Reslaj: Religion Education Social Laa Roiba Journal

Volume 8 Nomor 11 (2026) 1096 – 1118 P-ISSN 2656-274x E-ISSN 2656-4691
DOI: 10.47476/reslaj.v8i11.13659

Artificial Intelligence Chatbot dalam Pelayanan Kesehatan: Dampaknya terhadap Efisiensi Pelayanan dan Pengalaman Pasien

Intan Sekarlangit, Kosasih, Farida Yuliaty
Universitas Sangga Buana
i.sekarlangit@gmail.com, kosasih@usbypkp.ac.id, farida.yuliaty@usbypkp.ac.id

ABSTRACT

AI chatbots are increasingly adopted in healthcare, but evidence of their effects on service efficiency and patient experience remains fragmented. This systematic literature review synthesized 2021–2026 evidence using PRISMA 2020. Searches of PubMed/PMC, DOAJ, and publisher websites identified 14 relevant articles from 166 records. Findings showed efficiency gains in structured tasks, including a 57.3% reduction in history-taking time, while triage accuracy varied (pooled ChatGPT-4 accuracy: 0.86; I^2 = 93%). Patient experience was generally positive, with 73% satisfaction among perioperative patients, although chatbots were less effective than physicians in reducing anxiety. Key barriers included privacy, empathy, system integration, and limited real-patient data. AI chatbots can improve healthcare services when applied to structured processes with appropriate human oversight.

Keywords : Artificial intelligence chatbot, service efficiency, patient experience, hospital management, systematic literature review.

ABSTRAK

Chatbot AI semakin banyak digunakan di layanan kesehatan, tetapi bukti mengenai dampaknya terhadap efisiensi layanan dan pengalaman pasien masih beragam. Tinjauan literatur sistematis ini mensintesis bukti tahun 2021–2026 menggunakan PRISMA 2020. Dari 166 rekaman yang teridentifikasi, 14 artikel dianalisis. Hasil menunjukkan peningkatan efisiensi pada tugas terstruktur, termasuk penghematan waktu pengambilan riwayat sebesar 57,3%, sementara akurasi triase bervariasi (akurasi ChatGPT-4 sebesar 0,86; I^2 = 93%). Pengalaman pasien umumnya positif, dengan tingkat kepuasan 73% pada pasien perioperatif, meskipun chatbot kurang efektif dibandingkan dokter dalam mengurangi kecemasan. Hambatan utama meliputi privasi, empati, integrasi sistem, dan keterbatasan data pasien. Chatbot AI berpotensi meningkatkan layanan kesehatan jika diterapkan pada proses terstruktur dengan pengawasan manusia.

Kata kunci : Chatbot kecerdasan buatan, efisiensi pelayanan, pengalaman pasien, manajemen rumah sakit, tinjauan sistematis.

PENDAHULUAN

Rumah sakit menghadapi tuntutan pelayanan yang cepat, efisien, dan berpusat pada pasien di tengah beban administrasi dan keterbatasan tenaga kesehatan. Salah satu teknologi yang banyak dibahas untuk menjawab tuntutan tersebut adalah *chatbot* berbasis kecerdasan buatan (*artificial intelligence*/AI), yaitu program yang mensimulasikan percakapan manusia untuk berbagai kebutuhan layanan kesehatan (Laymouna dkk., 2024). Minat riset meningkat tajam; Hindelang dkk. (2024) menunjukkan publikasi dengan kata kunci *chatbot* dan kedokteran naik dari satu studi pada 2017 menjadi 445 studi pada 2023. Pandemi COVID-19 turut

Volume 8 Nomor 11 (2026) 1096 – 1118 P-ISSN 2656-274x E-ISSN 2656-4691
DOI: 10.47476/reslaj.v8i11.13659

mendorong pemanfaatan *chatbot* dalam respons kesehatan masyarakat (Amiri dan Karahanna, 2022), dan penggunaannya pada layanan kesehatan serta onkologi telah dikaji secara sistematis (Xu dkk., 2021).

Secara konseptual, *chatbot* berkembang dari program berbasis aturan menuju model bahasa besar (*large language model*/LLM) yang menghasilkan respons secara bebas (Adamopoulou dan Moussiades, 2020). Sistem berbasis aturan memberikan jawaban yang dapat diprediksi tetapi kaku, sedangkan sistem berbasis LLM lebih fleksibel tetapi dapat menghasilkan jawaban yang keliru. Sejak ChatGPT tersedia luas pada akhir 2022, sebagian studi evaluasi klinis memakai model semacam itu (Sarbay dkk., 2023; Colakca dkk., 2024; Zhang dkk., 2025).

Bagi manajemen rumah sakit, potensi *chatbot* terletak pada dua dimensi utama, yaitu efisiensi pelayanan dan pengalaman pasien. Laymouna dkk. (2024) mengelompokkan peran *chatbot* ke dalam layanan kesehatan jarak jauh dan bantuan administratif bagi penyedia layanan, dengan manfaat berupa peningkatan mutu, efisiensi, dan efektivitas biaya. Meski demikian, keterbatasan etika, medikolegal, keselamatan, teknis, dan pengalaman pengguna juga dilaporkan (Laymouna dkk., 2024). Evaluasi aplikasi model bahasa besar di bidang kesehatan pun masih terfragmentasi: dari 519 studi yang ditinjau, hanya 5% memakai data perawatan pasien nyata dan tugas administratif jarang diteliti (Bedi dkk., 2025).

Efisiensi pelayanan dan pengalaman pasien dipahami di sini sebagai dua luaran yang berbeda. Kerangka Donabedian (1988) membedakan struktur, proses, dan luaran layanan; efisiensi dalam tinjauan ini terutama menyangkut proses (waktu, beban kerja, waktu tunggu, akurasi keputusan awal, dan biaya), sedangkan pengalaman pasien menyangkut keseluruhan interaksi pasien dengan sistem layanan, termasuk komunikasi, informasi, dan keterlibatan, sehingga tidak identik dengan kepuasan (Wolf dkk., 2014). Keduanya selaras dengan dimensi mutu layanan yang tepat waktu, efisien, dan berpusat pada pasien (Institute of Medicine, 2001).

Di Indonesia, transformasi digital kesehatan diarahkan agar teknologi berfokus pada pelayanan bagi masyarakat, bukan sekadar pelaporan, sebagaimana tertuang dalam Cetak Biru Strategi Transformasi Digital Kesehatan 2024 (Kementerian Kesehatan Republik Indonesia, 2021) dan diperkuat melalui Keputusan Menteri Kesehatan Nomor HK.01.07/MENKES/1559/2022 (Kementerian Kesehatan Republik Indonesia, 2022a). Arah kebijakan ini membuka ruang bagi rumah sakit untuk menilai teknologi percakapan, tetapi keputusan adopsi tetap memerlukan bukti tentang manfaat dan risikonya. Pengalaman digitalisasi rumah sakit sebelumnya juga menunjukkan bahwa rekam medis elektronik dapat meningkatkan efisiensi pelayanan melalui penurunan biaya operasional dan optimalisasi proses layanan, tetapi manfaatnya bergantung pada kesiapan organisasi, kapabilitas sumber daya manusia, dan integrasi sistem (Kosasih dkk., 2026a).

Tinjauan yang ada cenderung berfokus pada domain sempit, seperti anamnesis (Hindelang dkk., 2024), perawatan perioperatif (Lin dkk., 2024), perubahan perilaku kesehatan (Aggarwal dkk., 2023), atau triase gawat darurat

Volume 8 Nomor 11 (2026) 1096 – 1118 P-ISSN 2656-274x E-ISSN 2656-4691
DOI: 10.47476/reslaj.v8i11.13659

(Kaboudi dkk., 2024). Protokol tinjauan yang baru dipublikasikan juga menegaskan masih terbatasnya sintesis tentang penerapan teknologi ini di berbagai konteks layanan dan luaran yang dilaporkan, termasuk kepuasan pasien dan efisiensi operasional (Goossens dkk., 2026). Oleh karena itu, penelitian ini bertujuan mensintesis secara sistematis bukti terbitan 2021–2026 mengenai dampak AI *chatbot* dalam pelayanan kesehatan terhadap efisiensi pelayanan dan pengalaman pasien, serta merumuskan implikasinya bagi manajemen rumah sakit. Tinjauan ini menjawab empat pertanyaan penelitian (*research question*/RQ) berikut.

RQ1. Bagaimana karakteristik studi dan bentuk penerapan AI *chatbot* di pelayanan kesehatan yang dilaporkan pada 2021–2026?
RQ2. Bagaimana dampak AI *chatbot* terhadap efisiensi pelayanan, yaitu waktu, beban kerja, akurasi triase, biaya, dan waktu tunggu?
RQ3. Bagaimana dampak AI *chatbot* terhadap pengalaman pasien, yaitu kepuasan, pengetahuan, komunikasi, empati, dan kecemasan?
RQ4. Apa hambatan, kesenjangan bukti, dan implikasi tata kelola bagi manajemen rumah sakit?

Kontribusi tinjauan ini ada tiga: efisiensi dan pengalaman pasien disintesis bersama lintas domain layanan; kualitas studi dan kepastian bukti dinilai secara eksplisit sehingga bukti primer dapat dibedakan dari angka sekunder; dan temuan diterjemahkan ke dalam tahapan penerapan dan indikator pemantauan bagi manajemen rumah sakit.

METODE PENELITIAN

Desain dan Protokol Penelitian

Penelitian ini merupakan *systematic literature review* (SLR) yang dilaporkan mengikuti pedoman *Preferred Reporting Items for Systematic Reviews and Meta-Analyses* (PRISMA) 2020 (Page dkk., 2021). Empat pertanyaan penelitian (RQ1–RQ4) dirumuskan dengan kerangka PICOS (Tabel 1), yaitu bagaimana karakteristik studi dan bentuk penerapan AI *chatbot* di pelayanan kesehatan, bagaimana dampaknya terhadap efisiensi pelayanan, bagaimana dampaknya terhadap pengalaman pasien, serta apa hambatan dan implikasinya bagi manajemen rumah sakit. Protokol tinjauan tidak didaftarkan pada PROSPERO.

Perumusan pertanyaan dan kriteria kelayakan mengacu pada *Cochrane Handbook* (Higgins dkk., 2023), sedangkan pelaporan sintesis tanpa meta-analisis mengacu pada pedoman *synthesis without meta-analysis* (SWiM) (Campbell dkk., 2020). Karena protokol tidak didaftarkan, kepatuhan terhadap rencana awal tidak dapat diverifikasi secara independen dan hal ini dicatat sebagai keterbatasan.

Strategi Pencarian dan Sumber Informasi

Pencarian dilakukan pada September 2026 melalui mesin pencari web terhadap literatur yang terindeks di PubMed/PMC, *Directory of Open Access Journals* (DOAJ), dan situs penerbit jurnal (JMIR, Frontiers, BMJ, JAMA Network, dan Springer

Volume 8 Nomor 11 (2026) 1096 – 1118 P-ISSN 2656-274x E-ISSN 2656-4691
DOI: 10.47476/reslaj.v8i11.13659

Nature). Kata kunci disusun dalam bahasa Inggris dan Indonesia, yaitu ("AI chatbot" OR "conversational agent" OR ChatGPT) AND (healthcare OR hospital OR outpatient OR triage) AND ("patient satisfaction" OR "patient experience" OR efficiency OR workload OR "waiting time"), serta padanannya "chatbot rumah sakit kepuasan pasien efisiensi pelayanan". Sebanyak 18 kueri dijalankan, termasuk kueri lanjutan untuk memverifikasi data bibliografis artikel yang teridentifikasi, dan seluruh hasil yang ditampilkan dicatat (n = 165). Satu artikel acuan ditambahkan melalui pencarian tambahan (n = 1).

Ringkasan strategi pencarian disajikan pada Tabel 2. Pencarian tidak memakai sintaks khusus basis data, misalnya *medical subject headings*, karena dilakukan melalui mesin pencari web; pilihan ini memengaruhi sensitivitas pencarian dan dibahas sebagai keterbatasan.

Tabel 2. Ringkasan Strategi Pencarian

Komponen	Rincian
Tanggal pencarian	September 2026
Sumber informasi	Literatur terindeks di PubMed/PMC dan DOAJ serta situs penerbit (JMIR, Frontiers, BMJ, JAMA Network, Springer Nature), diakses melalui mesin pencari web
Blok intervensi	("AI chatbot" OR "conversational agent" OR ChatGPT)
Blok konteks	AND (healthcare OR hospital OR outpatient OR triage)
Blok luaran	AND ("patient satisfaction" OR "patient experience" OR efficiency OR workload OR "waiting time")
Kueri berbahasa Indonesia	"chatbot rumah sakit kepuasan pasien efisiensi pelayanan"
Batasan (diterapkan saat penapisan)	Terbit 2021–2026; bahasa Inggris atau Indonesia; artikel *peer-reviewed*
Jumlah kueri dan hasil	18 kueri (termasuk kueri lanjutan untuk memverifikasi data bibliografis); 165 rekaman dari pencarian dan 1 rekaman dari pencarian tambahan

Sumber : Analisis peneliti, 2026

Kriteria Kelayakan

Studi dimasukkan apabila terbit pada 2021–2026 dalam jurnal *peer-reviewed*, membahas AI *chatbot* atau *conversational agent* di layanan kesehatan, dan melaporkan luaran efisiensi dan/atau pengalaman pasien (Tabel 1).

Tabel 1. Kriteria Kelayakan Studi (PICOS)

Komponen	Kriteria inklusi	Kriteria eksklusi
Population	Pasien atau penyedia/pengelola layanan kesehatan (rumah sakit, poliklinik, gawat darurat, perioperatif)	Populasi di luar layanan kesehatan

Reslaj: Religion Education Social Laa Roiba Journal

Volume 8 Nomor 11 (2026) 1096 – 1118 P-ISSN 2656-274x E-ISSN 2656-4691
DOI: 10.47476/reslaj.v8i11.13659

Komponen	Kriteria inklusi	Kriteria eksklusi
Intervention	AI *chatbot*/*conversational agent*, termasuk berbasis *large language model* (misalnya ChatGPT)	Sistem non-percakapan (misalnya kios atau aplikasi tanpa AI)
Comparator	Layanan konvensional oleh tenaga kesehatan, tanpa *chatbot*, atau tanpa pembanding	–
Outcome	Efisiensi pelayanan (waktu, beban kerja, akurasi triase, biaya, waktu tunggu) dan/atau pengalaman pasien (kepuasan, keterlibatan, empati, penerimaan)	Tidak melaporkan luaran tersebut
Study design	Studi primer dan tinjauan (sistematis, *scoping*, *rapid*) yang *peer-reviewed*	*Preprint*, protokol, prosiding/poster/abstrak, editorial, sumber non-ilmiah
Tahun dan bahasa	2021–2026; Inggris atau Indonesia	Sebelum 2021; bahasa lain
Ketersediaan data	Data hasil dapat diekstraksi dari teks/abstrak yang tersedia	Data hasil tidak dapat diekstraksi lengkap

Sumber : Analisis peneliti, 2026

Temuan studi primer diperlakukan sebagai bukti primer, sedangkan angka yang bersumber dari tinjauan dilaporkan sebagaimana dikutip dan diperlakukan sebagai bukti sekunder, sehingga angka gabungan tinjauan tidak dianggap setara dengan estimasi efek studi asli.

Seleksi Studi dan Ekstraksi Data

Seleksi dilakukan bertahap, yaitu penghapusan duplikat dan sumber non-ilmiah, penapisan judul dan abstrak, lalu penilaian kelayakan. Seleksi dan ekstraksi dilakukan oleh satu penelaah sehingga tidak ada penilaian antarpenelaah. Data diekstraksi dari teks atau abstrak yang dapat diakses ke dalam formulir terstruktur yang memuat: (1) identitas artikel (penulis, tahun, jurnal); (2) desain dan ukuran sampel atau jumlah studi yang ditinjau; (3) negara dan konteks layanan; (4) jenis intervensi *chatbot* dan pembanding bila ada; (5) luaran efisiensi; (6) luaran pengalaman pasien; serta (7) hambatan dan keterbatasan yang dilaporkan penulis studi asli. Data yang tidak dilaporkan dicatat sebagai tidak tersedia dan tidak diimputasi.

Penilaian Kualitas dan Kepastian Bukti

Karena berbagai desain studi dimasukkan, penilaian kualitas dilakukan secara naratif dengan kerangka yang sama untuk semua studi. Setiap studi dinilai pada empat pertimbangan: (1) kesesuaian desain untuk menjawab pertanyaan

Volume 8 Nomor 11 (2026) 1096 – 1118 P-ISSN 2656-274x E-ISSN 2656-4691
DOI: 10.47476/reslaj.v8i11.13659

penelitian; (2) sumber data, yaitu pasien atau pengguna nyata dibandingkan skenario atau forum daring; (3) ukuran sampel dan presisi yang dilaporkan; serta (4) keterbatasan yang dilaporkan penulis asli atau tampak dari desainnya. Studi lalu diberi peringkat kepercayaan Tinggi, Sedang, atau Rendah terhadap perannya dalam sintesis ini. Peringkat ini merupakan penilaian peneliti dan tidak menggantikan alat formal, yaitu RoB 2 untuk uji acak (Sterne dkk., 2019), AMSTAR 2 untuk tinjauan sistematis (Shea dkk., 2017), dan daftar periksa JBI untuk studi potong lintang (Moola dkk., 2020).

Tingkat kepastian bukti pada setiap kelompok luaran dinilai secara naratif dengan mengacu pada pertimbangan *Grading of Recommendations Assessment, Development and Evaluation* (GRADE) (Guyatt dkk., 2008), yaitu desain studi, risiko bias, inkonsistensi, ketidaklangsungan, dan ketidakpresisian, dengan empat tingkat: sangat rendah, rendah, sedang, dan tinggi. Penerapan ini tidak formal karena penilaian risiko bias pada tingkat studi individual tidak dilakukan.

Sintesis Data

Karena heterogenitas desain, populasi, dan ukuran luaran, sintesis dilakukan secara naratif-tematik tanpa meta-analisis. Temuan dikelompokkan ke dalam tema efisiensi pelayanan, pengalaman pasien, serta hambatan dan tata kelola. Angka yang berasal dari tinjauan sekunder dilaporkan sebagaimana dikutip dan tidak diperlakukan sebagai estimasi efek primer.

Sintesis naratif mengikuti tahapan yang dianjurkan Popay dkk. (2006), yaitu ringkasan awal, eksplorasi pola antarstudi, dan penilaian kekokohan sintesis. Tema utama diturunkan dari pertanyaan penelitian dan subtema dari data, dengan prinsip analisis tematik (Braun dan Clarke, 2006). Sesuai SWiM (Campbell dkk., 2020), metrik yang dipakai adalah proporsi gabungan dengan interval kepercayaan (IK) 95% dan koefisien kappa (κ), yang ditafsirkan dengan kategori Landis dan Koch (1977). Heterogenitas (I^2) dilaporkan sebagaimana dikutip dan ditafsirkan berdasarkan panduan Cochrane (Higgins dkk., 2023). Kekokohan sintesis diperiksa dengan memisahkan bukti primer dari sekunder dan mencermati kemungkinan tumpang tindih studi primer antartinjauan (Pieper dkk., 2014); tumpang tindih tidak dihitung secara kuantitatif.

HASIL DAN PEMBAHASAN

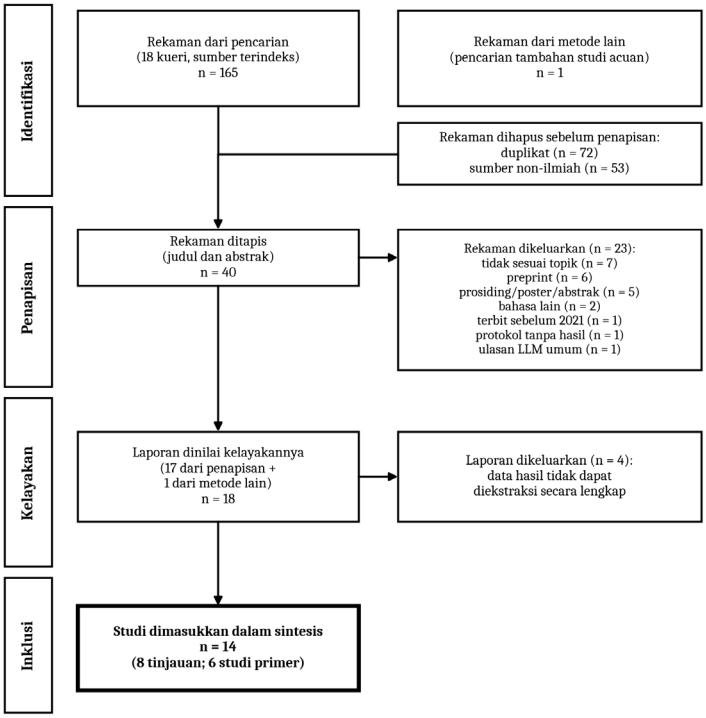
Hasil

Hasil Seleksi Studi

Pencarian menghasilkan 165 rekaman. Sebanyak 125 rekaman dikeluarkan sebelum penapisan, yaitu 72 duplikat dan 53 sumber non-ilmiah (blog vendor, berita, forum, skripsi atau dokumen instansi, registri uji klinis, dan halaman non-artikel). Sebanyak 40 rekaman ilmiah unik ditapis dan 23 dikeluarkan pada tahap judul-abstrak karena tidak sesuai topik (n = 7), berupa *preprint* (n = 6), prosiding/poster/abstrak (n = 5), berbahasa lain (n = 2), terbit sebelum 2021 (n = 1),

Volume 8 Nomor 11 (2026) 1096 – 1118 P-ISSN 2656-274x E-ISSN 2656-4691
DOI: 10.47476/reslaj.v8i11.13659

protokol tanpa hasil (n = 1), atau ulasan model bahasa umum di luar konteks layanan pasien (n = 1). Tujuh belas rekaman bersama satu rekaman dari pencarian tambahan dinilai kelayakannya; empat dikeluarkan karena data hasil tidak dapat diekstraksi secara lengkap. Sebanyak 14 artikel dimasukkan dalam sintesis (Gambar 1).



Sumber : Diadaptasi dari Page dkk. (2021), data peneliti, 2026

Gambar 1. Diagram Alir Seleksi Studi (PRISMA 2020)

Karakteristik Studi

Empat belas artikel terbit pada 2022–2025 (satu artikel pada 2022, tiga pada 2023, enam pada 2024, dan empat pada 2025). Terdapat delapan tinjauan (empat tinjauan sistematis, dua tinjauan sistematis dengan meta-analisis, satu *scoping review*, dan satu *rapid review*) serta enam studi primer (lima studi potong lintang dan satu uji acak terkontrol). Tiga studi primer dilakukan di Tiongkok, dua di Türkiye, dan satu memakai forum tanya-jawab medis daring publik; tidak ada studi dari Indonesia (Tabel 3 dan Gambar 2).

Menurut fokus layanannya, tiga artikel membahas triase gawat darurat berbasis *Emergency Severity Index* (ESI), satu anamnesis, dua perioperatif, tiga pengalaman atau komunikasi pasien, satu perubahan perilaku kesehatan, dan empat memberikan gambaran lintas layanan atau kerangka evaluasi. Empat dari enam studi primer mengevaluasi ChatGPT (Ayers dkk., 2023; Sarbay dkk., 2023; Colakca dkk., 2024; Zhang dkk., 2025). Tidak satu pun studi primer mengukur waktu tunggu, waktu proses, atau biaya secara langsung; indikator operasional tersebut hanya muncul pada tingkat tinjauan (Hindelang dkk., 2024; Laymouna dkk., 2024; Wah, 2025).

Reslaj: Religion Education Social Laa Roiba Journal

Volume 8 Nomor 11 (2026) 1096 – 1118 P-ISSN 2656-274x E-ISSN 2656-4691
DOI: 10.47476/reslaj.v8i11.13659

Tabel 3. Karakteristik dan Temuan Utama Studi yang Dimasukkan (n = 14)

No	Penulis (tahun)	Desain dan sampel	Fokus/konteks	Temuan utama
1	Li dkk. (2022)	Survei lapangan potong lintang; 172 kuesioner; PLS-SEM dan analisis tematik	Rawat jalan rumah sakit di Shanghai; niat keberlanjutan penggunaan *conversational agent*	Mengidentifikasi faktor penentu niat keberlanjutan berbasis model *expectation-confirmation* dan menyusun implikasi manajerial
2	Aggarwal dkk. (2023)	Tinjauan sistematis; 15 studi	Perubahan perilaku kesehatan	Efikasi tinggi pada gaya hidup sehat (n = 6), berhenti merokok (n = 4), kepatuhan pengobatan (n = 2); kelayakan, penerimaan, dan usabilitas beragam; risiko bias sedang–tinggi
3	Ayers dkk. (2023)	Potong lintang; 195 pertanyaan pasien dari forum daring; penilai tenaga kesehatan berlisensi	Komunikasi pasien–dokter (bukan setting klinis)	Respons *chatbot* dipilih pada 78,6% dari 585 penilaian; dinilai lebih berkualitas dan lebih empatik
4	Sarbay dkk. (2023)	Potong lintang berbasis skenario; 50 kasus ESI	Triase gawat darurat	κ = 0,341 (kesepakatan cukup); *over-triage* 22%, *under-triage* 18%
5	Colakca dkk. (2024)	Potong lintang; pasien dewasa IGD yang datang berturut-turut selama 24 jam; komite pakar	Triase IGD (ESI)	κ = 0,659; akurasi 76,6%; ESI 1–2: κ = 0,828, spesifisitas 95,63%, NPV 98,17%
6	Ding dkk. (2024)	*Systematic scoping review*	Evaluasi *conversational agent* AI pada intervensi kesehatan	Kerangka evaluasi selaras standar WHO untuk teknologi kesehatan digital: efektivitas, keselamatan, kepuasan pengguna
7	Hindelang dkk. (2024)	Tinjauan sistematis; 203 rekaman menjadi 18 studi	Anamnesis (pengambilan riwayat medis)	Pengumpulan data otomatis 24 jam meningkatkan efisiensi dan aksesibilitas; keterlibatan dan kepuasan meningkat; kualitas studi beragam

Volume 8 Nomor 11 (2026) 1096 – 1118 P-ISSN 2656-274x E-ISSN 2656-4691
DOI: 10.47476/reslaj.v8i11.13659

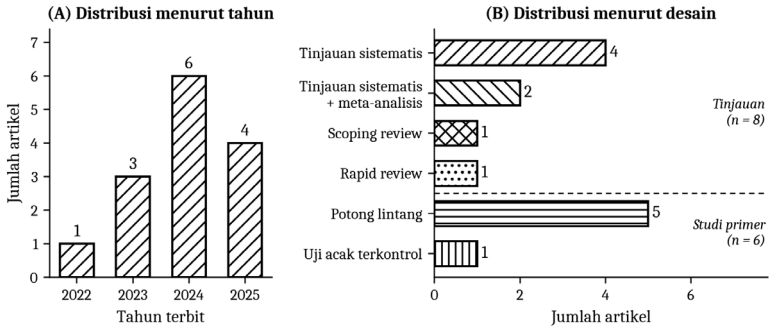
No	Penulis (tahun)	Desain dan sampel	Fokus/konteks	Temuan utama
8	Kaboudi dkk. (2024)	Tinjauan sistematis dan meta-analisis; 14 studi, 1.412 pasien/skenario	Triase gawat darurat	Akurasi gabungan ChatGPT-4 0,86 (IK 95% 0,64–0,98); I^2 = 93%
9	Laymouna dkk. (2024)	*Rapid review* 2017–2023	Peran *chatbot* di layanan kesehatan	Dua peran (layanan jarak jauh, bantuan administratif); manfaat mutu, efisiensi, efektivitas biaya; keterbatasan etika, medikolegal, teknis, pengalaman pengguna
10	Lin dkk. (2024)	Tinjauan sistematis dan meta-analisis; 8 uji, 1.073 dewasa	Perioperatif	73% puas (IK 95% 0,62–0,85); 80% merasa menambah pengetahuan (IK 95% 0,74–0,87); tidak ada bahaya dilaporkan
11	Bedi dkk. (2025)	Tinjauan sistematis; 519 studi (2022–Feb 2024)	Evaluasi aplikasi model bahasa besar	Hanya 5% memakai data pasien nyata; tugas administratif jarang dikaji; akurasi menjadi dimensi dominan
12	Li dkk. (2025)	Potong lintang; 394 pasien rawat jalan rumah sakit tersier	Pengalaman pasien terkait dokter (Tiongkok)	Pengguna memiliki skor pengalaman lebih tinggi; regresi: +7,51%; komunikasi, informasi kesehatan, luaran jangka pendek, kepuasan umum
13	Wah (2025)	Tinjauan sistematis; 29 studi dari 116 rekaman	*Chatbot* hibrida	Readmisi turun hingga 25%, keterlibatan +30%, waktu tunggu konsultasi −15% (dilaporkan); hambatan: kepercayaan, privasi, integrasi, biaya
14	Zhang dkk. (2025)	Uji acak tersamar tunggal; 75 pasien bedah urolitiasis (1:1:1)	Edukasi pra-operasi (Tiongkok)	Kecemasan turun pada semua kelompok; penurunan terbesar pada dokter, lalu ChatGPT, lalu Ernie Bot; kepuasan tinggi pada dokter dan ChatGPT

Reslaj: Religion Education Social Laa Roiba Journal

Volume 8 Nomor 11 (2026) 1096 – 1118 P-ISSN 2656-274x E-ISSN 2656-4691
DOI: 10.47476/reslaj.v8i11.13659

Sumber : Analisis peneliti, 2026

Keterangan : IGD = instalasi gawat darurat; NPV = *negative predictive value*; IK = interval kepercayaan; κ = kappa; PLS-SEM = *partial least squares structural equation modeling*.



Sumber : Analisis peneliti, 2026

Gambar 2. Distribusi Artikel yang Dimasukkan Menurut Tahun dan Desain (n = 14)

Penilaian Kualitas Studi

Hasil penilaian kualitas naratif disajikan pada Tabel 4. Satu artikel dinilai Tinggi untuk perannya dalam sintesis (Bedi dkk., 2025, sebagai penunjuk kesenjangan bukti), sepuluh dinilai Sedang, dan tiga dinilai Rendah (Ayers dkk., 2023; Sarbay dkk., 2023; Wah, 2025). Keterbatasan yang berulang adalah data pasien nyata yang terbatas (skenario dan forum daring), desain potong lintang yang tidak dapat menetapkan sebab-akibat, sampel kecil pada satu-satunya uji acak, dan heterogenitas tinggi pada bukti gabungan. Tidak ada studi dari Indonesia sehingga kesesuaian dengan konteks lokal tidak dapat dinilai langsung.

Tabel 4. Penilaian Kualitas Naratif Studi yang Dimasukkan (n = 14)

Studi (tahun)	Kekuatan utama	Keterbatasan utama	Peringkat
Li dkk. (2022)	Survei lapangan pada pasien rawat jalan nyata; model teoretis mapan (konfirmasi harapan)	172 kuesioner di satu kota; potong lintang; luaran berupa niat, bukan perilaku	Sedang
Aggarwal dkk. (2023)	Tinjauan sistematis 15 studi; efikasi dilaporkan per domain perilaku	Risiko bias sedang–tinggi; kelayakan, penerimaan, dan usabilitas beragam	Sedang
Ayers dkk. (2023)	Penilai tenaga kesehatan berlisensi; 585 penilaian atas 195 pertanyaan	Pertanyaan berasal dari forum publik, bukan setting klinis; tidak mewakili konsultasi nyata	Rendah
Sarbay dkk. (2023)	Perbandingan dengan spesialis pada kasus terstandar ESI	Hanya 50 skenario, bukan pasien nyata; model generasi awal	Rendah

Volume 8 Nomor 11 (2026) 1096 – 1118 P-ISSN 2656-274x E-ISSN 2656-4691
DOI: 10.47476/reslaj.v8i11.13659

Studi (tahun)	Kekuatan utama	Keterbatasan utama	Peringkat
Colakca dkk. (2024)	Pasien IGD nyata yang datang berturut-turut; komite pakar sebagai acuan; ESI 1–2 dilaporkan terpisah	Pengamatan hanya 24 jam; potong lintang; sensitivitas tidak diekstraksi	Sedang
Ding dkk. (2024)	Kerangka evaluasi selaras standar WHO; cakupan sistematis	*Scoping review* tidak menghasilkan estimasi efek dan umumnya tidak menilai mutu studi yang ditinjau	Sedang
Hindelang dkk. (2024)	Tinjauan sistematis (203 rekaman menjadi 18 studi) pada anamnesis	Kualitas studi beragam; angka efisiensi berasal dari studi tunggal	Sedang
Kaboudi dkk. (2024)	Meta-analisis 14 studi (1.412 pasien/skenario)	Heterogenitas sangat tinggi (I^2 = 93%); IK lebar (0,64–0,98); campuran pasien dan skenario	Sedang
Laymouna dkk. (2024)	Cakupan luas peran, pengguna, manfaat, dan keterbatasan	*Rapid review* dengan penyederhanaan metode; data 2017–2023; bersifat deskriptif	Sedang
Lin dkk. (2024)	Meta-analisis 8 uji, 1.073 dewasa; tidak ada bahaya dilaporkan	Luaran berupa proporsi gabungan, bukan selisih terhadap pembanding; sebagian besar intervensi berbasis aturan	Sedang
Bedi dkk. (2025)	Tinjauan sistematis sangat besar (519 studi) yang memetakan praktik evaluasi	Berfokus pada cara evaluasi, bukan estimasi dampak efisiensi atau pengalaman; data sampai Februari 2024	Tinggi
Li dkk. (2025)	394 pasien rawat jalan rumah sakit tersier; regresi berganda	Potong lintang; hubungan asosiatif dan pengguna mungkin berbeda dari nonpengguna; satu negara	Sedang
Wah (2025)	Tinjauan sistematis 29 studi; menyoroti model hibrida	Sebagian sumber berupa prapublikasi dan prosiding; angka dampak sekunder dan sulit diverifikasi	Rendah
Zhang dkk. (2025)	Uji acak 1:1:1; randomisasi terkomputer, penyembunyian alokasi, penilai luaran tersamar	n = 75 di satu pusat; seluruh peserta ASA I; telaah luaran AI retrospektif tanpa pemantauan keselamatan waktu nyata	Sedang

Sumber : Analisis peneliti, 2026

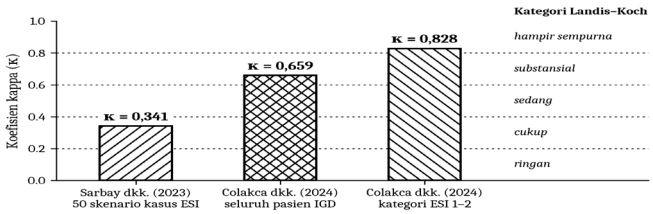
Volume 8 Nomor 11 (2026) 1096 – 1118 P-ISSN 2656-274x E-ISSN 2656-4691
DOI: 10.47476/reslaj.v8i11.13659

Dampak terhadap Efisiensi Pelayanan

Bukti efisiensi terbagi dua kelompok. Pertama, efisiensi alur kerja pada tugas terstruktur. Hindelang dkk. (2024) melaporkan bahwa pengumpulan data otomatis 24 jam meningkatkan efisiensi dan aksesibilitas layanan, dengan salah satu studi yang dirangkum mencatat penghematan waktu 57,3% pada anamnesis alergi sengatan serangga dengan kelengkapan data 73,3%. Laymouna dkk. (2024) mengidentifikasi bantuan administratif dan efektivitas biaya sebagai tema manfaat. Wah (2025) melaporkan penurunan readmisi hingga 25%, peningkatan keterlibatan pasien 30%, dan pemangkasan waktu tunggu konsultasi 15%; angka ini bersumber dari studi yang dirangkum, sebagian berupa prapublikasi dan prosiding, sehingga perlu ditafsirkan hati-hati.

Kedua, efisiensi pengambilan keputusan pada triase. Kaboudi dkk. (2024) melaporkan akurasi gabungan ChatGPT-4 sebesar 0,86 (IK 95% 0,64–0,98) dengan heterogenitas tinggi (I^2 = 93%). Pada 50 skenario berbasis ESI, kesepakatan ChatGPT dengan spesialis hanya tergolong cukup (κ = 0,341), dengan 22% kasus *over-triage* dan 18% *under-triage* (Sarbay dkk., 2023). Studi pada pasien IGD sungguhan menunjukkan kesepakatan lebih baik (κ = 0,659; akurasi 76,6%) dan tinggi untuk kategori gawat darurat ESI 1–2 (κ = 0,828; spesifisitas 95,63%; NPV 98,17%) (Colakca dkk., 2024). Kesepakatan triase disajikan pada Gambar 3.

Menurut kategori Landis dan Koch (1977), κ 0,21–0,40 tergolong cukup, 0,61–0,80 substansial, dan 0,81–1,00 hampir sempurna. Kesepakatan ChatGPT tampak meningkat dari cukup pada studi berbasis skenario (Sarbay dkk., 2023) menjadi substansial pada pasien nyata (Colakca dkk., 2024) dan hampir sempurna untuk ESI 1–2, tetapi perbaikan ini tidak dapat dipastikan sebagai efek generasi model karena desain, populasi, dan standar acuan berbeda. Akurasi keseluruhan 76,6% berarti sekitar seperempat pasien tidak terklasifikasi sesuai standar acuan, sehingga *chatbot* belum layak menjadi penentu tunggal. Sensitivitas tidak diekstraksi dalam tinjauan ini sehingga risiko pasien gawat darurat yang terlewat tidak dapat dinilai penuh.



Sumber : Analisis peneliti berdasarkan Sarbay dkk. (2023) dan Colakca dkk. (2024); kategori menurut Landis dan Koch (1977)

Gambar 3. Kesepakatan Triase ChatGPT dengan Standar Acuan (Koefisien Kappa)

Dampak terhadap Pengalaman Pasien

Lin dkk. (2024) melaporkan proporsi gabungan pasien perioperatif yang puas sebesar 0,73 (IK 95% 0,62–0,85) dan yang merasa memperoleh pengetahuan sebesar

Volume 8 Nomor 11 (2026) 1096 – 1118 P-ISSN 2656-274x E-ISSN 2656-4691
DOI: 10.47476/reslaj.v8i11.13659

0,80 (IK 95% 0,74–0,87) tanpa laporan kerugian. Pada pasien rawat jalan rumah sakit tersier di Tiongkok, pengguna *conversational agent* melaporkan skor pengalaman total lebih tinggi daripada nonpengguna, dengan peningkatan rata-rata 7,51% dalam regresi linear berganda, terutama pada komunikasi dokter–pasien, akses informasi kesehatan, luaran jangka pendek, dan kepuasan umum (Li dkk., 2025). Dalam hal komunikasi, respons ChatGPT dipilih pada 78,6% dari 585 penilaian dibandingkan respons dokter pada forum daring publik, meski bukan pada setting klinis (Ayers dkk., 2023). Namun, pada uji acak edukasi pra-operasi, ChatGPT menurunkan kecemasan tetapi tidak sebesar edukasi oleh urolog (Zhang dkk., 2025). Aggarwal dkk. (2023) melaporkan hasil beragam terkait kelayakan, penerimaan, dan usabilitas, sedangkan Li dkk. (2022) menganalisis niat keberlanjutan penggunaan berbasis konfirmasi harapan pasien. Hindelang dkk. (2024) mencatat keterlibatan dan kepuasan meningkat, tetapi sebagian pasien lebih memilih mengungkapkan informasi sensitif kepada dokter.

Zhang dkk. (2025) merupakan satu-satunya uji acak dalam sintesis ini. Sebanyak 75 pasien bedah urolitiasis dialokasikan 1:1:1 ke edukasi oleh ChatGPT, Ernie Bot, atau urolog, dan kecemasan diukur dengan *State–Trait Anxiety Inventory* (STAI) pada empat titik waktu. Kecemasan turun bermakna pada ketiga kelompok, dengan penurunan terbesar pada kelompok urolog, diikuti ChatGPT, lalu Ernie Bot. ChatGPT unggul dalam empati, keterbacaan, dan kedalaman informasi, serta memicu pertanyaan pasien sekitar dua kali lebih banyak. Tidak ada kejadian tidak diinginkan yang dilaporkan, tetapi tiga kesalahan informasi teridentifikasi pada kelompok ChatGPT.

Temuan Ayers dkk. (2023) perlu dibaca dengan hati-hati karena pertanyaan berasal dari forum publik dan dijawab dokter tanpa hubungan klinis dengan penanya, sedangkan respons yang lebih panjang dan rinci dapat memengaruhi penilaian mutu dan empati. Li dkk. (2022) memakai model konfirmasi harapan (Bhattacherjee, 2001), yang menyatakan bahwa kepuasan dan niat melanjutkan penggunaan bergantung pada sejauh mana harapan pengguna terkonfirmasi oleh pengalaman aktualnya; jadi pengalaman pasien ditentukan pula oleh kesesuaian antara harapan dan kemampuan yang benar-benar ditawarkan *chatbot*. Terakhir, proporsi gabungan kepuasan (0,73; IK 95% 0,62–0,85) dan pengetahuan (0,80; IK 95% 0,74–0,87) tidak dibandingkan langsung dengan layanan konvensional sehingga menggambarkan penerimaan, bukan keunggulan terhadap pelayanan standar.

Hambatan, Kesenjangan Bukti, dan Tata Kelola

Hambatan yang dilaporkan meliputi keamanan data, keterbatasan empati, dan integrasi ke alur kerja klinis (Hindelang dkk., 2024), masalah etika, medikolegal, dan keselamatan (Laymouna dkk., 2024), serta kepercayaan, biaya, dan adaptasi budaya-bahasa (Wah, 2025). Dari sisi bukti, hanya 5% evaluasi model bahasa besar memakai data pasien nyata (Bedi dkk., 2025), dan kerangka evaluasi yang mencakup efektivitas, keselamatan, dan kepuasan pengguna baru tersedia (Ding dkk., 2024). Hambatan tersebut mencakup keselamatan pasien (misalnya *under-triage* 18% pada

Volume 8 Nomor 11 (2026) 1096 – 1118 P-ISSN 2656-274x E-ISSN 2656-4691
DOI: 10.47476/reslaj.v8i11.13659

studi skenario dan tiga kesalahan informasi pada kelompok ChatGPT dalam uji acak; Sarbay dkk., 2023; Zhang dkk., 2025), privasi dan etika, integrasi serta biaya, dan penerimaan. Ringkasan sintesis tematik disajikan pada Tabel 5.

Tabel 5. Sintesis Tematik Temuan Studi

Tema	Ringkasan temuan	Studi pendukung	Kekuatan bukti
Efisiensi alur kerja (anamnesis, administrasi, biaya)	Penghematan waktu dan pengumpulan data 24 jam pada tugas terstruktur; manfaat administratif dan biaya dilaporkan pada tingkat tinjauan	Hindelang dkk. (2024); Laymouna dkk. (2024); Wah (2025)	Sedang; berasal dari tinjauan, data operasional primer terbatas
Efisiensi dan keselamatan triase	Akurasi gabungan tinggi tetapi heterogen; kesepakatan lebih baik pada pasien nyata dan kategori gawat darurat; ada *under-triage*	Kaboudi dkk. (2024); Sarbay dkk. (2023); Colakca dkk. (2024)	Campuran; bergantung desain dan model
Kepuasan, pengetahuan, dan pengalaman pasien	Mayoritas puas dan merasa memperoleh pengetahuan; skor pengalaman pengguna lebih tinggi; respons dinilai berkualitas dan empatik	Lin dkk. (2024); Li dkk. (2025); Ayers dkk. (2023); Aggarwal dkk. (2023)	Sedang, cenderung positif; desain potong lintang dan setting non-klinis
Penerimaan, kepercayaan, dan batas peran	*Chatbot* belum menyamai dokter dalam menurunkan kecemasan; sebagian pasien memilih dokter untuk informasi sensitif; niat keberlanjutan bergantung konfirmasi harapan	Zhang dkk. (2025); Li dkk. (2022); Hindelang dkk. (2024); Wah (2025)	Campuran
Hambatan dan tata kelola evaluasi	Privasi, etika, medikolegal, integrasi, biaya; minim data pasien nyata; kerangka evaluasi baru muncul	Laymouna dkk. (2024); Bedi dkk. (2025); Ding dkk. (2024); Wah (2025)	Konsisten pada tingkat tinjauan

Sumber : Analisis peneliti, 2026

Volume 8 Nomor 11 (2026) 1096 – 1118 P-ISSN 2656-274x E-ISSN 2656-4691
DOI: 10.47476/reslaj.v8i11.13659

Tingkat Kepastian Bukti

Tabel 6 merangkum tingkat kepastian bukti pada setiap kelompok luaran. Tidak ada kelompok luaran yang mencapai kepastian tinggi. Empat kelompok luaran efisiensi dan pengalaman pasien dinilai rendah karena bukti kuantitatifnya bersumber dari studi tunggal atau tinjauan sekunder, desain potong lintang atau berbasis skenario, inkonsistensi (I^2 = 93%), dan sampel kecil pada satu-satunya uji acak. Kepastian untuk hambatan tata kelola dinilai sedang karena konsistensinya lintas tinjauan, meskipun sifatnya deskriptif.

Tabel 6. Tingkat Kepastian Bukti Menurut Kelompok Luaran

Kelompok luaran	Dasar bukti	Arah temuan	Alasan penilaian	Kepastian
Efisiensi alur kerja (waktu, biaya, administrasi)	3 tinjauan (Hindelang; Laymouna; Wah); tidak ada studi primer operasional	Menguntungkan	Data operasional primer tidak ada; angka berasal dari studi tunggal atau sumber non-*peer-reviewed*	Rendah
Akurasi dan keselamatan triase	1 meta-analisis dan 2 studi potong lintang	Campuran	Inkonsistensi (I^2 = 93%); sebagian berbasis skenario; *under-triage* 18% pada satu studi	Rendah
Kepuasan dan pengetahuan pasien	1 meta-analisis (8 uji); 1 studi potong lintang (394 pasien); 1 studi forum; 1 tinjauan	Umumnya menguntungkan	Proporsi gabungan tanpa pembanding langsung; desain asosiatif; setting non-klinis pada satu studi	Rendah
Kecemasan dan empati	1 uji acak (n = 75)	Menguntungkan, di bawah dokter	Sampel kecil; satu pusat; seluruh peserta ASA I	Rendah
Hambatan dan tata kelola	5 tinjauan (Hindelang; Laymouna; Wah; Bedi; Ding)	Konsisten	Deskriptif tetapi konsisten lintas tinjauan; kemungkinan tumpang tindih studi	Sedang

Sumber : Analisis peneliti, 2026

Volume 8 Nomor 11 (2026) 1096 – 1118 P-ISSN 2656-274x E-ISSN 2656-4691
DOI: 10.47476/reslaj.v8i11.13659

Pembahasan

Tinjauan ini menjawab empat pertanyaan penelitian. Terkait RQ1, bukti terpusat pada studi dari Tiongkok dan Türkiye, didominasi desain potong lintang dan tinjauan, dan tidak mencakup Indonesia. Terkait RQ2 dan RQ3, dampak positif terlihat pada tugas terstruktur dan kepuasan pasien, tetapi kepastian buktinya umumnya rendah (Tabel 6). Terkait RQ4, hambatan yang konsisten adalah privasi, empati, integrasi sistem, dan minimnya data pasien nyata. Temuan ini sejalan dengan tinjauan sebelumnya (Hindelang dkk., 2024; Laymouna dkk., 2024; Wah, 2025) dan menambahkan pemisahan bukti primer dari sekunder, penilaian kepastian bukti per luaran, serta penegasan bahwa efisiensi operasional belum diukur langsung oleh studi primer, sejalan dengan kritik Bedi dkk. (2025) dan protokol tinjauan lingkup yang baru terbit (Goossens dkk., 2026).

Pola bukti menunjukkan bahwa AI *chatbot* paling meyakinkan pada tugas yang terstruktur, berulang, dan berbasis pengumpulan informasi. Pada anamnesis, pertanyaan yang sederhana dan sering berupa jawaban ya atau tidak membuat otomatisasi lebih mudah dan mengurangi beban tenaga kesehatan (Hindelang dkk., 2024). Ketersediaan 24 jam juga melepaskan layanan dari keterbatasan jam kerja petugas. Bagi manajer rumah sakit, implikasinya adalah bahwa titik awal penerapan yang paling rasional adalah proses terstruktur seperti pengumpulan riwayat, edukasi pra-tindakan, dan penyaringan informasi awal.

Terdapat paradoks yang perlu dicermati. Bantuan administratif sering dipromosikan sebagai manfaat utama (Laymouna dkk., 2024), tetapi tugas administratif nyaris tidak dievaluasi dalam penelitian model bahasa besar (Bedi dkk., 2025). Angka penurunan readmisi, waktu tunggu, dan keterlibatan (Wah, 2025) pun berasal dari sintesis sekunder yang kualitas sumbernya beragam. Dengan demikian, rumah sakit tidak dapat mengasumsikan bahwa manfaat serupa akan terjadi di institusinya; uji coba terbatas dengan data dasar sendiri, misalnya waktu tunggu, waktu proses pendaftaran atau anamnesis, dan tingkat eskalasi ke petugas, diperlukan sebelum ekspansi.

Perbedaan hasil pada triase tampaknya dipengaruhi desain studi dan generasi model. Studi awal berbasis skenario menunjukkan kesepakatan yang cukup dengan proporsi kesalahan klasifikasi 40% (Sarbay dkk., 2023), sedangkan studi pada pasien nyata dan tinjauan yang menilai ChatGPT-4 melaporkan kinerja lebih baik (Colakca dkk., 2024; Kaboudi dkk., 2024). Namun, heterogenitas yang tinggi dan adanya *under-triage* menegaskan bahwa kesalahan pada tugas ini berdampak langsung pada keselamatan pasien. Kinerja yang tinggi pada kategori gawat darurat menunjukkan peran pendukung, misalnya penanda awal pasien berisiko tinggi, dengan verifikasi manusia sebagai syarat mutlak dan bukan sebagai pengganti perawat triase.

Pada sisi pengalaman pasien, kepuasan umumnya tinggi tetapi bergantung pada konteks. Proporsi puas 73% pada pasien perioperatif berarti sekitar seperempat pasien tidak puas, dan sebagian besar intervensinya berbasis aturan (Lin dkk., 2024). Skor pengalaman yang lebih tinggi pada pengguna rawat jalan (Li dkk.,

Volume 8 Nomor 11 (2026) 1096 – 1118 P-ISSN 2656-274x E-ISSN 2656-4691
DOI: 10.47476/reslaj.v8i11.13659

2025) bersifat asosiatif; pengguna mungkin berbeda dari nonpengguna sehingga hubungan sebab-akibat tidak dapat ditegakkan. Keunggulan respons ChatGPT pada forum daring (Ayers dkk., 2023) tidak otomatis berlaku dalam pelayanan klinis yang bersifat pribadi. Uji acak menunjukkan bahwa edukasi oleh dokter tetap lebih unggul dalam menurunkan kecemasan meskipun kepuasan terhadap ChatGPT tinggi (Zhang dkk., 2025), dan pasien cenderung memilih dokter untuk informasi sensitif (Hindelang dkk., 2024). Temuan ini mendukung model komplementer, yaitu *chatbot* menangani informasi rutin sementara tenaga kesehatan menangani percakapan yang membutuhkan empati dan penilaian klinis, sejalan dengan gagasan *chatbot* hibrida (Wah, 2025).

Penerimaan jangka panjang juga bergantung pada pengelolaan harapan pasien. Li dkk. (2022) menempatkan konfirmasi harapan sebagai dasar niat keberlanjutan penggunaan, sehingga rumah sakit perlu menyampaikan dengan jelas apa yang dapat dan tidak dapat dilakukan *chatbot*, menyediakan jalur eskalasi ke petugas, dan mengumpulkan umpan balik pengguna. Wah (2025) menyoroti hambatan budaya dan bahasa, yang relevan bagi Indonesia dengan keragaman bahasa dan tingkat literasi digital, meskipun tidak ada studi Indonesia dalam sintesis ini. Arah kebijakan transformasi digital yang berorientasi pelayanan (Kementerian Kesehatan Republik Indonesia, 2021) sejalan dengan pemanfaatan teknologi percakapan, tetapi memerlukan studi lokal untuk memastikan kesesuaian. Kesiapan sumber daya manusia juga menentukan: tinjauan tentang penerapan AI dalam manajemen sumber daya manusia rumah sakit menunjukkan bahwa manfaat AI terhadap kinerja pegawai dan efisiensi operasional bergantung pada infrastruktur teknologi, kompetensi digital, dan budaya kerja yang adaptif (Kosasih dkk., 2026b).

Tata kelola evaluasi menjadi syarat penerapan. Kerangka yang menilai efektivitas, keselamatan, dan kepuasan pengguna (Ding dkk., 2024) dan seruan agar evaluasi memakai data pasien nyata (Bedi dkk., 2025) dapat diterjemahkan menjadi papan indikator rumah sakit yang memuat waktu tunggu, beban kerja petugas, akurasi dan tingkat eskalasi, serta skor pengalaman pasien. Isu privasi, etika, dan medikolegal (Hindelang dkk., 2024; Laymouna dkk., 2024) perlu ditangani melalui kebijakan persetujuan pasien, perlindungan data, dan pencatatan audit. Integrasi chatbot dengan rekam medis elektronik juga perlu direncanakan sejak awal karena efisiensi pelayanan digital rumah sakit sangat dipengaruhi oleh interoperabilitas sistem dan pelatihan staf (Kosasih dkk., 2026a).

Implikasi bagi Manajemen Rumah Sakit

Bagi manajer rumah sakit, temuan di atas mengarah pada penerapan bertahap. Tabel 7 menyusun empat tahap beserta indikator pemantauan yang dapat diukur dari data rutin rumah sakit dan menutup kesenjangan bukti, terutama pengukuran waktu tunggu, waktu proses, dan biaya.

Volume 8 Nomor 11 (2026) 1096 – 1118 P-ISSN 2656-274x E-ISSN 2656-4691
DOI: 10.47476/reslaj.v8i11.13659

Tabel 7. Tahapan Penerapan AI *Chatbot* dan Indikator Pemantauan bagi Rumah Sakit

Tahap	Cakupan	Indikator pemantauan	Kontrol keselamatan dan tata kelola
1. Persiapan dan tata kelola	Menetapkan tujuan, peran *chatbot*, penanggung jawab klinis, kebijakan persetujuan, dan pelindungan data	Data dasar: waktu tunggu, waktu pendaftaran atau anamnesis, beban kerja petugas, skor pengalaman pasien	Kajian kepatuhan terhadap UU PDP; pertimbangan penilaian dampak pelindungan data pribadi; pencatatan audit
2. Uji coba proses terstruktur	Anamnesis, edukasi pra-tindakan, dan informasi administratif pada satu unit	Waktu proses, kelengkapan data, tingkat eskalasi ke petugas, kepuasan dan pengetahuan pasien	Jalur eskalasi ke petugas; telaah sampel percakapan; pemberitahuan bahwa pasien berinteraksi dengan AI
3. Pendukung triase (bila dipertimbangkan)	*Chatbot* memberi penanda risiko; keputusan tetap oleh perawat triase	Kesepakatan dengan keputusan perawat (κ), *over-triage* dan *under-triage*, kinerja pada ESI 1–2	Verifikasi manusia wajib; tidak menjadi penentu tunggal; audit setiap kasus *under-triage*
4. Evaluasi dan keputusan ekspansi	Membandingkan kondisi sebelum dan sesudah; memutuskan lanjut, ubah, atau hentikan	Perubahan indikator terhadap data dasar; keluhan dan insiden keselamatan; biaya	Tinjauan berkala oleh komite mutu; umpan balik pasien; dokumentasi hasil

Sumber : Disusun peneliti berdasarkan sintesis pada Tabel 3–6, 2026; belum divalidasi secara empiris

Konteks Indonesia dan Kerangka Regulasi

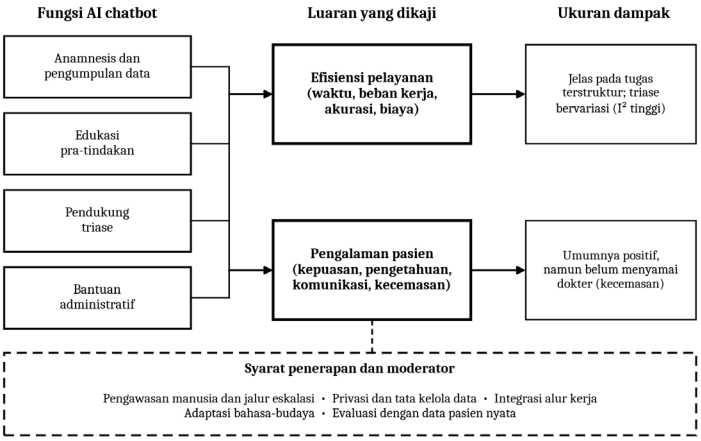
Penerapan *chatbot* di rumah sakit Indonesia perlu selaras dengan kerangka regulasi yang berlaku. Data dan informasi kesehatan tergolong data pribadi yang bersifat spesifik menurut Undang-Undang Nomor 27 Tahun 2022 tentang Pelindungan Data Pribadi (Republik Indonesia, 2022), dengan masa penyesuaian bagi pengendali data yang berakhir pada 17 Oktober 2024. Rumah sakit umumnya berperan sebagai pengendali data pribadi, sedangkan penyedia teknologi dapat menjadi prosesor data. Integrasi dengan rekam medis perlu memperhatikan Peraturan Menteri Kesehatan Nomor 24 Tahun 2022 tentang Rekam Medis

Volume 8 Nomor 11 (2026) 1096 – 1118 P-ISSN 2656-274x E-ISSN 2656-4691
DOI: 10.47476/reslaj.v8i11.13659

(Kementerian Kesehatan Republik Indonesia, 2022b), sedangkan Undang-Undang Nomor 17 Tahun 2023 tentang Kesehatan memuat pengaturan tentang sistem informasi dan teknologi kesehatan (Republik Indonesia, 2023). Pada tataran etika, panduan *World Health Organization* menekankan transparansi, akuntabilitas, dan inklusivitas dalam penggunaan AI untuk kesehatan (World Health Organization, 2021, 2024). Bagi rumah sakit, prinsip tersebut dapat diterjemahkan menjadi pemberitahuan kepada pasien bahwa mereka berinteraksi dengan AI, penetapan tanggung jawab klinis atas keluaran *chatbot*, dan mekanisme pengaduan.

Kerangka Konseptual Hasil Sintesis

Gambar 4 merangkum hubungan antara fungsi *chatbot*, luaran yang dikaji, dan syarat penerapan. Arah dan besar dampaknya bergantung pada pengawasan manusia dengan jalur eskalasi, perlindungan privasi, integrasi ke alur kerja, adaptasi bahasa dan budaya, serta evaluasi dengan data pasien nyata. Kerangka ini disusun peneliti sebagai ringkasan sintesis, bukan model teoretis yang telah divalidasi secara empiris.



Sumber : Analisis peneliti berdasarkan sintesis studi, 2026

Gambar 4. Kerangka Konseptual Hasil Sintesis

KESIMPULAN

Bukti terbitan 2021–2026 menunjukkan bahwa AI *chatbot* berpotensi meningkatkan efisiensi pelayanan, terutama pada tugas terstruktur seperti anamnesis, edukasi pra-operasi, dan pendukung triase, serta memperbaiki pengalaman pasien berupa kepuasan, akses informasi, dan komunikasi. Namun, kinerjanya belum menggantikan tenaga kesehatan, akurasi triase masih bervariasi, dan sebagian besar bukti berasal dari studi potong lintang atau tinjauan sekunder dengan data pasien nyata yang terbatas. Tingkat kepastian bukti pada seluruh kelompok luaran efisiensi dan pengalaman pasien dinilai rendah, dan tidak ada studi primer yang mengukur waktu tunggu atau biaya secara langsung.

Volume 8 Nomor 11 (2026) 1096 – 1118 P-ISSN 2656-274x E-ISSN 2656-4691
DOI: 10.47476/reslaj.v8i11.13659

DAFTAR PUSTAKA

Adamopoulou E. & Moussiades L. (2020). Chatbots: history, technology, and applications. *Machine Learning with Applications*, 2:100006. doi:10.1016/j.mlwa.2020.100006

Aggarwal A., Tam C.C., Wu D. dkk. (2023). Artificial intelligence–based chatbots for promoting health behavioral changes: systematic review. *Journal of Medical Internet Research*, 25:e40789. doi:10.2196/40789

Amiri P. & Karahanna E. (2022). Chatbot use cases in the COVID-19 public health response. *Journal of the American Medical Informatics Association*, 29(5):1000-1010. doi:10.1093/jamia/ocac014

Ayers J.W., Poliak A., Dredze M. dkk. (2023). Comparing physician and artificial intelligence chatbot responses to patient questions posted to a public social media forum. *JAMA Internal Medicine*, 183(6):589-596. doi:10.1001/jamainternmed.2023.1838

Bedi S., Liu Y., Orr-Ewing L. dkk. (2025). Testing and evaluation of health care applications of large language models: a systematic review. *JAMA*, 333(4):319-328. doi:10.1001/jama.2024.21700

Bhattacherjee A. (2001). Understanding information systems continuance: an expectation-confirmation model. *MIS Quarterly*, 25(3):351-370. doi:10.2307/3250921

Braun V. & Clarke V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2):77-101. doi:10.1191/1478088706qp063oa

Campbell M., McKenzie J.E., Sowden A. dkk. (2020). Synthesis without meta-analysis (SWiM) in systematic reviews: reporting guideline. *BMJ*, 368:l6890. doi:10.1136/bmj.l6890

Colakca C., Ergın M., Ozensoy H.S. dkk. (2024). Emergency department triaging using ChatGPT based on emergency severity index principles: a cross-sectional study. *Scientific Reports*, 14(1):1-7. doi:10.1038/s41598-024-73229-7

Ding H., Simmich J., Vaezipour A. dkk. (2024). Evaluation framework for conversational agents with artificial intelligence in health interventions: a systematic scoping review. *Journal of the American Medical Informatics Association*, 31(3):746-761. doi:10.1093/jamia/ocad222

Donabedian A. (1988). The quality of care: how can it be assessed? *JAMA*, 260(12):1743-1748. doi:10.1001/jama.1988.03410120089033

Goossens E., Mfouth Kemajou P., Coorevits P. dkk. (2026). Clinical outcomes, patient satisfaction and operational efficiency of AI-powered chatbots in medicine and healthcare: protocol for an AI-aided scoping review. *BMJ Open*, 16(6):e113547.

Guyatt G.H., Oxman A.D., Vist G.E. dkk. (2008). GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ*, 336(7650):924-926. doi:10.1136/bmj.39489.470347.AD

Volume 8 Nomor 11 (2026) 1096 – 1118 P-ISSN 2656-274x E-ISSN 2656-4691
DOI: 10.47476/reslaj.v8i11.13659

Higgins J.P.T., Thomas J., Chandler J. dkk. (Eds.). (2023). *Cochrane Handbook for Systematic Reviews of Interventions* version 6.4. Cochrane.

Hindelang M., Sitaru S. & Zink A. (2024). Transforming health care through chatbots for medical history-taking and future directions: comprehensive systematic review. *JMIR Medical Informatics*, 12:e56628. doi:10.2196/56628

Institute of Medicine. (2001). *Crossing the Quality Chasm: A New Health System for the 21st Century*. Washington, DC: National Academy Press.

Kaboudi N., Firouzbakht S., Shahir Eftekhar M. dkk. (2024). Diagnostic accuracy of ChatGPT for patients' triage; a systematic review and meta-analysis. *Archives of Academic Emergency Medicine*, 12(1):e60. doi:10.22037/aaem.v12i1.2384

Kementerian Kesehatan Republik Indonesia. (2021). *Cetak Biru Strategi Transformasi Digital Kesehatan 2024*. Jakarta: Kementerian Kesehatan Republik Indonesia.

Kementerian Kesehatan Republik Indonesia. (2022a). *Keputusan Menteri Kesehatan Nomor HK.01.07/MENKES/1559/2022 tentang Penerapan Sistem Pemerintahan Berbasis Elektronik Bidang Kesehatan dan Strategi Transformasi Digital Kesehatan*. Jakarta: Kementerian Kesehatan Republik Indonesia.

Kementerian Kesehatan Republik Indonesia. (2022b). *Peraturan Menteri Kesehatan Nomor 24 Tahun 2022 tentang Rekam Medis*. Jakarta: Kementerian Kesehatan Republik Indonesia.

Kosasih K., Septiani D., Anggraini W. dkk. (2026a). Analysis of the efficiency of Hospital X services based on electronic medical records: systematic literature review, PRISMA, and VOSviewer. *Advances in Human Resource Management Research*, 4(1):147-163. doi:10.60079/ahrmr.v4i1.790

Kosasih K., Yesika M., Lustina F.A. dkk. (2026b). Artificial intelligence in hospital human resource management: a systematic review and bibliometric analysis. *Advances in Human Resource Management Research*, 4(2):267-285. doi:10.60079/ahrmr.v4i2.776

Landis J.R. & Koch G.G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33(1):159-174. doi:10.2307/2529310

Laymouna M., Ma Y., Lessard D. dkk. (2024). Roles, users, benefits, and limitations of chatbots in health care: rapid review. *Journal of Medical Internet Research*, 26:e56930. doi:10.2196/56930

Li D., Zhang H., Lu C. dkk. (2025). Use of artificial intelligence–assisted conversational agents to improve patient experience related to physicians: cross-sectional study in China. *Journal of Medical Internet Research*, 27:e76540. doi:10.2196/76540

Li X., Xie S., Ye Z. dkk. (2022). Investigating patients' continuance intention toward conversational agents in outpatient departments: cross-sectional field survey. *Journal of Medical Internet Research*, 24(11):e40681. doi:10.2196/40681

Volume 8 Nomor 11 (2026) 1096 – 1118 P-ISSN 2656-274x E-ISSN 2656-4691
DOI: 10.47476/reslaj.v8i11.13659

Lin S., Sun C., Chen D. dkk. (2024). Perioperative application of chatbots: a systematic review and meta-analysis. *BMJ Health & Care Informatics*, 31(1):e100985. doi:10.1136/bmjhci-2023-100985

Moola S., Munn Z., Tufanaru C. dkk. (2020). Chapter 7: systematic reviews of etiology and risk. Dalam: Aromataris E. & Munn Z. (Eds.), *JBI Manual for Evidence Synthesis*. JBI. doi:10.46658/JBIMES-20-08

Page M.J., McKenzie J.E., Bossuyt P.M. dkk. (2021). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ*, 372:n71. doi:10.1136/bmj.n71

Pieper D., Antoine S.L., Mathes T. dkk. (2014). Systematic review finds overlapping reviews were not mentioned in every other overview. *Journal of Clinical Epidemiology*, 67(4):368-375. doi:10.1016/j.jclinepi.2013.11.007

Popay J., Roberts H., Sowden A. dkk. (2006). *Guidance on the Conduct of Narrative Synthesis in Systematic Reviews: A Product from the ESRC Methods Programme*. Lancaster: Lancaster University.

Republik Indonesia. (2022). *Undang-Undang Nomor 27 Tahun 2022 tentang Pelindungan Data Pribadi*. Jakarta: Sekretariat Negara.

Republik Indonesia. (2023). *Undang-Undang Nomor 17 Tahun 2023 tentang Kesehatan*. Jakarta: Sekretariat Negara.

Sarbay İ., Berikol G.B. & Özturan İ.U. (2023). Performance of emergency triage prediction of an open access natural language processing based chatbot application (ChatGPT): a preliminary, scenario-based cross-sectional study. *Turkish Journal of Emergency Medicine*, 23(3):156-161. doi:10.4103/tjem.tjem_79_23

Shea B.J., Reeves B.C., Wells G. dkk. (2017). AMSTAR 2: a critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. *BMJ*, 358:j4008. doi:10.1136/bmj.j4008

Sterne J.A.C., Savović J., Page M.J. dkk. (2019). RoB 2: a revised tool for assessing risk of bias in randomised trials. *BMJ*, 366:l4898. doi:10.1136/bmj.l4898

Wah J.N.K. (2025). Revolutionizing e-health: the transformative role of AI-powered hybrid chatbots in healthcare solutions. *Frontiers in Public Health*, 13:1530799. doi:10.3389/fpubh.2025.1530799

Wolf J.A., Niederhauser V., Marshburn D. & LaVela S.L. (2014). Defining patient experience. *Patient Experience Journal*, 1(1):7-19.

World Health Organization. (2021). *Ethics and Governance of Artificial Intelligence for Health: WHO Guidance*. Geneva: World Health Organization.

World Health Organization. (2024). *Ethics and Governance of Artificial Intelligence for Health: Guidance on Large Multi-Modal Models*. Geneva: World Health Organization.

Xu L., Sanders L., Li K. dkk. (2021). Chatbot for health care and oncology applications using artificial intelligence and machine learning: systematic review. *JMIR Cancer*, 7(4):e27850. doi:10.2196/27850

Volume 8 Nomor 11 (2026) 1096 – 1118 P-ISSN 2656-274x E-ISSN 2656-4691
DOI: 10.47476/reslaj.v8i11.13659

Zhang H., Wang X., Luo H. dkk. (2025). Comparison of preoperative education by artificial intelligence versus traditional physicians in perioperative management of urolithiasis surgery: a prospective single-blind randomized controlled trial conducted in China. *Frontiers in Medicine*, 12:1543630. doi:10.3389/fmed.2025.1543630