalah *felt shortness of breath* yang merupakan *trigger case* untuk domain kasus asma. Kemampuan ini sebenarnya juga yang diharapkan dapat dilakukan oleh siswa saat kelas PBL berlangsung (Burgun et al., 2006). Siswa diharapkan mampu melakukan konstruksi (Anggraeni et al., 2023) dan sekaligus menegaskan “framing” domain diagnosa. Menelisik lebih jauh terhadap presisi respon terlihat bahwa ChatGPT mampu melakukan reduksi terhadap jumlah diagnosa banding, hal ini merupakan respon hasil dari akumulasi pembelajaran *machine learning* karena bertambahnya informasi yang diberikan pada *page* sebelumnya. Terlihat dari 5 DD di halaman pertama (*page 1*) menjadi 2 DD di halaman terakhir (*page 4*). Sehingga secara keseluruhan akurasi diagnosa dan diagnosa banding ChatGPT pada PBL 1 1 adalah

100% dengan tingkat variasi diagnosa banding 3x lebih banyak dibandingkan diagnosa banding faktual dari kasus PBL yang sudah diujikan ke siswa.

Kasus PBL 2

Pada kasus PBL 2 untuk page 1 ChatGPT memberikan respon angina pectoris untuk diagnosa dan 3 DD. Respon ini dihasilkan berdasarkan kata kunci *ER with intermittent chest pain*. Adapun kunci jawaban diagnosa pada Kasus PBL 2 ini adalah Infark Miokard Akut (IMA). Menariknya, jawaban IMA pada page 1 oleh ChatGPT dimasukkan pada DD namun dijadikan diagnosa pada page 2-4. Ini berarti pembelajaran dan akurasi ChatGPT tinggi dalam memberikan respon. Secara keseluruhan respon ChatGPT adalah 100% untuk diagnosa dan diagnosa banding dengan tingkat variasi diagnosa banding 3x lebih banyak dibandingkan diagnosa banding faktual dari kasus PBL yang sudah diujikan ke siswa.

Kasus PBL 3

Pada kasus PBL 3 jawaban diagnosanya adalah Idiopatik Trombositopenia Purpura dengan trigger kata kunci *baby with complaints of red spots on the body since 1 day*. Adapun ChatGPT memberikan respon dermatitis atopik untuk diagnosa dan DD sebanyak 5. Pada page 1 respon ChatGPT masih belum begitu baik karena keterbatasan informasi. Meskipun demikian, akurasi ChatGPT meningkat dan konsisten pada page 2-4. Sehingga secara keseluruhan akurasi diagnosa dan diagnosa banding berada pada angka 75% dengan tingkat variasi DD 2,3X lebih banyak dibandingkan diagnosa banding faktual dari kasus PBL yang sudah diujikan ke siswa.

Kasus PBL 4

Pada kasus PBL 4 jawaban diagnosanya adalah *Meningomyelocele* suatu kondisi gangguan pada tabung saraf. ChatGPT diminta membaca kasus page 1 dengan key trigger yaitu *with mass of her lower back since birth*. Hasilnya ChatGPT cukup optimal ini dikarenakan keterbatasan informasi. Secara keseluruhan ChatGPT menghasilkan akurasi diagnosa dan diagnosa banding pada angka 75% adapun variasi diagnosanya adalah 1,2x lebih banyak dibandingkan diagnosa banding faktual dari kasus PBL yang sudah diujikan ke siswa. Kasus PBL 4 ini dekat dengan embriologi yang juga berkaitan dengan obstetri dan ginekologi atau ilmu kandungan. Menariknya peneliti mulai melihat bahwa penggunaan ChatGPT sudah saat menjadi perhatian pada departemen ini. Para klinisi pengguna diharapkan dapat melakukan validasi pada setiap respon yang dihasilkan (Grünebaum et al., 2023).

Kasus PBL 5

Pada kasus PBL 5 diberikan informasi yang domain kasusnya pada organ mata. Adapun jawaban kasusnya adalah miopia atau rabun jauh. Untuk kasus ini jumlah halamannya hanya 3 page. Berdasarkan informasi yang diberikan ChatGPT mampu memberikan akurasi 100% untuk ketepatan diagnosa dan diagnosa banding. Sedangkan variasi diagnosanya adalah 2x lebih banyak dibandingkan diagnosa banding faktual dari kasus PBL yang sudah diujikan ke siswa.

Kasus PBL 6

Pada kasus PBL 6 domain kasusnya berada pada indra telinga. ChatGPT diminta memberikan respon diagnosa dan diagnosa banding dengan key trigger *felt pain in her right ear since yesterday*. Adapun respon yang dihasilkan adalah 100% akurasi diagnosa pada page 1-4. Meskipun demikian terjadi perluasan diagnosa banding pada Page 4 meningkat menjadi 4 DD. Sedangkan variasi diagnosanya adalah 1,5x lebih banyak dibandingkan diagnosa banding faktual dari kasus PBL yang sudah diujikan ke siswa.

Berdasarkan akumulasi data terlihat ChatGPT memberikan akurasi respon diagnosa sekitar 92% untuk 6 kasus PBL dengan akurasi DD sekitar 88%. Angka ini diperkuat dengan peningkatan jumlah

variasi DD sekitar 2x lebih banyak dibandingkan diagnosa banding faktual dari kasus PBL yang sudah diujikan ke siswa. Temuan ini sejalan dengan penelitian performa teknologi large language model yang signifikan menyelesaikan soal teks dengan konteks klinis (Peng et al., 2023). Dari sini kita dapat melihat bahwa ChatGPT dapat dijadikan sumber informasi dalam menambah wawasan tentunya dengan tetap mengedepankan validasi informasi. Penggunaan teknologi dalam ranah PBL ini kalau dilihat sejalan dengan penelitian oleh Rintaro yang dapat meningkatkan skill siswa (Imafuku, 2012) tentunya dalam menemukan informasi.

Melihat dari kemampuan ChatGPT dalam memberikan respon terhadap kasus PBL memberikan fakta kepada kita bahwa sistem *machine learning* dan *training data* telah berjalan dan semakin disempurnakan oleh pengembang. Apalagi kemampuan menghasilkan akurasi diagnosa dan diagnosa banding memperlihatkan keluasan informasi kesehatan yang dimiliki Gambar 2. Karena itu kemampuan menghasilkan teks pada ChatGPT memang menjadi salah satu tantangan saat ini (NewYork Times, 2023). Apalagi terkait bias informasi dan etik menjadikan aplikasi ini sebagai objek riset penting dalam dunia pendidikan tinggi terutama dalam bidang kedokteran. Banyak aspek yang masih harus didalami dalam penggunaan aplikasi ini (Tsang, 2023). Namun, di sisi lain teknologi OpenAi pada ChatGPT memang sangat maju. Dalam berbagai testing salah satunya ujian progres test pada Fakultas Kedokteran di German berhasil dilalui dengan baik (Friederichs et al., 2023).

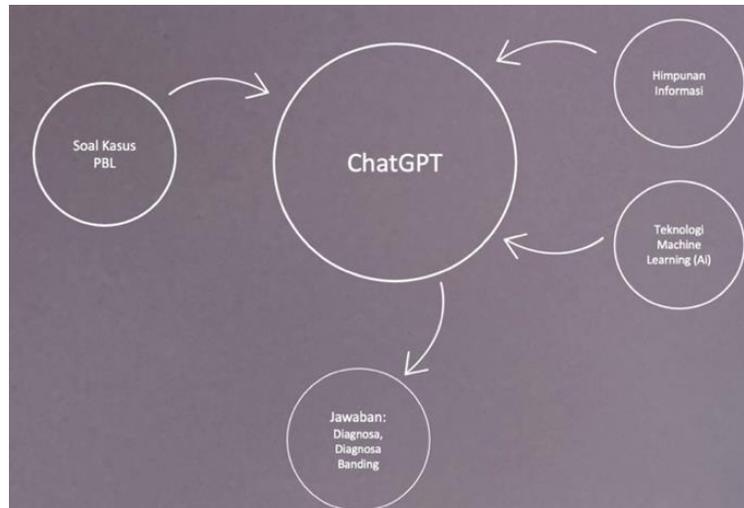
Korelasi dan Temuan Penting

Berdasarkan ulasan dari penelitian ini kita melihat bahwa teknologi ChatGPT memang mampu memberikan potensi solusi terhadap. Hasil penelitian ini sesuai dengan kajian oleh Malik Sallam, yang menguraikan tentang peningkatan personalisasi dalam pembelajaran dan berfikir kritis dalam *Problem Based Learning* (Sallam, 2023). Meskipun demikian, kehandalan penyertaan dan akurasi sumber referensi memang masih menjadi perhatian utama dalam penggunaan Ai ChatGPT (Chen, 2023). Untuk itulah manusia dibutuhkan pada bagian akhir dalam sintesis tulisan. Manusia harus berperan sebagai *human control*. Temuan penting dalam riset ini adalah hasil pengujian performa ChatGPT pada area *medical education* khususnya pada *Problem Based Learning* dengan fokus pada pengukuran presisi diagnosa dan diagnosa banding. Pengujian performa pada area *medical education* telah pernah dilakukan namun hanya pengujian soal ujian tingkat akhir dalam bentuk USMLE (Kung et al., 2023). Sehingga penelitian ini memberikan kontribusi perspektif yang lebih luas pada *medical education*. Terutama untuk menyelesaikan permasalahan medis yang memiliki kompleksitas yang tinggi (Sallam et al., 2023).

4. KESIMPULAN

Hasil analisis akurasi ChatGPT pada kasus PBL diperoleh akurasi respon diagnosa sekitar 92% untuk 6 kasus PBL dengan akurasi DD sekitar 88%. Angka ini diperkuat dengan peningkatan jumlah variasi DD sekitar 2x lebih banyak dibandingkan diagnosa banding faktual dari kasus PBL yang sudah diujikan ke siswa. Dari sini kita dapat melihat bahwa ChatGPT dapat dijadikan sumber informasi dalam menambah wawasan namun dengan tetap mengedepankan validasi informasi. ChatGPT juga dapat mendukung belajar mandiri siswa karena dapat menjadi mesin pengetahuan tentunya harus dibarengi dengan kemampuan validitas informasi.

Penelitian ini masih memiliki peluang untuk dikembangkan dan disempurnakan. Oleh karena itu, dibutuhkan eksplorasi lanjut berikut ini penjabaran riset yang dapat dilakukan. Berikut ini beberapa saran penelitian yang dapat dilakukan untuk melanjutkan penelitian ini:



Gambar 1. Skema proses kerja dan rasionalitas kemampuan ChatGPT dalam menjawab soal-soal kasus PBL.

1. **Perluasan Kasus PBL:** Dalam penelitian ini, ChatGPT telah diuji dengan 6 kasus PBL. Penelitian berikutnya dapat memperluas jumlah dan variasi kasus PBL yang digunakan untuk mengetahui sejauh mana keakuratan dan konsistensi ChatGPT dalam berbagai konteks dan tingkat kesulitan.
2. **Analisis Komparatif dengan Model AI Lainnya:** Penelitian ini hanya mengevaluasi performa ChatGPT. Penelitian berikutnya dapat membandingkan kinerja ChatGPT dengan model AI lainnya untuk mengetahui mana yang memberikan respon diagnosa dan DD terbaik.

TERIMA KASIH

1. Fakultas Kedokteran Universitas Surabaya
2. Lembaga Penelitian dan Pengabdian kepada Masyarakat, Universitas Surabaya

DAFTAR PUSTAKA

- Anggraeni, D. M., Prahani, B. K., Suprpto, N., Shofiyah, N., & Jatmiko, B. (2023). Systematic review of problem based learning research in fostering critical thinking skills. *Thinking Skills and Creativity*, 49, 101334. <https://doi.org/10.1016/J.TSC.2023.101334>
- Burgun, A., Darmoni, S., Duff, F. Le, & Wéber, J. (2006). Problem-based learning in medical informatics for undergraduate medical students: An experiment in two medical schools. *International Journal of Medical Informatics*, 75(5), 396–402. <https://doi.org/10.1016/J.IJMEDINF.2005.07.014>
- Cain, J., Malcom, D. R., & Aungst, T. D. (2023). The Role of Artificial Intelligence in the Future of Pharmacy Education. *American Journal of Pharmaceutical Education*, 100135. <https://doi.org/10.1016/j.ajpe.2023.100135>
- Friederichs, H., Friederichs, W. J., & März, M. (2023). ChatGPT in medical school: how successful is AI in progress testing? *https://doi.org/10.1080/10872981.2023.2220920*, 28(1). <https://doi.org/10.1080/10872981.2023.2220920>

- Grünebaum, A., Chervenak, J., Pollet, S. L., Katz, A., & Chervenak, F. A. (2023). The exciting potential for ChatGPT in obstetrics and gynecology. *American Journal of Obstetrics and Gynecology*, 228(6), 696–705. <https://doi.org/10.1016/j.ajog.2023.03.009>
- Howard S. Barrows, & Robyn M. Tamblyn. (1980). *Problem-Based Learning: An Approach to Medical Education - Howard S. Barrows, MD, Robyn M. Tamblyn, BScN - Google Buku*. Springer Publishing Company. https://books.google.co.id/books/about/Problem_Based_Learning.html?id=9u-5DJuQq2UC&redir_esc=y
- Imafuku, R. (2012). Japanese first-Year PBL students' learning processes: A classroom discourse analysis. *Problem-Based Learning in Clinical Education: The Next Generation*, 153–170. https://doi.org/10.1007/978-94-007-2515-7_10/COVER
- NewYork Times. (2023). *Don't Ban ChatGPT in Schools. Teach With It. - The New York Times*. Newyork Times. <https://www.nytimes.com/2023/01/12/technology/chatgpt-schools-teachers.html>
- O'Connor, S., & ChatGPT. (2023). Open artificial intelligence platforms in nursing education: Tools for academic progress or abuse? *Nurse Education in Practice*, 66, 103537. <https://doi.org/10.1016/J.NEPR.2022.103537>
- OPEN AI. (2023). *New and improved content moderation tooling*. Website. <https://openai.com/blog/new-and-improved-content-moderation-tooling>
- Peng, C., Yang, X., Chen, A., Smith, K. E., Pournajatian, N., Costa, A. B., Martin, C., Flores, M. G., Zhang, Y., Magoc, T., Lipori, G., Mitchell, D. A., Ospina, N. S., Ahmed, M. M., Hogan, W. R., Shenkman, E. A., Guo, Y., Bian, J., & Wu, Y. (2023). *A Study of Generative Large Language Model for Medical Research and Healthcare*.
- Servant-Miklos, V. F. C. (2019). A Revolution in its Own Right: How Maastricht University Reinvented Problem-Based Learning. *Health Professions Education*, 5(4), 283–293. <https://doi.org/10.1016/j.hpe.2018.12.005>
- Tsang, R. (2023). Practical Applications of ChatGPT in Undergraduate Medical Education. <https://doi.org/10.1177/23821205231178449>, 10, 23821205231178450. <https://doi.org/10.1177/23821205231178449>
- Yilmaz, R., & Karaoglan Yilmaz, F. G. (2023). The effect of generative artificial intelligence (AI)-based tool use on students' computational thinking skills, programming self-efficacy and motivation. *Computers and Education: Artificial Intelligence*, 4, 100147. <https://doi.org/10.1016/j.caeai.2023.100147>