https://doi.org/10.3348/kjr.2024.0788

Supplement

Supplementary Table 1. List of 159 studies analyzed

No.	Author	Year	Title	Journal
1	V. Yaneva, P. Baldwin, D. P. Jurich, K. Swygert and B. E. Clauser	2024	Examining ChatGPT Performance on USMLE Sample Items and Implications for Assessment	Acad Med
2	M. C. Laupichler, J. F. Rother, I. C. Grunwald Kadow, S. Ahmadi and T. Raupach	2024	Large Language Models in Medical Education: Comparing ChatGPT- to Human-Generated Exam Questions	Acad Med
3	K. Suchman, S. Garg and A. J. Trindade	2023	Chat Generative Pretrained Transformer Fails the Multiple-Choice American College of Gastroenterology Self-Assessment Test	Am J Gastroenterol
4	J. B. Henson, J. R. Glissen Brown, J. P. Lee, A. Patel and D. A. Leiman	2023	Evaluation of the Potential Utility of an Artificial Intelligence Chatbot in Gastroesophageal Reflux Disease Management	Am J Gastroenterol
5	A. Levartovsky, S. Ben-Horin, U. Kopylov, E. Klang and Y. Barash	2023	Towards AI-Augmented Clinical Decision-Making: An Examination of ChatGPT's Utility in Acute Ulcerative Colitis Presentations	Am J Gastroenterol
6	L. Saba, C. L. Fu, J. Khouri, B. Faiman, F. Anwer and C. P. Chaulagain	2024	Evaluating ChatGPT as an educational resource for patients with multiple myeloma: A preliminary investigation	Am J Hematol
7	T. Hirosawa, K. Mizuta, Y. Harada and T. Shimizu	2023	Comparative Evaluation of Diagnostic Accuracy Between Google Bard and Physicians	Am J Med
8	S. W. Li, M. W. Kemp, S. J. S. Logan, P. S. Dimri, N. Singh, C. N. Z. Mattar, P. Dashraath, H. Ramlal, A. P. Mahyuddin, S. Kanayan, S. W. D. Carter, S. P. T. Thain, E. L. Fee, S. E. Illanes and M. A. Choolani	2023	ChatGPT outscored human candidates in a virtual objective structured clinical examination in obstetrics and gynecology	Am J Obstet Gynecol
9	E. T. Pan and M. Florian-Rodriguez	2024	Human vs machine: identifying ChatGPT-generated abstracts in Gynecology and Urogynecology	Am J Obstet Gynecol
10	L. Z. Cai, A. Shaheen, A. Jin, R. Fukui, J. S. Yi, N. Yannuzzi and C. Alabiad	2023	Performance of Generative Large Language Models on Ophthalmology Board-Style Questions	Am J Ophthalmol
11	Q. Dihan, M. Z. Chauhan, T. K. Eleiwa, A. K. Hassan, A. B. Sallam, A. S. Khouri, T. C. Chang and A. M. Elhusseiny	2024	Using Large Language Models to Generate Educational Materials on Childhood Glaucoma	Am J Ophthalmol
12	X. Huang, H. Raja, Y. Madadi, M. Delsoz, A. Poursoroush, M. Y. Kahook and S. Yousefi	2024	Predicting glaucoma before onset using a large language model chatbot	Am J Ophthalmol
13	Y. Barak-Corren, R. Wolf, R. Rozenblum, J. K. Creedon, S. C. Lipsett, T. W. Lyons, K. A. Michelson, K. A. Miller, D. J. Shapiro, B. Y. Reis and A. M. Fine	2024	Harnessing the Power of Generative AI for Clinical Summaries: Perspectives From Emergency Physicians	Ann Emerg Med

https://doi.org/10.3348/kjr.2024.0788

14	J. Hake, M. Crowley, A. Coy, D. Shanks, A. Eoff, K. Kirmer-Voss, G. Dhanda and D. J. Parente	2024	Quality, Accuracy, and Bias in ChatGPT-Based Summarization of Medical Abstracts	Ann Fam Med
15	T. H. McCoy and R. H. Perlis	2024	Dimensional measures of psychopathology in children and adolescents using large language models	Biol Psychiatry
16	S. Koga, N. B. Martin and D. W. Dickson	2024	Evaluating the performance of large language models: ChatGPT and Google Bard in generating differential diagnoses in clinicopathological conferences of neurodegenerative disorders	Brain Pathol
17	N. Pugliese, V. Wai-Sun Wong, J. M. Schattenberg, M. Romero-Gomez, G. Sebastiani and A. Aghemo	2024	Accuracy, Reliability, and Comprehensibility of ChatGPT-Generated Medical Responses for Patients With Nonalcoholic Fatty Liver Disease	Clin Gastroenterol Hepatol
18	A. Kerbage, J. Kassab, J. El Dahdah, C. A. Burke, J. P. Achkar and C. Rouphael	2024	Accuracy of ChatGPT in Common Gastrointestinal Diseases: Impact for Patients and Providers	Clin Gastroenterol Hepatol
19	P. W. Chang, M. M. Amini, R. O. Davis, D. D. Nguyen, J. L. Dodge, H. Lee, S. Sheibani, J. Phan, J. L. Buxbaum and A. B. Sahakian	2024	ChatGPT4 Outperforms Endoscopists for Determination of Postcolonoscopy Rescreening and Surveillance Recommendations	Clin Gastroenterol Hepatol
20	J. Miao, C. Thongprayoon, O. A. Garcia Valencia, P. Krisanapan, M. S. Sheikh, P. W. Davis, P. Mekraksakit, M. G. Suarez, I. M. Craici and W. Cheungpasitporn	2023	Performance of ChatGPT on Nephrology Test Questions	Clin J Am Soc Nephrol
21	W. I. Wei, C. L. K. Leung, A. Tang, E. B. McNeil, S. Y. S. Wong and K. O. Kwok	2024	Extracting symptoms from free-text responses using ChatGPT among COVID-19 cases in Hong Kong	Clin Microbiol Infect
22	Y. H. Yeo, J. S. Samaan, W. H. Ng, P. S. Ting, H. Trivedi, A. Vipani, W. Ayoub, J. D. Yang, O. Liran, B. Spiegel and A. Kuo	2023	Assessing the performance of ChatGPT in answering questions regarding cirrhosis and hepatocellular carcinoma	Clin Mol Hepatol
23	J. M. M. Rogasch, H. V. Jochens, G. Metzger, C. Wetz, J. Kaufmann, C. Furth, H. Amthauer and I. Schatka	2024	Keeping Up With ChatGPT: Evaluating Its Recognition and Interpretation of Nuclear Medicine Images	Clin Nucl Med
24	Z. C. Lum	2023	Can Artificial Intelligence Pass the American Board of Orthopaedic Surgery Examination? Orthopaedic Residents Versus ChatGPT	Clin Orthop Relat Res
25	L. Ying, Z. Liu, H. Fang, R. Kusko, L. Wu, S. Harris and W. Tong	2024	Text summarization with ChatGPT for drug labeling documents	Drug Discov Today
26	Z. W. Lim, K. Pushpanathan, S. M. E. Yew, Y. Lai, C. H. Sun, J. S. H. Lam, D. Z. Chen, J. H. L. Goh, M. C. J. Tan, B. Sheng, C. Y. Cheng, V. T. C. Koh and Y. C. Tham	2023	Benchmarking large language models' performances for myopia care: a comparative analysis of ChatGPT-3.5, ChatGPT-4.0, and Google Bard	EBioMedicine
27	L. Van Bulck and P. Moons	2024	What if your patient switches from Dr. Google to Dr. ChatGPT? A vignette-based survey of the trustworthiness, value, and danger of ChatGPT-generated responses to health questions	Eur J Cardiovasc Nurs

https://doi.org/10.3348/kjr.2024.0788

28	P. Moons and L. Van Bulck	2024	Using ChatGPT and Google Bard to improve the readability of written patient information: a proof of concept	Eur J Cardiovasc Nurs
29	D. Musheyev, A. Pan, S. Loeb and A. E. Kabarriti	2024	How Well Do Artificial Intelligence Chatbots Respond to the Top Search Queries About Urological Malignancies?	Eur Urol
30	C. E. Hermann, J. M. Patel, L. Boyd, W. B. Growdon, E. Aviki and M. Stasenko	2023	Let's chat about cervical cancer: Assessing the accuracy of ChatGPT responses to cervical cancer questions	Gynecol Oncol
31	J. M. Patel, C. E. Hermann, W. B. Growdon, E. Aviki and M. Stasenko	2024	ChatGPT accurately performs genetic counseling for gynecologic cancers	Gynecol Oncol
32	A. Suárez, V. Díaz-Flores García, J. Algar, M. Gómez Sánchez, M. Llorente de Pedro and Y. Freire	2024	Unveiling the ChatGPT phenomenon: Evaluating the consistency and accuracy of endodontic question answers	Int Endod J
33	Y. Yoshiyasu, F. Wu, A. K. Dhanda, D. Gorelik, M. Takashima and O. G. Ahmed	2023	GPT-4 accuracy and completeness against International Consensus Statement on Allergy and Rhinology: Rhinosinusitis	Int Forum Allergy Rhinol
34	A. D. Workman, V. K. Rathi, D. K. Lerner, J. N. Palmer, N. D. Adappa and N. A. Cohen	2024	Utility of a LangChain and OpenAI GPT-powered chatbot based on the international consensus statement on allergy and rhinology: Rhinosinusitis	Int Forum Allergy Rhinol
35	G. S. Hill, J. L. Fischer, N. L. Watson, C. A. Riley and A. M. Tolisano	2024	Assessing the quality of artificial intelligence-generated patient counseling for rhinosinusitis	Int Forum Allergy Rhinol
36	G. Levin, R. Pareja, D. Viveros-Carreño, E. Sanchez Diaz, E. M. Yates, B. Zand and P. T. Ramirez	2024	Association of reviewer experience with discriminating human-written versus ChatGPT-written abstracts	Int J Gynecol Cancer
37	A. P. Mika, J. R. Martin, S. M. Engstrom, G. G. Polkowski and J. M. Wilson	2023	Assessing ChatGPT Responses to Common Patient Questions Regarding Total Hip Arthroplasty	J Bone Joint Surg Am
38	V. S. Kasthuri, J. Glueck, H. Pham, M. Daher, M. Balmaceno-Criss, C. L. McDonald, B. G. Diebo and A. H. Daniels	2024	Assessing the Accuracy and Reliability of AI-Generated Responses to Patient Questions Regarding Spine Surgery	J Bone Joint Surg Am
39	M. C. Williams and J. Shambrook	2023	How will artificial intelligence transform cardiovascular computed tomography? A conversation with an AI model	J Cardiovasc Comput Tomogr
40	E. Turan, A. E. Baydemir, F. G. Özcan and A. S. Şahin	2024	Evaluating the accuracy of ChatGPT-4 in predicting ASA scores: A prospective multicentric study ChatGPT-4 in ASA score prediction	J Clin Anesth
41	M. Sciberras, Y. Farrugia, H. Gordon, F. Furfaro, M. Allocca, J. Torres, N. Arebi, G. Fiorino, M. Iacucci, B. Verstockt, F. Magro, K. Katsanos, J. Busuttil, K. De Giovanni, V. A. Fenech, S. Chetcuti Zammit and P. Ellul	2024	Accuracy of Information given by ChatGPT for patients with Inflammatory Bowel Disease in relation to ECCO Guidelines	J Crohns Colitis

https://doi.org/10.3348/kjr.2024.0788

42	H. Burnette, A. Pabani, M. S. von Itzstein, B. Switzer, R. Fan, F. Ye, I. Puzanov, J. Naidoo, P. A. Ascierto, D. E. Gerber, M. S. Ernstoff and D. B. Johnson	2024	Use of artificial intelligence chatbots in clinical management of immune-related adverse events	J Immunother Cancer
43	J. Chen, A. Cadiente, L. J. Kasselman and B. Pilkington	2024	Assessing the performance of ChatGPT in bioethics: a large language model's moral compass in medicine	J Med Ethics
44	K. Sridharan and G. Sivaramakrishnan	2024	Leveraging artificial intelligence to detect ethical concerns in medical research: a case study	J Med Ethics
45	M. Májovský, M. Černý, M. Kasal, M. Komarc and D. Netuka	2023	Artificial Intelligence Can Generate Fraudulent but Authentic-Looking Scientific Medical Articles: Pandora's Box Has Been Opened	J Med Internet Res
46	H. L. Walker, S. Ghani, C. Kuemmerli, C. A. Nebiker, B. P. Müller, D. A. Raptis and S. M. Staubli	2023	Reliability of Medical Information Provided by ChatGPT: Assessment Against Clinical Guidelines and Patient Information Quality Instrument	J Med Internet Res
47	A. Rao, M. Pang, J. Kim, M. Kamineni, W. Lie, A. K. Prasad, A. Landman, K. Dreyer and M. D. Succi	2023	Assessing the Utility of ChatGPT Throughout the Entire Clinical Workflow: Development and Usability Study	J Med Internet Res
48	V. Hristidis, N. Ruggiano, E. L. Brown, S. R. R. Ganta and S. Stewart	2023	ChatGPT vs Google for Queries Related to Dementia and Other Cognitive Decline: Comparison of Results	J Med Internet Res
49	T. I. Wilhelm, J. Roos and R. Kaczmarczyk	2023	Large Language Models for Therapy Recommendations Across 3 Clinical Specialties: Comparative Study	J Med Internet Res
50	G. Wang, K. Gao, Q. Liu, Y. Wu, K. Zhang, W. Zhou and C. Guo	2023	Potential and Limitations of ChatGPT 3.5 and 4.0 as a Source of COVID-19 Information: Comprehensive Comparative Analysis of Generative and Authoritative Information	J Med Internet Res
51	S. Ziegelmayer, A. W. Marka, N. Lenhart, N. Nehls, S. Reischl, F. Harder, A. Sauter, M. Makowski, M. Graf and J. Gawlitza	2023	Evaluation of GPT-4's Chest X-Ray Impression Generation: A Reader Study on Performance and Perception	J Med Internet Res
52	S. L. Cheng, S. J. Tsai, Y. M. Bai, C. H. Ko, C. W. Hsu, F. C. Yang, C. K. Tsai, Y. K. Tu, S. N. Yang, P. T. Tseng, T. W. Hsu, C. S. Liang and K. P. Su	2023	Comparisons of Quality, Correctness, and Similarity Between ChatGPT-Generated and Human-Written Abstracts for Basic Research: Cross-Sectional Study	J Med Internet Res
53	M. S. Deiner, N. A. Deiner, V. Hristidis, S. D. McLeod, T. Doan, T. M. Lietman and T. C. Porco	2024	Use of Large Language Models to Assess the Likelihood of Epidemics From the Content of Tweets: Infodemiology Study	J Med Internet Res
54	Z. Xue, Y. Zhang, W. Gan, H. Wang, G. She and X. Zheng	2024	Quality and Dependability of ChatGPT and DingXiangYuan Forums for Remote Orthopedic Consultations: Comparative Analysis	J Med Internet Res
55	X. Liu, J. Wu, A. Shao, W. Shen, P. Ye, Y. Wang, J. Ye, K. Jin and J. Yang	2024	Uncovering Language Disparity of ChatGPT on Retinal Vascular Disease Classification: Cross-Sectional Study	J Med Internet Res

https://doi.org/10.3348/kjr.2024.0788

56	A. Herrmann-Werner, T. Festl-Wietek, F. Holderried, L. Herschbach, J. Griewatz, K. Masters, S. Zipfel and M. Mahling	2024	Assessing ChatGPT's Mastery of Bloom's Taxonomy Using Psychosomatic Medicine Exam Questions: Mixed-Methods Study	J Med Internet Res
57	L. Masanneck, L. Schmidt, A. Seifert, T. Kölsche, N. Huntemann, R. Jansen, M. Mehsin, M. Bernhard, S. G. Meuth, L. Böhm and M. Pawlitzki	2024	Triage Performance Across Large Language Models, ChatGPT, and Untrained Doctors in Emergency Medicine: Comparative Study	J Med Internet Res
58	W. H. K. Chiu, W. S. K. Ko, W. C. S. Cho, S. Y. J. Hui, W. C. L. Chan and M. D. Kuo	2024	Evaluating the Diagnostic Performance of Large Language Models on Complex Multimodal Medical Cases	J Med Internet Res
59	A. Kernberg, J. A. Gold and V. Mohan	2024	Using ChatGPT-4 to Create Structured Medical Notes From Audio Recordings of Physician-Patient Encounters: Comparative Study	J Med Internet Res
60	L. Wang, Y. Ma, W. Bi, H. Lv and Y. Li	2024	An Entity Extraction Pipeline for Medical Text Records Using Large Language Models: Analytical Study	J Med Internet Res
61	L. Zhu, W. Mou, K. Wu, Y. Lai, A. Lin, T. Yang, J. Zhang and P. Luo	2024	Multimodal ChatGPT-4V for ECG Interpretation: Promise and Limitations	J Med Internet Res
62	W. He, W. Zhang, Y. Jin, Q. Zhou, H. Zhang and Q. Xia	2024	Physician Versus Large Language Model Chatbot Responses to Web-Based Questions From Autistic Patients in Chinese: Cross-Sectional Comparative Analysis	J Med Internet Res
63	E. Xue, D. Bracken-Clarke, G. M. Iannantuono, H. Choo-Wosoba, J. L. Gulley and C. S. Floudas	2024	Utility of Large Language Models for Health Care Professionals and Patients in Navigating Hematopoietic Stem Cell Transplantation: Comparison of the Performance of ChatGPT-3.5, ChatGPT-4, and Bard	J Med Internet Res
64	F. Busch, T. Han, M. R. Makowski, D. Truhn, K. K. Bressem and L. Adams	2024	Integrating Text and Image Analysis: Exploring GPT-4V's Capabilities in Advanced Radiological Applications Across Subspecialties	J Med Internet Res
65	V. Mishra, A. Sarraju, N. M. Kalwani and J. P. Dexter	2024	Evaluation of Prompts to Simplify Cardiovascular Disease Information Generated Using a Large Language Model: Cross-Sectional Study	J Med Internet Res
66	X. Lv, X. Zhang, Y. Li, X. Ding, H. Lai and J. Shi	2024	Leveraging Large Language Models for Improved Patient Access and Self-Management: Assessor-Blinded Comparison Between Expert- and AI-Generated Content	J Med Internet Res
67	L. Li, P. Li, K. Wang, L. Zhang, H. Zhao and H. Ji	2024	Benchmarking State-of-the-Art Large Language Models for Migraine Patient Education: A Comparison of Performances on the Responses to Common Queries	J Med Internet Res
68	D. Yoon, C. Han, D. W. Kim, S. Kim, S. Bae, J. A. Ryu and Y. Choi	2024	Redefining Health Care Data Interoperability: Empirical Exploration of Large Language Models in Information Exchange	J Med Internet Res

https://doi.org/10.3348/kjr.2024.0788

69	Z. He, B. Bhasuran, Q. Jin, S. Tian, K. Hanna, C. Shavor, L. G. Arguello, P. Murray and Z. Lu	2024	Quality of Answers of Generative Large Language Models Versus Peer Users for Interpreting Laboratory Test Results for Lay Patients: Evaluation Study	J Med Internet Res
70	T. C. Chen, M. W. Couldwell, J. Singer, A. Singer, L. Koduri, E. Kaminski, K. Nguyen, E. Multala, A. S. Dumont and A. Wang	2024	Assessing the clinical reasoning of ChatGPT for mechanical thrombectomy in patients with stroke	J Neurointerv Surg
71	T. Pedro, J. M. Sousa, L. Fonseca, M. G. Gama, G. Moreira, M. Pintalhão, P. C. Chaves, A. Aires, G. Alves, L. Augusto, L. Pinheiro Albuquerque, P. Castro and M. L. Silva	2024	Exploring the use of ChatGPT in predicting anterior circulation stroke functional outcomes after mechanical thrombectomy: a pilot study	J Neurointerv Surg
72	D. Truhn, C. M. Loeffler, G. Müller-Franzes, S. Nebelung, K. J. Hewitt, S. Brandner, K. K. Bressem, S. Foersch and J. N. Kather	2024	Extracting structured information from unstructured histopathology reports using generative pre-trained transformer 4 (GPT-4)	J Pathol
73	R. Davis, M. Eppler, O. Ayo-Ajibola, J. C. Loh-Doyle, J. Nabhani, M. Samplaski, I. Gill and G. E. Cacciamani	2023	Evaluating the Effectiveness of Artificial Intelligence-powered Large Language Models Application in Disseminating Appropriate and Readable Health Information in Urology	J Urol
74	Z. Kanjee, B. Crowe and A. Rodman	2023	Accuracy of a Generative Artificial Intelligence Model in a Complex Diagnostic Challenge	JAMA
75	D. von Wedel, R. A. Schmitt, M. Thiele, R. Leuner, D. Shay, S. Redaelli and M. S. Schaefer	2024	Affiliation Bias in Peer Review of Abstracts by a Large Language Model	JAMA
76	T. Han, L. C. Adams, K. K. Bressem, F. Busch, S. Nebelung and D. Truhn	2024	Comparative Analysis of Multimodal Large Language Model Performance on Clinical Vignette Questions	JAMA
77	J. W. Ayers, A. Poliak, M. Dredze, E. C. Leas, Z. Zhu, J. B. Kelley, D. J. Faix, A. M. Goodman, C. A. Longhurst, M. Hogarth and D. M. Smith	2023	Comparing Physician and Artificial Intelligence Chatbot Responses to Patient Questions Posted to a Public Social Media Forum	JAMA Intern Med
78	A. Nayak, M. S. Alkaitis, K. Nayak, M. Nikolov, K. P. Weinfurt and K. Schulman	2023	Comparison of History of Present Illness Summaries Generated by a Chatbot and Senior Internal Medicine Residents	JAMA Intern Med
79	E. Strong, A. DiGiammarino, Y. Weng, A. Kumar, P. Hosamani, J. Hom and J. H. Chen	2023	Chatbot vs Medical Student Performance on Free-Response Clinical Reasoning Examinations	JAMA Intern Med
80	S. Cabral, D. Restrepo, Z. Kanjee, P. Wilson, B. Crowe, R. E. Abdulnour and A. Rodman	2024	Clinical Reasoning of a Generative Artificial Intelligence Model Compared With Physicians	JAMA Intern Med
81	J. W. Ayers, Z. Zhu, A. Poliak, E. C. Leas, M. Dredze, M. Hogarth and D. M. Smith	2023	Evaluating Artificial Intelligence Responses to Public Health Questions	JAMA Netw Open
82	Y. F. Shea, C. M. Y. Lee, W. C. T. Ip, D. W. A. Luk and S. S. W. Wong	2023	Use of GPT-4 to Analyze Medical Records of Patients With Extensive Investigations and Delayed Diagnosis	JAMA Netw Open
83	A. Chen and D. O. Chen	2023	Accuracy of Chatbots in Citing Journal Articles	JAMA Netw Open

https://doi.org/10.3348/kjr.2024.0788

84	I. A. Bernstein, Y. V. Zhang, D. Govil, I. Majid, R. T. Chang, Y. Sun, A. Shue, J. C. Chou, E. Schehlein, K. L. Christopher, S. L. Groth, C. Ludwig and S. Y. Wang	2023	Comparison of Ophthalmologist and Large Language Model Chatbot Responses to Online Patient Eye Care Questions	JAMA Netw Open
85	R. S. Goodman, J. R. Patrinely, C. A. Stone, Jr., E. Zimmerman, R. R. Donald, S. S. Chang, S. T. Berkowitz, A. P. Finn, E. Jahangir, E. A. Scoville, T. S. Reese, D. L. Friedman, J. A. Bastarache, Y. F. van der Heijden, J. J. Wright, F. Ye, N. Carter, M. R. Alexander, J. H. Choe, C. A. Chastain, J. A. Zic, S. N. Horst, I. Turker, R. Agarwal, E. Osmundson, K. Idrees, C. M. Kiernan, C. Padmanabhan, C. E. Bailey, C. E. Schlegel, L. B. Chambless, M. K. Gibson, T. J. Osterman, L. E. Wheless and D. B. Johnson	2023	Accuracy and Reliability of Chatbot Responses to Physician Questions	JAMA Netw Open
86	H. Decker, K. Trang, J. Ramirez, A. Colley, L. Pierce, M. Coleman, T. Bongiovanni, G. B. Melton and E. Wick	2023	Large Language Model-Based Chatbot vs Surgeon-Generated Informed Consent Documentation for Common Procedures	JAMA Netw Open
87	J. Kim, Z. R. Cai, M. L. Chen, J. F. Simard and E. Linos	2023	Assessing Biases in Medical Decisions via Clinician and AI Chatbot Responses to Patient Vignettes	JAMA Netw Open
88	M. Benary, X. D. Wang, M. Schmidt, D. Soll, G. Hilfenhaus, M. Nassir, C. Sigler, M. Knödler, U. Keller, D. Beule, U. Keilholz, U. Leser and D. T. Rieke	2023	Leveraging Large Language Models for Decision Support in Personalized Oncology	JAMA Netw Open
89	M. C. Schubert, W. Wick and V. Venkataramani	2023	Performance of Large Language Models on a Neurology Board-Style Examination	JAMA Netw Open
90	J. Zaretsky, J. M. Kim, S. Baskharoun, Y. Zhao, J. Austrian, Y. Aphinyanaphongs, R. Gupta, S. B. Blecker and J. Feldman	2024	Generative Artificial Intelligence to Transform Inpatient Discharge Summaries to Patient-Friendly Language and Format	JAMA Netw Open
91	H. Lai, L. Ge, M. Sun, B. Pan, J. Huang, L. Hou, Q. Yang, J. Liu, J. Liu, Z. Ye, D. Xia, W. Zhao, X. Wang, M. Liu, J. R. Talukdar, J. Tian, K. Yang and J. Estill	2024	Assessing the Risk of Bias in Randomized Clinical Trials With Large Language Models	JAMA Netw Open
92	E. Steimetz, J. Minkowitz, E. C. Gabutan, J. Ngichabe, H. Attia, M. Hershkop, F. Ozay, M. G. Hanna and R. Gupta	2024	Use of Artificial Intelligence Chatbots in Interpretation of Pathology Reports	JAMA Netw Open
93	G. Lorenzoni, D. Gregori, S. Bressan, H. Ocagli, D. Azzolina, L. Da Dalt and P. Berchialla	2024	Use of a Large Language Model to Identify and Classify Injuries With Free-Text Emergency Department Data	JAMA Netw Open
94	J. B. Longwell, I. Hirsch, F. Binder, G. A. Gonzalez Conchas, D. Mau, R. Jang, R. G. Krishnan and R. C. Grant	2024	Performance of Large Language Models on Medical Oncology Examination Questions	JAMA Netw Open

https://doi.org/10.3348/kjr.2024.0788

95	P. Garcia, S. P. Ma, S. Shah, M. Smith, Y. Jeong, A. Devon-Sand, M. Tai-Seale, K. Takazawa, D. Clutter, K. Vogt, C. Lugtu, M. Rojo, S. Lin, T. Shanafelt, M. A. Pfeffer and C. Sharp	2024	Artificial Intelligence-Generated Draft Replies to Patient Inbox Messages	JAMA Netw Open
96	A. Yalamanchili, B. Sengupta, J. Song, S. Lim, T. O. Thomas, B. B. Mittal, M. E. Abazeed and P. T. Teo	2024	Quality of Large Language Model Responses to Radiation Oncology Patient Care Questions	JAMA Netw Open
97	C. Y. K. Williams, T. Zack, B. Y. Miao, M. Sushil, M. Wang, A. E. Kornblith and A. J. Butte	2024	Use of a Large Language Model to Assess Clinical Acuity of Adults in the Emergency Department	JAMA Netw Open
98	S. Habib, H. Butt, S. R. Goldenholz, C. Y. Chang and D. M. Goldenholz	2024	Large Language Model Performance on Practice Epilepsy Board Examinations	JAMA Neurol
99	A. Pan, D. Musheyev, D. Bockelman, S. Loeb and A. E. Kabarriti	2023	Assessment of Artificial Intelligence Chatbot Responses to Top Searched Queries About Cancer	JAMA Oncol
100	D. Chen, R. Parsa, A. Hope, B. Hannon, E. Mak, L. Eng, F. F. Liu, N. Fallah-Rad, A. M. Heesters and S. Raman	2024	Physician and Artificial Intelligence Chatbot Responses to Cancer Questions From Social Media	JAMA Oncol
101	A. Mihalache, M. M. Popovic and R. H. Muni	2023	Performance of an Artificial Intelligence Chatbot in Ophthalmic Knowledge Assessment	JAMA Ophthalmol
102	A. Mihalache, R. S. Huang, M. M. Popovic and R. H. Muni	2023	Performance of an Upgraded Artificial Intelligence Chatbot for Ophthalmic Knowledge Assessment	JAMA Ophthalmol
103	H. U. Hua, A. H. Kaakour, A. Rachitskaya, S. Srivastava, S. Sharma and D. A. Mammo	2023	Evaluation and Comparison of Ophthalmic Scientific Abstracts and References by Current Artificial Intelligence Chatbots	JAMA Ophthalmol
104	J. T. Caranfa, N. K. Bommakanti, B. K. Young and P. Y. Zhao	2023	Accuracy of Vitreoretinal Disease Information From an Artificial Intelligence Chatbot	JAMA Ophthalmol
105	A. Taloni, V. Scorcia and G. Giannaccare	2023	Large Language Model Advanced Data Analysis Abuse to Create a Fake Data Set in Medical Research	JAMA Ophthalmol
106	A. S. Huang, K. Hirabayashi, L. Barna, D. Parikh and L. R. Pasquale	2024	Assessment of a Large Language Model's Responses to Questions and Cases About Glaucoma and Retina Management	JAMA Ophthalmol
107	A. Mihalache, R. S. Huang, M. M. Popovic, N. S. Patil, B. U. Pandya, R. Shor, A. Pereira, J. M. Kwok, P. Yan, D. T. Wong, P. J. Kertes and R. H. Muni	2024	Accuracy of an Artificial Intelligence Chatbot's Interpretation of Clinical Ophthalmic Images	JAMA Ophthalmol
108	F. Antaki, R. Chopra and P. A. Keane	2024	Vision-Language Models for Feature Detection of Macular Diseases on Optical Coherence Tomography	JAMA Ophthalmol
109	K. Beam, P. Sharma, B. Kumar, C. Wang, D. Brodsky, C. R. Martin and A. Beam	2023	Performance of a Large Language Model on Practice Questions for the Neonatal Board Examination	JAMA Pediatr
110	J. Barile, A. Margolis, G. Cason, R. Kim, S. Kalash, A. Tchaconas and R. Milanaik	2024	Diagnostic Accuracy of a Large Language Model in Pediatric Case Studies	JAMA Pediatr

https://doi.org/10.3348/kjr.2024.0788

111	N. Rabbani, C. Brown, M. Bedgood, R. L. Goldstein, J. L. Carlson, N. M. Pageler and K. E. Morse	2024	Evaluation of a Large Language Model to Identify Confidential Content in Adolescent Encounter Notes	JAMA Pediatr
112	P. Sheeran, A. Kenny, A. Bermudez, K. Gray, E. F. Galper, M. Boynton and S. M. Noar	2024	Artificial Intelligence Simulation of Adolescents' Responses to Vaping-Prevention Messages	JAMA Pediatr
113	A. Grigorian, J. Shipley, J. Nahmias, N. Nguyen, A. C. Schwed, B. A. Petrie and C. de Virgilio	2023	Implications of Using Chatbots for Future Surgical Education	JAMA Surg
114	P. Chung, C. T. Fong, A. M. Walters, N. Aghaeepour, M. Yetisgen and V. N. O'Reilly-Shah	2024	Large Language Model Capabilities in Perioperative Risk Prediction and Prognostication	JAMA Surg
115	H. Fraser, D. Crossland, I. Bacher, M. Ranney, T. Madsen and R. Hilliard	2023	Comparison of Diagnostic and Triage Accuracy of Ada Health and WebMD Symptom Checkers, ChatGPT, and Physicians for Patients in an Emergency Department: Clinical Data Analysis Study	JMIR Mhealth Uhealth
116	D. Van Veen, C. Van Uden, L. Blankemeier, J. B. Delbrouck, A. Aali, C. Bluethgen, A. Pareek, M. Polacin, E. P. Reis, A. Seehofnerová, N. Rohatgi, P. Hosamani, W. Collins, N. Ahuja, C. P. Langlotz, J. Hom, S. Gatidis, J. Pauly and A. S. Chaudhari	2024	Adapted large language models can outperform medical experts in clinical text summarization	Nat Med
117	R. H. Perlis, J. F. Goldberg, M. J. Ostacher and C. D. Schneck	2024	Clinical decision support for bipolar depression using large language models	Neuropsychopharmacology
118	M. R. Mejia, J. S. Arroyave, M. Saturno, L. C. M. Ndjonko, B. Zaidat, R. Rajjoub, W. Ahmed, I. Zapolsky and S. K. Cho	2024	Use of ChatGPT for Determining Clinical and Surgical Treatment of Lumbar Disc Herniation With Radiculopathy: A North American Spine Society Guideline Comparison	Neurospine
119	B. Zaidat, N. Shrestha, A. M. Rosenberg, W. Ahmed, R. Rajjoub, T. Hoang, M. R. Mejia, A. H. Duey, J. E. Tang, J. S. Kim and S. K. Cho	2024	Performance of a Large Language Model in the Generation of Clinical Guidelines for Antibiotic Prophylaxis in Spine Surgery	Neurospine
120	R. Ali, O. Y. Tang, I. D. Connolly, J. S. Fridley, J. H. Shin, P. L. Zadnik Sullivan, D. Cielo, A. A. Oyelese, C. E. Doberstein, A. E. Telfeian, Z. L. Gokaslan and W. F. Asaad	2023	Performance of ChatGPT, GPT-4, and Google Bard on a Neurosurgery Oral Boards Preparation Question Bank	Neurosurgery
121	A. Mishra, S. L. Begley, A. Chen, M. Rob, I. Pelcher, M. Ward and M. Schulder	2023	Exploring the Intersection of Artificial Intelligence and Neurosurgery: Let us be Cautious With ChatGPT	Neurosurgery
122	R. Ali, O. Y. Tang, I. D. Connolly, P. L. Zadnik Sullivan, J. H. Shin, J. S. Fridley, W. F. Asaad, D. Cielo, A. A. Oyelese, C. E. Doberstein, Z. L. Gokaslan and A. E. Telfeian	2023	Performance of ChatGPT and GPT-4 on Neurosurgery Written Board Examinations	Neurosurgery

https://doi.org/10.3348/kjr.2024.0788

123	A. A. Gajjar, R. P. Kumar, E. D. Paliwoda, C. C. Kuo, S. Adida, A. D. Legarreta, H. Deng, S. K. Anand, D. K. Hamilton, T. J. Buell, N. Agarwal, P. C. Gerszten and J. S. Hudson	2024	Usefulness and Accuracy of Artificial Intelligence Chatbot Responses to Patient Questions for Neurosurgical Procedures	Neurosurgery
124	M. Ward, P. Unadkat, D. Toscano, A. Kashanian, D. G. Lynch, A. C. Horn, R. S. D'Amico, M. Mittler and G. R. Baum	2024	A Quantitative Assessment of ChatGPT as a Neurosurgical Triaging Tool	Neurosurgery
125	C. A. Gao, F. M. Howard, N. S. Markov, E. C. Dyer, S. Ramesh, Y. Luo and A. T. Pearson	2023	Comparing scientific abstracts generated by ChatGPT to real abstracts with detectors and blinded human reviewers	NPJ Digit Med
126	L. Tang, Z. Sun, B. Idnay, J. G. Nestor, A. Soroush, P. A. Elias, Z. Xu, Y. Ding, G. Durrett, J. F. Rousseau, C. Weng and Y. Peng	2023	Evaluating large language models on medical evidence summarization	NPJ Digit Med
127	J. A. Omiye, J. C. Lester, S. Spichak, V. Rotemberg and R. Daneshjou	2023	Large language models propagate race-based medicine	NPJ Digit Med
128	C. Peng, X. Yang, A. Chen, K. E. Smith, N. PourNejatian, A. B. Costa, C. Martin, M. G. Flores, Y. Zhang, T. Magoc, G. Lipori, D. A. Mitchell, N. S. Ospina, M. M. Ahmed, W. R. Hogan, E. A. Shenkman, Y. Guo, J. Bian and Y. Wu	2023	A study of generative large language model for medical research and healthcare	NPJ Digit Med
129	T. Savage, A. Nayak, R. Gallo, E. Rangan and J. H. Chen	2024	Diagnostic reasoning prompts reveal the potential for large language model interpretability in medicine	NPJ Digit Med
130	L. Wang, X. Chen, X. Deng, H. Wen, M. You, W. Liu, Q. Li and J. Li	2024	Prompt engineering in consistency and reliability with the evidence-based guideline for LLMs	NPJ Digit Med
131	R. Ali, I. D. Connolly, O. Y. Tang, F. N. Mirza, B. Johnston, H. F. Abdulrazeq, R. K. Lim, P. F. Galamaga, T. J. Libby, N. R. Sodha, M. W. Groff, Z. L. Gokaslan, A. E. Telfeian, J. H. Shin, W. F. Asaad, J. Zou and C. E. Doberstein	2024	Bridging the literacy gap for surgical consents: an AI-human expert collaborative approach	NPJ Digit Med
132	J. Huang, D. M. Yang, R. Rong, K. Nezafati, C. Treager, Z. Chi, S. Wang, X. Cheng, Y. Guo, L. J. Klesse, G. Xiao, E. D. Peterson, X. Zhan and Y. Xie	2024	A critical assessment of using ChatGPT for extracting structured data from clinical notes	NPJ Digit Med
133	R. C. L. Brewster, P. Gonzalez, R. Khazanchi, A. Butler, R. Selcer, D. Chu, B. P. Aires, M. Luercio and J. D. Hron	2024	Performance of ChatGPT and Google Translate for Pediatric Discharge Instruction Translation	Pediatrics
134	J. H. Hollman, B. A. Cloud-Biebl, D. A. Krause and D. Q. Calley	2024	Detecting Artificial Intelligence-Generated Personal Statements in Professional Physical Therapist Education Program Applications: A Lexical Analysis	Phys Ther

https://doi.org/10.3348/kjr.2024.0788

135	A. Cocci, M. Pezzoli, M. Lo Re, G. I. Russo, M. G. Asmundo, M. Fode, G. Cacciamani, S. Cimino, A. Minervini and E. Durukan	2024	Quality of information and appropriateness of ChatGPT outputs for urology patients	Prostate Cancer Prostatic Dis
136	R. Lombardo, G. Gallo, J. Stira, B. Turchi, G. Santoro, S. Riolo, M. Romagnoli, A. Cicione, G. Tema, A. Pastore, Y. Al Salhi, A. Fuschi, G. Franco, A. Nacchia, A. Tubaro and C. De Nunzio	2024	Quality of information and appropriateness of Open AI outputs for prostate cancer	Prostate Cancer Prostatic Dis
137	J. S. Hershenhouse, D. Mokhtar, M. B. Eppler, S. Rodler, L. Storino Ramacciotti, C. Ganjavi, B. Hom, R. J. Davis, J. Tran, G. I. Russo, A. Cocci, A. Abreu, I. Gill, M. Desai and G. E. Cacciamani	2024	Accuracy, readability, and understandability of large language models for prostate cancer information to the public	Prostate Cancer Prostatic Dis
138	A. K. Puerto Nino, V. Garcia Perez, S. Secco, C. De Nunzio, R. Lombardo, K. A. O. Tikkinen and D. S. Elterman	2024	Can ChatGPT provide high-quality patient information on male lower urinary tract symptoms suggestive of benign prostate enlargement?	Prostate Cancer Prostatic Dis
139	C. A. Mallio, A. C. Sertorio, C. Bernetti and B. Beomonte Zobel	2023	Large language models for structured reporting in radiology: performance of GPT-4, ChatGPT-3.5, Perplexity and Bing	Radiol Med
140	R. Bhayana, S. Krishna and R. R. Bleakney	2023	Performance of ChatGPT on a Radiology Board-style Examination: Insights into Current Strengths and Limitations	Radiology
141	R. J. Gertz, A. C. Bunck, S. Lennartz, T. Dratsch, A. I. Iuga, D. Maintz and J. Kottlors	2023	GPT-4 for Automated Determination of Radiological Study and Protocol based on Radiology Request Forms: A Feasibility Study	Radiology
142	A. A. Rahsepar, N. Tavakoli, G. H. J. Kim, C. Hassani, F. Abtin and A. Bedayat	2023	How AI Responds to Common Lung Cancer Questions: ChatGPT vs Google Bard	Radiology
143	A. Rau, S. Rau, D. Zoeller, A. Fink, H. Tran, C. Wilpert, J. Nattenmueller, J. Neubauer, F. Bamberg, M. Reisert and M. F. Russe	2023	A Context-based Chatbot Surpasses Trained Radiologists and Generic ChatGPT in Following the ACR Appropriateness Guidelines	Radiology
144	P. Mukherjee, B. Hou, R. B. Lanfredi and R. M. Summers	2023	Feasibility of Using the Privacy-preserving Large Language Model Vicuna for Labeling Radiology Reports	Radiology
145	M. A. Fink, A. Bischoff, C. A. Fink, M. Moll, J. Kroschke, L. Dulz, C. P. Heußel, H. U. Kauczor and T. F. Weber	2023	Potential of ChatGPT and GPT-4 for Data Mining of Free-Text CT Reports on Lung Cancer	Radiology
146	R. Doshi, K. S. Amin, P. Khosla, S. S. Bajaj, S. Chheang and H. P. Forman	2024	Quantitative Evaluation of Large Language Models to Streamline Radiology Report Impressions: A Multimodal Retrospective Analysis	Radiology
147	A. Cozzi, K. Pinker, A. Hidber, T. Zhang, L. Bonomo, R. Lo Gullo, B. Christianson, M. Curti, S. Rizzo, F. Del Grande, R. M. Mann and S. Schiaffino	2024	BI-RADS Category Assignments by GPT-3.5, GPT-4, and Google Bard: A Multilanguage Study	Radiology

https://doi.org/10.3348/kjr.2024.0788

148	S. H. Wu, W. J. Tong, M. D. Li, H. T. Hu, X. Z. Lu, Z. R. Huang, X. X. Lin, R. F. Lu, M. D. Lu, L. D. Chen and W. Wang	2024	Collaborative Enhancement of Consistency and Accuracy in US Diagnosis of Thyroid Nodules Using Large Language Models	Radiology
149	R. J. Gertz, T. Dratsch, A. C. Bunck, S. Lennartz, A. I. Iuga, M. G. Hellmich, T. Persigehl, L. Pennig, C. H. Gietzen, P. Fervers, D. Maintz, R. Hahnfeldt and J. Kottlors	2024	Potential of GPT-4 for Detecting Errors in Radiology Reports: Implications for Reporting Accuracy	Radiology
150	S. Krishna, N. Bhambra, R. Bleakney and R. Bhayana	2024	Evaluation of Reliability, Repeatability, Robustness, and Confidence of GPT-3.5 and GPT-4 on a Radiology Board-style Examination	Radiology
151	N. C. Lehnen, F. Dorn, I. C. Wiest, H. Zimmermann, A. Radbruch, J. N. Kather and D. Paech	2024	Data Extraction from Free-Text Reports on Mechanical Thrombectomy in Acute Ischemic Stroke Using ChatGPT: A Retrospective Analysis	Radiology
152	R. Bhayana, B. Nanda, T. Dehkharghanian, Y. Deng, N. Bhambra, G. Elias, D. Datta, A. Kambadakone, C. G. Shwaartz, C. A. Moulton, D. Henault, S. Gallinger and S. Krishna	2024	Large Language Models for Automated Synoptic Reports and Resectability Categorization in Pancreatic Cancer	Radiology
153	Y. Zhou, H. Ong, P. Kennedy, C. C. Wu, J. Kazam, K. Hentel, A. Flanders, G. Shih and Y. Peng	2024	Evaluating GPT-V4 (GPT-4 with Vision) on Detection of Radiologic Findings on Chest Radiographs	Radiology
154	M. Guckenberger, N. Andratschke, M. Ahmadsei, S. M. Christ, A. E. Heusel, S. Kamal, T. E. Kroese, E. L. Looman, S. Reichl, E. Vlaskou Badra, J. von der Grün, J. Willmann, S. Tanadini-Lang and M. Mayinger	2023	Potential of ChatGPT in facilitating research in radiation oncology?	Radiother Oncol
155	D. J. Wu and J. E. Bibault	2024	Pilot applications of GPT-4 in radiation oncology: Summarizing patient symptom intake and targeted chatbot applications	Radiother Oncol
156	T. Scquizzato, F. Semeraro, P. Swindell, R. Simpson, M. Angelini, A. Gazzato, U. Sajjad, E. G. Bignami, G. Landoni, T. R. Keeble and M. Mion	2024	Testing ChatGPT ability to answer laypeople questions about cardiac arrest and cardiopulmonary resuscitation	Resuscitation
157	C. C. Y. Gue, N. D. A. Rahim, W. Rojas-Carabali, R. Agrawal, P. Rk, J. Abisheganaden and W. F. Yip	2024	Evaluating the OpenAI's GPT-3.5 Turbo's performance in extracting information from scientific articles on diabetic retinopathy	Syst Rev
158	F. Dennstädt, J. Zink, P. M. Putora, J. Hastings and N. Cihoric	2024	Title and abstract screening for literature reviews using large language models: an exploratory study in the biomedical domain	Syst Rev
159	M. Shiraishi, K. Kanayama, D. Kurita, Y. Moriwaki and M. Okazaki	2024	Performance of artificial intelligence chatbots in interpreting clinical images of pressure injuries	Wound Repair Regen