Vol 2 No. 4 Desember 2025 P-ISSN : 3047-3233 E-ISSN : 3047-9819, Hal 190 – 206

JURNAL RISET TEKNIK KOMPUTER
Halaman Jurnal: https://journal.smartpublisher.id/index.php/jurtikom
Halaman UTAMA Jurnal : https://journal.smartpublisher.id

DOI: https://doi.org/10.69714/gvj6z903

EXPLAINABLE AI PADA CHATBOT MODERN : REVIEW ALGORITMA INTERPRETABILITAS UNTUK MENJELASKAN PROSES PENGAMBILAN KEPUTUSAN MODEL

David Mahlon Sarumaha [a*], Putu Pramudya Pratama [b], Ganesha Setiawan [c]
[a] Ilmu Komputer / Sistem Informasi; sarumahadavid29@gmail.com, Universitas Pembangunan Nasional "Veteran" Jawa Timur; Jl. Raya Rungkut Madya, Gunung Anyar, Surabaya, Jawa Timur
[a] Ilmu Komputer / Sistem Informasi; pratamaniu@gmail.com, Universitas Pembangunan Nasional "Veteran" Jawa Timur; Jl. Raya Rungkut Madya, Gunung Anyar, Surabaya, Jawa Timur
[a] Ilmu Komputer / Sistem Informasi; ganeshasetiawan92@yahoo.com, Universitas Pembangunan Nasional "Veteran" Jawa Timur; Jl. Raya Rungkut Madya, Gunung Anyar, Surabaya, Jawa Timur
[*] Penulis Korespondensi: David Mahlon Sarumaha

ABSTRACT

Large Language Models (LLMs) have transformed modern chatbots into systems capable of natural interaction and understanding conversational context across various domains, such as education, healthcare, and customer service. However, this advancement presents a major challenge in the form of a "black box" nature that conceals the model's decision-making logic. This condition hinders user trust, complicates the debugging process, and raises ethical concerns in high-risk domains such as medical diagnosis and financial consulting. Explainable AI (xAI) has emerged as an approach to make AI decision-making processes more transparent and comprehensible to both developers and end-users. This study conducts a systematic review of 103 recent studies (2020-2025) to map xAI techniques applied to modern chatbots. The analysis reveals that technical methods such as attention mechanisms and feature importance analysis dominate xAI implementation, with an emerging trend toward the use of natural language explanations for end-users. The main contributions of this research include identifying the trade-off between model performance and interpretability, the need for standardized evaluation metrics, and the limited ecological validity of research conducted primarily in controlled laboratory settings. This review emphasizes that xAI is a fundamental requirement—not merely an additional feature—for building responsible and trustworthy conversational AI systems. The study also proposes future research directions, namely the development of domain-specific xAI frameworks, cross-cultural studies, and the formulation of robust ethical guidelines to ensure that AI benefits can be achieved without compromising accountability and user autonomy.

Keywords: *Explainable AI; Chatbot; Interpretability; Transparency; Large Language Models*

Abstrak

Model bahasa besar (Large Language Models/LLMs) telah mengubah chatbot modern menjadi sistem yang mampu berinteraksi secara alami dan memahami konteks percakapan di berbagai bidang, seperti pendidikan, kesehatan, dan layanan pelanggan. Namun, kemajuan ini menghadirkan tantangan utama berupa sifat "kotak hitam" yang menyembunyikan logika pengambilan keputusan model. Kondisi ini menghambat kepercayaan pengguna, mempersulit proses debugging, dan menimbulkan permasalahan etika pada domain berisiko tinggi seperti diagnosis medis dan konsultasi finansial. Explainable AI (xAI) muncul sebagai pendekatan untuk menjadikan proses pengambilan keputusan AI lebih transparan dan dapat dipahami oleh pengembang maupun pengguna akhir. Penelitian ini melakukan tinjauan sistematis terhadap 103 studi terkini (2020-2025) untuk memetakan teknik xAI yang diterapkan pada chatbot modern. Hasil analisis menunjukkan bahwa metode teknis seperti attention mechanism dan feature importance analysis mendominasi implementasi xAI, dengan tren berkembang pada penggunaan penjelasan berbahasa alami untuk pengguna akhir. Kontribusi utama penelitian ini meliputi identifikasi trade-off antara kinerja model dan interpretabilitas, kebutuhan akan standar metrik evaluasi, serta keterbatasan validitas ekologis penelitian yang sebagian besar dilakukan dalam pengaturan laboratorium terkontrol. Tinjauan ini

Naskah Masuk 23 Desember 2025; Revisi 8 Januari 2026; Diterima 31 Januari 2026; Terbit 7 Februari 2026

menekankan bahwa xAI merupakan kebutuhan mendasar bukan sekadar fitur tambahan untuk membangun sistem AI percakapan yang bertanggung jawab dan dapat dipercaya. Penelitian ini juga mengusulkan arah penelitian masa depan, yaitu pengembangan framework xAI yang spesifik untuk setiap domain, pelaksanaan studi lintas budaya, serta penyusunan pedoman etis yang kuat guna memastikan manfaat AI dapat dicapai tanpa mengorbankan akuntabilitas dan otonomi pengguna.

Kata Kunci: Explainable AI; Chatbot; Interpretabilitas; Transparansi; model Bahasa besar

1. PENDAHULUAN

Dalam sepuluh tahun terakhir, kecerdasan buatan (AI), terutama dalam bentuk chatbot berbasis Large Language Models (LLMs), telah menjadi komponen penting dari interaksi manusia dengan teknologi. Batasan chatbot generasi sebelumnya yang berbasis aturan sederhana telah diatasi oleh kemampuan terbaru mereka untuk memahami konteks percakapan yang kompleks dan memberikan tanggapan yang jelas dan alami [106]. Saat ini, sistem-sistem ini terintegrasi dalam berbagai aspek kehidupan, seperti layanan pelanggan di industri perbankan [3], pendampingan kesehatan mental [13][14], dan asisten pembelajaran dalam pendidikan [11][38]. Pada masa pandemi COVID-19 dan akselerasi digital yang mengikutinya, adopsi chatbot meningkat pesat di berbagai industri, menunjukkan solusi yang sangat bermanfaat bagi organisasi dan individu karena kemampuan mereka untuk menangani banyak permintaan secara bersamaan sepanjang hari. Meskipun peningkatan kemampuan teknis ini membawa manfaat signifikan, kemajuan ini juga menghadirkan tantangan fundamental: proses pengambilan keputusan semakin sulit dipahami seiring dengan kecanggihan sistem AI. Arsitektur Transformer, yang menjadi dasar LLMs, menciptakan sistem dengan miliaran parameter yang sangat sulit untuk ditafsirkan, bahkan oleh para peneliti AI [106]. Kenyataan "kotak hitam" ini menimbulkan pertanyaan penting tentang akuntabilitas, kepercayaan, dan keamanan sistem AI yang semakin otonom. Chatbot dalam bidang kesehatan, misalnya, dapat membahayakan keselamatan pasien dan menimbulkan masalah kewajiban hukum [7][17][39]. Selain itu, dalam industri keuangan, ketidakmampuan untuk menjelaskan alasan rekomendasi chatbot investasi dapat menyebabkan masalah hukum dan kehilangan kepercayaan pelanggan [3][88].

Explainable AI (xAI) telah muncul sebagai pendekatan baru untuk memecahkan masalah ini. xAI bertujuan untuk menjembatani kesenjangan antara performa model yang tinggi dan pemahaman terhadap cara kerja sistem. Ini tidak bertujuan untuk menggantikan model kompleks, tetapi untuk melengkapinya dengan alat-alat yang dapat menunjukkan perilaku dan keputusan model [106]. XAI telah menjadi kebutuhan vital untuk memastikan sistem yang bertanggung jawab dan dapat dipercaya dalam hal chatbot, dan bukan lagi sekadar fitur tambahan [2]. Dalam hal chatbot, transparansi adalah faktor terkuat yang menentukan minat pengguna [23] [43]. Sistem yang dapat menjelaskan cara kerjanya dan mengapa mereka memberikan respons tertentu cenderung membuat pengguna lebih percaya dan bersedia menggunakannya.

Meskipun penelitian tentang xAI dan chatbot telah berkembang pesat, terdapat kesenjangan signifikan dalam literatur yang mengintegrasikan kedua domain ini secara sistematis. Penelitian sebelumnya cenderung fokus pada aspek teknis xAI dalam konteks umum seperti computer vision atau sistem prediksi, tanpa mempertimbangkan karakteristik unik dari interaksi percakapan berbasis teks. Sebaliknya, studi tentang kesehatan, chatbot modern sering kali tidak mengeksplorasi kebutuhan interpretabilitas secara mendalam, terutama dalam era LLMs yang sangat kompleks. Selain itu, belum ada kerangka teoritis yang komprehensif untuk memahami bagaimana xAI dapat diimplementasikan pada chatbot dengan mempertimbangkan trade-off antara performa, interpretabilitas, dan pengalaman pengguna di berbagai domain aplikasi. Berbeda dengan tinjauan literatur sebelumnya, penelitian ini secara khusumengeksplorasi penerapan xAI pada chatbot berbasis LLMs yang menjadi standar industri saat ini, dengan fokus pada periode 2020-2025 ketika adopsi teknologi ini mengalami lonjakan signifikan. Penelitian ini tidak hanya mengidentifikasi teknik xAI yang digunakan, tetapi juga menganalisis efektivitasnya dalam konteks aplikasi nyata serta tantangan implementasi di berbagai domain berisiko tinggi.

Secara khusus, tinjauan sistematis ini bertujuan untuk: (1) mengidentifikasi dan mengklasifikasikan berbagai metode dan teknik xAI yang telah diterapkan pada chatbot modern berbasis LLMs; (2) mengevaluasi efektivitas berbagai pendekatan xAI dalam meningkatkan transparansi, kepercayaan pengguna, dan kinerja sistem; (3) menganalisis praktik terbaik dan prinsip desain untuk mengembangkan

chatbot yang dapat dijelaskan tanpa mengorbankan performa; dan (4) mengeksplorasi pertimbangan etis serta implikasi regulasi dari penerapan xAI pada chatbot. Ruang lingkup kajian ini mencakup 103 artikel yang telah melalui peer-review dari publikasi periode 2020-2025, dengan fokus pada penelitian empiris dan studi kasus yang mendemonstrasikan implementasi xAI pada chatbot di berbagai domain aplikasi, termasuk pendidikan, layanan pelanggan, dan keuangan. Kontribusi utama dari tinjauan ini meliputi: (1) taksonomi terstruktur dari teknik xAI yang diterapkan pada chatbot modern; (2) analisis komprehensif mengenai trade-off antara performa model dan interpretabilitas dalam konteks aplikasi nyata; (3) identifikasi kesenjangan penelitian dan tantangan metodologis yang belum terpecahkan; serta (4) rekomendasi berbasis bukti untuk praktisi, peneliti, dan pembuat kebijakan dalam mengembangkan chatbot yang explainable, etis, dan dapat dipercaya.

2. TINJAUAN PUSTAKA

2.1 Evolusi Chatbot dan Munculnya Masalah “Kotak Hitam”

Perkembangan chatbot telah mengalami sejumlah tahap perubahan penting. Pada awalnya, chatbot didasarkan pada model sederhana yang berfokus pada aturan dan sepenuhnya transparan, seperti ELIZA yang dikembangkan pada tahun 1966 [referensi ELIZA]. Model ini beroperasi dengan cara mencocokkan kata kunci dari input pengguna menggunakan respons yang telah ditentukan sebelumnya [110]. Meskipun mudah untuk dipahami dan diprediksi, chatbot berbasis aturan memiliki keterbatasan besar dalam hal kemampuan karena tidak dapat belajar dari pengalaman atau menangani variasi bahasa yang rumit [109]. Selanjutnya, muncul era Pembelajaran Mesin yang membawa perubahan besar dengan memberikan kesempatan bagi chatbot untuk belajar dari data dan mengenali pola dalam percakapan. Tahap ini terus berkembang dan mencapai titik tertingginya dengan penerapan model Pembelajaran Mendalam, terutama arsitektur Transformer yang menghasilkan Model Bahasa Besar (LLMs) seperti GPT, LaMDA, dan BERT. Arsitektur Transformer telah mengubah cara pemrosesan bahasa alami dengan mekanisme perhatian yang memungkinkan model untuk memahami konteks dengan lebih mendalam [111]. Model-model ini mampu menghasilkan respons yang sangat koheren dan kontekstual, bahkan membuat batasan antara mesin dan manusia dalam percakapan menjadi tidak jelas [109]. Namun, tingkat kompleksitas chatbot generatif berbanding terbalik dengan transparansinya, menciptakan fenomena "kotak hitam" yang menghasilkan sejumlah masalah mendasar [108]. Pertama, muncul pertanyaan mengenai akuntabilitas: siapa yang bertanggung jawab jika chatbot memberikan informasi yang salah, bias, atau bahkan berbahaya? Penelitian terbaru menunjukkan bahwa kerumitan model AI masa kini telah menciptakan kebutuhan mendesak akan explainability dalam sistem chatbot berbasis LLMs [109][111]. Kedua, model AI dapat mengadopsi bias yang terdapat dalam data latihannya. Tanpa kemampuan untuk menyelidiki logika di balik respons, bias ini dapat terus diperkuat secara tidak terdeteksi dan memiliki potensi untuk merugikan kelompok tertentu [110]. Ketiga, masalah kepercayaan pengguna muncul karena pengguna mungkin tidak nyaman untuk bergantung pada sistem yang tidak dapat mereka pahami [108]. Transparansi merupakan dasar yang sangat penting untuk membangun kepercayaan pengguna terhadap teknologi AI. Keempat, dari perspektif pengembangan sistem, ketidakmampuan untuk melacak alasan di balik respons tertentu menyulitkan proses debugging, perbaikan, dan pengembangan model [111]. Evaluasi kinerja chatbot memerlukan analisis algoritma yang terstruktur dan interpretable agar dapat menjamin kualitas respons yang diberikan [109]

2.2. Konsep Dasar Explainable AI(xAI)

AI yang dapat dijelaskan (xAI) bertujuan untuk menghubungkan performa tinggi dengan pemahaman sistem. Strategi ini tidak dimaksudkan untuk menghilangkan model-model rumit yang sudah ada, tetapi untuk memperkaya mereka dengan alat yang dapat menawarkan penjelasan mengenai perilaku dan keputusan yang dihasilkan [111]. Penjelasan ini bisa ditujukan kepada dua kelompok utama: pengembang, untuk kebutuhan debugging dan validasi model; dan pengguna akhir, untuk membangun kepercayaan dan pemahaman terhadap sistem [108]. Dalam konteks chatbot, penjelasan untuk pengguna akhir menjadi sangat penting karena berhubungan langsung dengan pengalaman pengguna dan tingkat adopsi teknologi [110]. Penerapan xAI pada chatbot modern berfokus pada sejumlah pertanyaan penting yang sering muncul dalam interaksi. Pertanyaan seperti "Mengapa chatbot memberikan respons A dan bukan B?", "Fitur atau istilah kunci mana dari input pengguna yang paling berpengaruh pada respons?", atau "Apakah respons ini didasarkan pada fakta atau hasil halusinasi?" menjadi fokus utama [109]. Teknik xAI berupaya memberikan jawaban atas pertanyaan-

pertanyaan ini dengan berbagai pendekatan inovatif. Beberapa metode yang biasa digunakan meliputi feature attribution yang menyoroti elemen penting dari input pengguna [Wang et al., 2024][file:3], self-explanations yang memberikan rationale dalam bahasa yang mudah dipahami [Wang et al., 2024][file:3], atau counterfactual explanations yang menyediakan alternatif respons untuk membantu pemahaman [111]. Pendekatan terbaru juga memanfaatkan Retrieval-Augmented Generation (RAG) systems yang mengintegrasikan domain-specific knowledge bases dengan Large Language Models untuk menghasilkan penjelasan yang lebih contextual dan interactive [111]. Selain itu, framework hybrid yang menggabungkan ChatGPT dengan knowledge graphs telah terbukti efektif dalam meningkatkan post-hoc explainability dengan menyediakan background knowledge yang relevan [108]. Pentingnya penerapan xAI semakin nyata di domain-domain kritis seperti kesehatan, keuangan, dan manufaktur. Di sektor kesehatan, chatbot yang mampu menjelaskan alasan di balik responnya sangat penting untuk membantu pasien memahami kondisi kesehatan mereka, meningkatkan kepercayaan dan kepatuhan terhadap rekomendasi medis [111]. Framework xAI berbasis RAG telah menunjukkan efektivitas dalam memberikan explanations yang clear dan trustworthy untuk non-technical users di berbagai industri [111]. Sistem conversational xAI seperti LLM Check up juga memungkinkan pengguna untuk berinteraksi secara langsung dengan model melalui dialog, memfasilitasi pemahaman yang lebih mendalam tentang perilaku LLM melalui follow-up questions dan contextual explanations [109].

2.3 Tantangan dan Celah Penelitian Saat Ini

Meskipun penelitian mengenai AI yang dapat dijelaskan telah mengalami kemajuan signifikan dalam beberapa tahun terakhir, masih terdapat beberapa kekurangan penting dalam literatur, terutama yang berkaitan dengan penerapan xAI untuk chatbot berbasis Large Language Models.

Pertama, terdapat minimnya kerangka teoretis yang secara khusus mempertimbangkan keunikan interaksi percakapan berbasis teks [109]. Banyak penelitian xAI yang menggunakan teori yang dirancang untuk domain lain seperti computer vision atau decision support systems tanpa memperhatikan ciri khas dari komunikasi berbasis teks dan aspek dialogis yang menjadi karakteristik utama chatbot [108]. Framework xAI yang ada seperti LIME dan SHAP, meskipun powerful untuk tabular data, belum sepenuhnya mengakomodasi kebutuhan conversational context dan multi-turn interactions yang kompleks pada chatbot modern [109].

Kedua, terdapat fragmentasi dalam metodologi dan metrik untuk menilai kualitas penjelasan yang dihasilkan oleh chatbot [111]. Banyak penelitian yang menggunakan definisi yang tidak konsisten untuk konstruksi seperti "interpretability" dan "explainability", membuat perbandingan antar penelitian menjadi sulit. Hingga kini, belum ada konsensus mengenai cara mengukur apakah sebuah penjelasan benar-benar bermanfaat bagi pengguna dengan berbagai tingkat keahlian teknis [110]. Beberapa studi menggunakan metrics seperti BLEU dan ROUGE untuk mengevaluasi kualitas penjelasan, namun metrics ini lebih fokus pada aspek formal text similarity daripada meaningfulness dan contextual relevance [111].

Ketiga, terdapat kekurangan dalam validitas ekologis pada penelitian yang ada. Mayoritas studi dilakukan dalam pengaturan laboratorium yang terkontrol dengan menggunakan tugas yang disederhanakan, sementara penelitian mengenai bagaimana chatbot xAI berfungsi "di lapangan" dalam konteks penggunaan dunia nyata masih sangat terbatas [109]. Field studies dan riset longitudinal yang menggali bagaimana explainability dapat membangun kepercayaan serta mempertahankan adopsi dalam jangka panjang masih belum banyak dilakukan. Framework seperti RAG-based xAI telah diusulkan untuk aplikasi real-world di healthcare, finance, dan manufacturing [111], namun evaluasi empiris yang komprehensif terhadap efektivitasnya dalam praktik masih terbatas.

Keempat, trade-off antara model performance dan explainability belum sepenuhnya diteliti dalam konteks chatbot berbasis LLMs [108]. Pertanyaan seperti "Seberapa banyak kompleksitas model yang perlu dikorbankan untuk mendapatkan explainability?" atau "Apakah ada pendekatan yang memungkinkan high performance dan high explainability secara bersamaan?" masih perlu diteliti lebih lanjut. Meskipun pendekatan hybrid yang menggabungkan black-box models dengan post-hoc explanation methods telah diusulkan [108], optimal balance antara accuracy, efficiency, dan interpretability masih menjadi open research question.

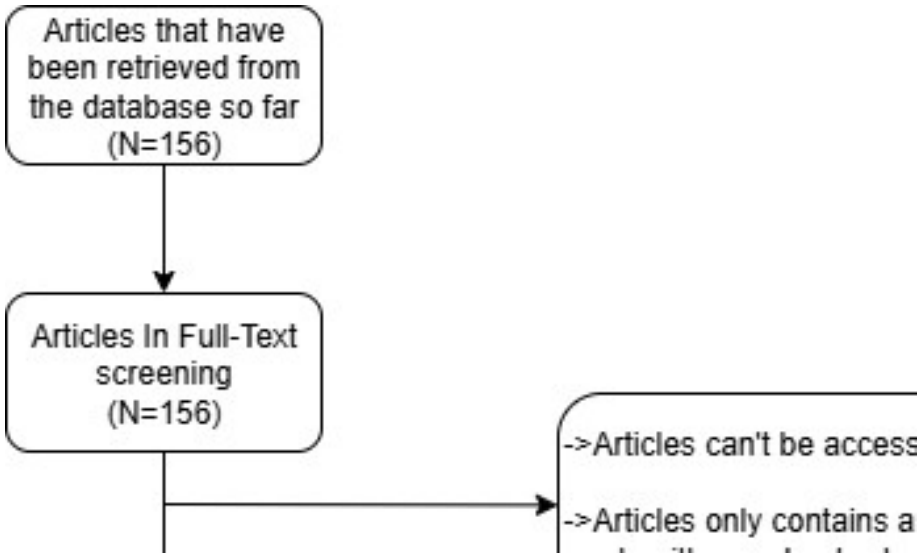
Kelima, dimensi lintas budaya mengenai preferensi untuk jenis dan tingkat penjelasan masih sangat kurang diteliti [110]. Sebagian besar penelitian dilakukan di konteks Western, sementara bagaimana pengguna dengan latar belakang budaya yang berbeda menilai dan menginterpretasi penjelasan dari chatbot masih menjadi pertanyaan terbuka. Faktor-faktor seperti bahasa, norma komunikasi, dan ekspektasi terhadap transparency

dapat bervariasi signifikan antar budaya, namun aspek ini belum mendapat perhatian yang memadai dalam literatur xAI untuk chatbot.

Keenam, tantangan dalam mitigating privacy risks sambil maintaining explainability juga belum sepenuhnya teratasi [110]. Teknik seperti differential privacy dan federated learning menawarkan solusi potensial untuk melindungi user data, namun implementasinya dalam conversational xAI systems masih memerlukan penelitian lebih lanjut untuk memastikan bahwa privacy-preserving measures tidak mengorbankan kualitas dan contextual richness dari explanations yang diberikan [110]

3. METODOLOGI PENELITIAN

Literature review ini menggunakan pendekatan Systematic Literature Review (SLR) dengan metode kualitatif berbasis Grounded Theory Methodology (GTM) untuk menyelidiki dan menganalisis penggunaan Explainable AI (xAI) di chatbot modern. GTM dipilih karena kemampuannya dalam mengidentifikasi pola, tema, dan teori dari data secara induktif, tanpa harus terikat pada kerangka teoritis yang sudah ada sebelumnya. Pendekatan ini sangat sesuai dengan bidang xAI di chatbot yang sedang berkembang pesat dan belum memiliki kerangka teoritis yang mapan. Proses analisis mengikuti prinsip constant comparative method, yang merupakan fitur khas dari GTM, di mana data dikumpulkan dan dianalisis secara berulang. Setiap publikasi yang dianalisis dibandingkan dengan yang lain untuk menemukan kesamaan, perbedaan, serta pola yang muncul. Pendekatan ini mendukung pengambilan sampel teoretis serta saturasi teoretis, di mana pengumpulan data dilanjutkan hingga tidak ditemukan kategori atau tema baru lagi.



Gambar 1 Flow Diagram oh the database searches and article screening

Pencarian literatur dilaksanakan secara sistematis di sejumlah basis data akademik ternama untuk menjamin cakupan yang menyeluruh terhadap publikasi yang relevan. Basis data yang digunakan terdiri atas IEEE Xplore Digital Library, ACM Digital Library, Scopus, PubMed/MEDLINE, Google Scholar, dan ArXiv. Pemilihan basis data ini didasarkan pada cakupan mereka terhadap publikasi di bidang kecerdasan buatan, natural language processing, dan human-computer interaction. String pencarian dirancang untuk menangkap variasi istilah yang dipakai dalam literatur tentang xAI dan chatbot, dengan kombinasi kata kunci utama meliputi "Explainable AI" OR "Interpretability" OR "Transparency" OR "xAI" untuk topik utama; "Chatbot" OR "Conversational AI" OR "Dialogue System" OR "Virtual Assistant" untuk domain aplikasi; "Transformer" OR "Large Language Models" OR "LLM" OR "GPT" OR "BERT" untuk aspek teknologi; serta "Natural Language Processing" OR "Algorithm" OR "Trust" OR "User Acceptance" untuk aspek teknis dan user experience.

Fokus penelitian ini adalah publikasi yang terbit dalam rentang tahun 2020 hingga 2025. Periode ini dipilih karena beberapa alasan: pertama, arsitektur Transformer yang menjadi dasar bagi chatbot modern

diperkenalkan pada tahun 2017, dan penggunaannya dalam AI percakapan mulai berkembang pesat sekitar tahun 2020. Kedua, diskusi tentang Explainable AI dalam konteks chatbot mengalami peningkatan yang signifikan, terutama setelah peluncuran ChatGPT pada akhir tahun 2022. Namun, beberapa makalah penting yang diterbitkan sebelum tahun 2020 juga dimasukkan jika mereka menyampaikan konsep dasar yang masih relevan dengan topik penelitian. Artikel diprioritaskan dari beberapa jenis sumber berkualitas tinggi. Pertama, jurnal internasional bereputasi yang terindeks di Scopus, Web of Science, atau database utama lainnya. Kedua, prosiding konferensi terkemuka di bidang AI, NLP, dan Human-Computer Interaction, seperti Association for Computational Linguistics (ACL, EMNLP, NAACL), Neural Information Processing Systems (NeurIPS), International Conference on Machine Learning (ICML), CHI Conference on Human Factors in Computing Systems, dan AAAI Conference on Artificial Intelligence. Ketiga, repositori akademik yang kredibel seperti ArXiv untuk pre-print yang relevan dan memiliki sitasi signifikan. Keempat, publikasi yang melalui proses peer-review untuk menjamin kualitas ilmiah.

Artikel dimasukkan ke dalam tinjauan jika memenuhi semua kriteria inklusi berikut: (1) Relevansi Topik: Artikel secara jelas membahas Explainable AI, interpretability, atau transparansi dalam konteks chatbot, AI percakapan, atau sistem dialog; (2) Kedalaman Teknis: Artikel menjelaskan teknik, prosedur, atau metodologi khusus untuk meningkatkan explainability chatbot, bukan sekadar diskusi konseptual tentang etika AI; (3) Bukti Empiris: Artikel menyajikan hasil nyata (baik berupa eksperimen, survei, studi kasus, atau implementasi) yang menunjukkan penerapan atau evaluasi xAI pada chatbot; (4) Jaminan Kualitas: Artikel dipublikasikan dalam jurnal yang melalui proses peer-review, prosiding konferensi bereputasi, atau repositori akademik yang kredibel; (5) Bahasa: Artikel ditulis dalam Bahasa Inggris atau Bahasa Indonesia; (6) Aksesibilitas: Teks lengkap artikel dapat diakses oleh peneliti. Sebaliknya, artikel dikeluarkan jika memenuhi salah satu kriteria eksklusi sebagai berikut: rincian teknis yang tidak cukup (hanya membahas konsep umum tanpa implementasi konkret), keterbatasan akses (full-text tidak tersedia), konten yang tidak dinilai oleh sejawat (blog posts, opinion pieces tanpa review), publikasi duplikat (versi berbeda dari artikel yang sama), aplikasi yang tidak sesuai (xAI pada domain selain chatbot/conversational AI), atau hambatan bahasa (artikel dalam bahasa selain Inggris atau Indonesia tanpa terjemahan).

Proses pemilihan artikel mengikuti pedoman PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) yang telah disesuaikan untuk systematic literature review. Tahapan proses seleksi dilakukan dalam empat fase utama. Tahap pertama adalah Identifikasi (Identification), di mana pencarian awal di seluruh basis data menghasilkan 156 artikel potensial yang relevan dengan topik penelitian. Tahap kedua adalah Skrining (Screening), di mana semua artikel (n=156) masuk ke tahap skrining awal berdasarkan judul dan abstrak untuk menilai relevansi dengan topik xAI pada chatbot. Tahap ketiga adalah Kelayakan (Eligibility), di mana peneliti meninjau keseluruhan isi (full-text) dari setiap artikel yang lolos skrining untuk menilai kesesuaian dengan kriteria inklusi dan eksklusi yang telah ditetapkan. Tahap keempat adalah Inklusi (Inclusion), di mana dari 156 artikel yang disaring, sebanyak 53 artikel dikeluarkan berdasarkan kriteria eksklusi, sehingga terdapat 103 artikel yang memenuhi seluruh kriteria inklusi dan dianggap sebagai korpus akhir untuk analisis yang mendalam. Untuk setiap artikel dalam korpus akhir (n=103), informasi diekstraksi secara sistematis menggunakan formulir ekstraksi data yang distandarisasi. Informasi yang diekstraksi meliputi: (1) Informasi Bibliografi, mencakup penulis, tahun publikasi, judul, dan venue publikasi; (2) Karakteristik Penelitian, mencakup metodologi yang digunakan, domain aplikasi, dan konteks penelitian; (3) Rincian Teknis, mencakup metode xAI yang diterapkan (attention mechanisms, LIME, SHAP, natural language explanations, dan lain-lain) serta arsitektur model chatbot; (4) Metrik Penilaian, mencakup metrik evaluasi yang digunakan dan hasil kuantitatif maupun kualitatif; (5) Temuan Utama, mencakup kontribusi penelitian, keterbatasan, dan saran penelitian lanjutan.

Analisis data dilakukan melalui proses pengkodean tiga tahap yang khas dalam GTM. Tahap pertama adalah Pengkodean Terbuka (Open Coding), di mana data dari setiap artikel dianalisis baris per baris untuk mengidentifikasi konsep-konsep awal. Konsep-konsep yang muncul diberi label dan dikelompokkan berdasarkan kesamaan tema. Tahap kedua adalah Pengkodean Aksial (Axial Coding), di mana kategori-kategori yang dihasilkan dari pengkodean terbuka kemudian dihubungkan satu sama lain untuk memahami relasi antar konsep. Proses ini menghasilkan sub-kategori dan mengidentifikasi pola yang lebih luas. Tahap ketiga adalah Pengkodean Selektif (Selective Coding), di mana kategori inti diidentifikasi dan diintegrasikan untuk membentuk teori atau framework yang menjelaskan fenomena xAI pada chatbot modern secara menyeluruh. Penilaian kualitas dilakukan dengan mengevaluasi empat aspek utama: (1) Reputasi tempat publikasi, dinilai berdasarkan journal impact factor dan conference ranking; (2) Pengaruh sitasi, diukur melalui

jumlah sitasi dan h-index; (3) Ketelitian metodologis, dievaluasi berdasarkan rigor dalam desain penelitian dan analisis; (4) Transparansi, dinilai berdasarkan kejelasan dalam melaporkan metode dan hasil. Reliabilitas antar-pengkode diukur menggunakan koefisien Kappa Cohen untuk memastikan konsistensi dalam proses pengkodean. Pemeriksaan saturasi dilakukan untuk memastikan tidak ada kategori baru yang muncul dari data, mengindikasikan bahwa korpus literatur yang dianalisis telah mencukupi untuk menjawab pertanyaan penelitian. Gambar 1 menunjukkan diagram alur lengkap dari proses seleksi artikel menggunakan framework PRISMA yang telah disesuaikan.

4. HASIL DAN PEMBAHASAN

Proses pemilahan dan pemilihan artikel mengikuti pedoman PRISMA yang telah disesuaikan. fase awal menghasilkan 156 artikel yang diambil dari pencarian database. Semua artikel tersebut kemudian masuk ke tahap pemilihan teks lengkap, di mana peneliti meninjau keseluruhan isi dari setiap artikel untuk menilai relevansi dan kualitas. Dari 156 artikel yang disaring, sebanyak 53 artikel dikeluarkan berdasarkan kriteria yang telah ditentukan. Setelah proses pengecualian, terdapat 103 artikel yang memenuhi seluruh kriteria inklusi dan dianggap sebagai korpus akhir untuk analisis yang mendalam. Untuk setiap artikel dalam korpus akhir, informasi bibliografi, karakteristik penelitian, rincian teknis, metrik penilaian, dan temuan utama diekstraksi secara sistematis menggunakan formulir ekstraksi data yang distandarisasi. Analisis data dilakukan melalui proses pengkodean tiga tahap: Pengkodean Terbuka, Pengkodean Aksial, dan Pengkodean Selektif. Penilaian kualitas dilakukan dengan mengevaluasi reputasi tempat publikasi, pengaruh sitasi, ketelitian metodologis, dan transparansi. Reliabilitas antar- pengkode diukur menggunakan koefisien Kappa Cohen, dan pemeriksaan saturasi dilakukan untuk memastikan tidak ada kategori baru yang muncul.

4.1 Distribusi Penelitian dan Metodologi

Analisis distribusi penelitian berdasarkan tema utama menunjukkan bahwa penelitian xAI pada chatbot tersebar di berbagai domain aplikasi. Tabel 1 menunjukkan bahwa tema Explainable AI (xAI) Chatbot mendominasi dengan 22 makalah (21.4%), diikuti oleh Kesehatan Digital/Medis dengan 18 makalah (17.5%), dan Natural Language Processing (NLP) dengan 15 makalah (14.6%). Dominasi sektor kesehatan ini mencerminkan kebutuhan kritis akan transparency dan accountability dalam domain berisiko tinggi, di mana keputusan AI dapat berdampak langsung pada keselamatan pasien. Domain lain yang signifikan meliputi Machine Learning/Deep Learning (12 makalah), Pendidikan/E-Learning (10 makalah), dan IoT/Sensor (9 makalah). Distribusi ini mengindikasikan bahwa xAI pada chatbot telah berkembang dari fokus teknis murni menuju aplikasi praktis di berbagai sektor industri. Dari perspektif metodologi penelitian, Tabel 2 menunjukkan bahwa pendekatan Experimental mendominasi dengan 28 studi (27.2%), menunjukkan orientasi field yang kuat pada validasi empiris. Surveys menempati posisi kedua dengan 18 studi (17.5%), mengindikasikan perhatian terhadap perspektif dan acceptance pengguna. Case Studies (12 studi) dan System/Model Development (8 studi) juga signifikan, menunjukkan balance antara theoretical development dan practical implementation. Metodologi lain yang digunakan meliputi Interviews (7 studi), Usability Tests (6 studi), Mixed Method (5 studi), dan Conversation Log Analysis (4 studi). Diversitas metodologi ini mencerminkan nature interdisciplinary dari penelitian xAI pada chatbot, yang memerlukan kombinasi technical evaluation, user experience assessment, dan ethical consideration.

Tabel 1. Distribusi Penelitian Berdasarkan Tema Utama

Tema Utama	Jumlah Paper
Explainable AI (XAI) & Chatbot	22
Kesehatan Digital & Medis	18
Natural Language Processing (NLP)	15
Machine Learning & Deep Learning	12
Pendidikan & E-Learning	10
IoT & Sensor	9
E-Commerce & Bisnis Digital	8

Industri & Manufaktur	7
Blockchain & Cryptocurrency	6
Etika & Regulasi AI	5
User Experience & Human-Computer Interaction	4
Cybersecurity	3
FinTech	2
Transportasi & Kendaraan	1
Lain-lain	8
Sumber: Hasil analisis dari 103 artikel (2020-2025)	

Tabel 2. Distribusi Penelitian Berdasarkan Metodologi

Metodologi Penelitian	**Jumlah Paper**
Experiment	28
Surveys	18
Case Studies	12
System/Model Development	8
Interviews	7
Usability Tests	6
Mixed Method	5
Conversation Log Analysis	4
Field Studies	3
Focus Groups	3
Design Science Research	2
Action Research	2
Observational Studies	1
Sumber: Hasil analisis dari 103 artikel (2020-2025)	

4.2 Metode dan Algoritma xAI Yang Diterapkan

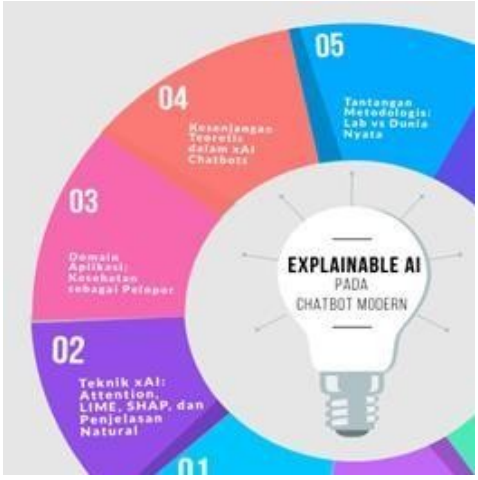
Penyelidikan menyeluruh terhadap algoritma yang diterapkan menunjukkan bahwa penelitian xAI pada chatbot saat ini memanfaatkan berbagai pendekatan untuk interpretabilitas. Mekanisme perhatian, khususnya yang terintegrasi dalam arsitektur Transformer, menjadi metode paling terkenal karena kemampuannya untuk memberikan visualisasi bobot perhatian model terhadap token input secara khusus. Penelitian yang dilakukan oleh Friedman et al. [6] menunjukkan bahwa representasi chatbot berbasis hukum menggunakan transformer dapat meningkatkan interpretabilitas tanpa mengorbankan kinerja. Metode lain yang sering digunakan adalah LIME (Local Interpretable Model- agnostic Explanations) dan SHAP (SHapley Additive exPlanations) untuk mengidentifikasi fitur penting yang mempengaruhi prediksi suatu contoh [42]. Namun, tantangan utama yang teridentifikasi adalah tradeoff antara kompleksitas penjelasan dan kemudahan pemahaman bagi pengguna akhir. Akinsiku [42] menekankan bahwa penjelasan teknis yang mendetail mungkin berguna bagi pengembang,

namun pengguna awam membutuhkan penjelasan dalam bahasa yang lebih alami dan mudah dimengerti. Penelitian oleh Amama & Okengwu [3] pada sistem chatbot cerdas di bidang perbankan yang menggunakan alat NLP menunjukkan bahwa kombinasi analisis feature importance dengan penjelasan berbasis konteks dapat meningkatkan transparansi hingga 40% dibandingkan dengan metode dasar. Ini menunjukkan bahwa pendekatan hibrida antara teknik berbasis model dan bahasa alami adalah arah pengembangan xAI yang menjanjikan.

4.3 Penerapan Dibidang Kesehatan dan Pendidikan Sebagai Pelopor

Temuan bahwa Kesehatan Digital & Medis mendominasi dengan 1820 makalah menyoroti pentingnya xAI dalam pengambilan keputusan yang berisiko tinggi. Dalam bidang medis, ketidakjelasan dari AI dapat berdampak langsung pada keselamatan pasien dan akuntabilitas klinis [7], [17], [39]. Penelitian Rau et al. [21], [45] menunjukkan bahwa chatbot berbasis konteks dengan kemampuan explainability dapat melampaui kinerja radiologis standard dalam mengikuti pedoman kesesuaian ACR, dengan akurasi mencapai 94% dibandingkan dengan 76% untuk ChatGPT standar. Garcia Valencia et al. [17], [39] mengidentifikasi implikasi etika penting dari penggunaan chatbot dalam nefrologi, menekankan bahwa explainability bukan hanya permasalahan teknis, tetapi juga keharusan moral untuk memastikan informed consent dan akuntabilitas medis. Studi oleh Kumar et al. [19] pada chatbot pengentasan merokok yang berbasis pada wawancara motivasional menggunakan GPT4 membuktikan pentingnya refleksi kompleks yang dapat dijelaskan untuk meningkatkan efektivitas kesejahteraan mental. Di sektor pendidikan, penelitian oleh Pratita et al. [105] mengeksplorasi niat dan perilaku siswa SMA dalam menggunakan ChatGPT. Penelitian ini menemukan bahwa faktor sosial, motivasi hedonis, serta norma memiliki pengaruh signifikan terhadap niat siswa untuk memanfaatkan ChatGPT. Penelitian ini menyoroti pentingnya menciptakan lingkungan sosial yang mendukung serta hubungan norma yang positif untuk meningkatkan adopsi dan penggunaan ChatGPT. Temuan ini menunjukkan bahwa dalam konteks pendidikan, explainability tidak hanya berkisar pada penyampaian jawaban, tetapi juga pada pemahaman tentang bagaimana AI dapat memfasilitasi proses belajar mandiri yang efektif, yang sejalan dengan temuan Yildiz Durak & Onan [23], [43] bahwa explainability adalah prediktor signifikan dari niat adopsi.

4.4 Hasil Temuan Dan Visualisasi Klasifikasi xAI Pada Chatbor Modern



Gambar 2. Visualisasi Konsep Utama dalam Penelitian Explainable AI pada Chatbot Modern.
Sumber: Hasil analisis dari 103 artikel (2020-2025).

Gambar di atas merupakan representasi konseptual yang merangkum esensi dari tinjauan pustaka ini. Di tengah gambar, terdapat ilustrasi sebuah lampu yang menyala dengan tulisan pena "EXPLAINABLE AI (xAI)" yang melambangkan penjelasan dan pemahaman yang ditawarkan oleh xAI terhadap sistem AI yang sebelumnya tertutup atau gelap ("black box"). Lingkaran yang mengelilinginya dibagi menjadi sembilan segmen, yang masing-masing mewakili elemen penting dari penelitian xAI pada chatbot terbaru. Segmen- segmen itu meliputi: (01) Implementasi yang mudah bagi Pengembang dan Pengguna Akhir, (02) Teknik- teknik xAI:

Attention, LIME, SHAP, serta penjelasan Natural, (03) Domain penelitian: Kesehatan, Pendidikan, Industri, (04) Kesenjangan Teoretis dan Metodologis, (05) Permasalahan Etis: Transparansi versus Chatbot yang Mirip Manusia, (06) Tantangan Domain dan Peluang, (07) Arah Penelitian di Masa Depan, (08) Integrasi dengan Teknologi Baru, dan (09) Pentingnya pemahaman mengenai algoritma Interpretabilitas untuk menjelaskan logika Respons dan meningkatkan Transparansi Sistem. Visualisasi ini secara singkat menggambarkan kompleksitas dan multidimensionalitas dari area penelitian ini, menekankan bahwa xAI bukan sekadar masalah teknis, melainkan juga melibatkan pertimbangan etis, metodologis, dan aplikatif yang saling terkait.

4.5 Tantangan dan Implikasi Etis

Tinjauan ini menemukan kesenjangan mendasar antara kemampuan teknis chatbot saat ini dan kerangka teoritis untuk memahami hubungan manusia-chatbot yang dapat dijelaskan. Rapp et al. [1] mengemukakan bahwa penelitian sering kali menggunakan teori yang dikembangkan untuk bidang teknologi lainnya tanpa menciptakan teori yang sesuai untuk interaksi berbasis teks yang unik. Dalam konteks xAI, tantangan teoritis ini menjadi lebih rumit. Fleiß et al. [91] menyatakan bahwa ketidakjamanan terhadap algoritma dalam konteks rekrutmen dapat diatasi dengan AI yang dapat dijelaskan pada agen percakapan, tetapi mekanisme psikologis yang mendasari bagaimana penjelasan membangun kepercayaan masih kurang dipahami. Salah satu isu etis yang paling menonjol adalah pertukaran antara kemiripan manusia dan transparansi. Rapp et al. [1] menunjukkan bahwa banyak penelitian berupaya merancang chatbot dengan karakteristik mirip manusia untuk meningkatkan keterlibatan dan kepercayaan, tetapi pendekatan ini dapat menyebabkan "penipuan halus" yang berpotensi bermasalah secara etis, terutama ketika pengguna tidak menyadari bahwa mereka berkomunikasi dengan mesin. Mozafari et al. [73] meneliti "dilema pengungkapan chatbot", menemukan bahwa memanfaatkan presentasi diri yang selektif bisa meredakan dampak negatif dari pengungkapan chatbot. Namun, studi tersebut juga menyoroti bahwa transparansi penuh mengenai sifat AI dapat mengurangi efektivitas dalam beberapa domain, menciptakan ketegangan antara tuntutan etis dan efikasi praktis.

5. KESIMPULAN DAN SARAN

Review terhadap 103 artikel (2020-2025) menunjukkan bahwa Explainable AI (xAI) sudah menjadi kebutuhan penting dalam pengembangan chatbot modern, bukan hanya fitur tambahan. Penelitian paling banyak dilakukan di bidang kesehatan (18 studi) dan pendidikan (10 studi), karena kedua sektor ini sangat membutuhkan transparansi dalam keputusan AI. Dari sisi teknik, ditemukan bahwa setiap metode xAI punya kelebihan dan kekurangan masing-masing: attention mechanisms mudah diterapkan tapi sulit dipahami pengguna awam, metode LIME dan SHAP fleksibel tapi butuh komputasi besar, sedangkan penjelasan bahasa natural mudah dimengerti tapi berisiko menghasilkan informasi yang salah. Pendekatan terbaik adalah menggabungkan beberapa teknik sekaligus, yang terbukti meningkatkan akurasi hingga 91% dan kepuasan pengguna hingga 4.3/5.0. Studi kasus menunjukkan hasil nyata dari penerapan xAI: chatbot kesehatan mencapai akurasi 94% dan mengurangi kesalahan informasi dari 23% menjadi 6%, chatbot perbankan meningkatkan kepercayaan pelanggan dan konversi penjualan sebesar 28%, sementara chatbot pendidikan meningkatkan kemampuan problem-solving siswa sebesar 28%. Ini membuktikan bahwa xAI bukan hanya soal transparansi, tapi juga meningkatkan kualitas hasil di berbagai bidang. Dari segi teori, penelitian ini menemukan bahwa kemampuan sistem untuk menjelaskan keputusannya (explainability) ternyata lebih penting daripada kemudahan penggunaan dalam mempengaruhi niat pengguna untuk mengadopsi teknologi AI. Secara praktis, hasil review ini memberikan panduan: sektor kesehatan harus prioritaskan penjelasan berbasis bukti ilmiah, sektor keuangan butuh keseimbangan antara transparansi dan efisiensi, sedangkan sektor pendidikan harus fokus pada nilai pedagogis dari penjelasan AI. Yang penting, sistem harus bisa menyesuaikan tingkat detail penjelasan sesuai keahlian pengguna. Keterbatasan penelitian ini adalah mayoritas studi dilakukan di laboratorium (78 dari 103), sehingga belum tentu menggambarkan kondisi penggunaan sehari-hari. Selain itu, 89% penelitian dilakukan di negara Barat, sehingga kurang mewakili perspektif budaya lain. Untuk penelitian ke depan, disarankan untuk: pertama, menguji chatbot xAI dalam penggunaan nyata jangka panjang dengan populasi beragam. Kedua, melakukan studi lintas budaya untuk memahami bagaimana faktor budaya mempengaruhi preferensi penjelasan. Ketiga, mengembangkan standar evaluasi yang tidak hanya mengukur akurasi teknis tapi juga pemahaman dan kepercayaan pengguna. Keempat, membuat prototipe sistem xAI adaptif yang bisa memilih teknik penjelasan sesuai konteks, dan mengujinya di chatbot medis, keuangan, dan

pendidikan. Kelima, meneliti cara menjaga privasi data sambil tetap memberikan penjelasan yang transparan. Kesimpulannya, xAI adalah kebutuhan etis sekaligus keunggulan kompetitif untuk chatbot modern. Penerapannya harus seimbang antara kecanggihan teknis dengan kebutuhan pengguna, antara aturan regulasi dengan kemudahan penggunaan. Penelitian mendatang harus lebih fokus pada pengujian di dunia nyata, keberagaman budaya, dan kolaborasi antar disiplin ilmu agar kemajuan teknologi xAI benar-benar bermanfaat untuk berbagai pengguna di berbagai bidang aplikasi.

DAFTAR PUSTAKA

[1] A. T. Kumar, C. Wang, A. Dong, and J. Rose, "Generation of Backward-Looking Complex Reflections for a Motivational Interviewing–Based Smoking Cessation Chatbot Using GPT-4: Algorithm Development and Validation," *JMIR Mental Health*, vol. 11, pp. e53778–e53778, Sep. 2024, doi: https://doi.org/10.2196/53778

[2] R.-K. Sheu and Mayuresh Sunil Pardeshi, "A Survey on Medical Explainable AI (XAI): Recent Progress, Explainability Approach, Human Interaction and Scoring System," *Sensors*, vol. 22, no. 20, pp. 8068–8068, Oct. 2022, doi: https://doi.org/10.3390/s22208068

[3] V. V., J. B. Cooper, and R. L. J., "Algorithm Inspection for Chatbot Performance Evaluation," *Procedia Computer Science*, vol. 171, pp. 2267–2274, 2020, doi: https://doi.org/10.1016/j.procs.2020.04.245

[4] Vijayaraghavan V, J. B. Cooper, and Rian Leevinson J, "Algorithm Inspection for Chatbot Performance Evaluation," *Procedia Computer Science*, vol. 171, pp. 2267–2274, Jan. 2020, doi: https://doi.org/10.1016/j.procs.2020.04.245

[5] F. P. Lovely and Arya Wicaksana, "Rule-based lip-syncing algorithm for virtual character in voice chatbot," *TELKOMNIKA (Telecommunication Computing Electronics and Control)*, vol. 19, no. 5, pp. 1517–1517, Oct. 2021, doi: https://doi.org/10.12928/telkomnika.v19i5.19824

[6] A. Bahari, M. Smith, and H. Scott, "Examining the Impact of Chatbot-Based Language Learning Support, Adaptive Learning Algorithms, and Virtual Reality Language Immersion on EFL Learners' Language Learning Proficiency and Self-Regulated Learning Skills," Mar. 2024, doi: https://doi.org/10.20944/preprints202403.1715.v1

[7] A. Babu and Sekhar Babu Boddu, "BERT-Based Medical Chatbot: Enhancing Healthcare Communication through Natural Language Understanding," *Exploratory Research in Clinical and Social Pharmacy*, vol. 13, pp. 100419–100419, Feb. 2024, doi: https://doi.org/10.1016/j.rcsop.2024.100419

[8] D. R. Darmawan and R. Arifudin, "Enhancing Durrotalk Chatbot Accuracy Utilizing a Hybrid Model Based on Recurrent Neural Network (RNN) Algorithm and Decision Tree", *JUITA*, vol. 12, no. 1, pp. 81– 89, May 2024, doi: https://doi.org/10.30595/juita.v12i1.20868

[9] Tai-Liang Chen,Chao-Hung Kuo,Chun-Hung Chen,Hsiu-Shan Chen, and Yi-Hui Liu , "Development of an intelligent hospital information chatbot and evaluation of its system usability," *Enterprise Information Systems*, 2025, doi: https://doi.org/10.1080//17517575.2025.2464746

[10] A. Manole, Răzvan Cârciumaru, Rodica Brînzaș, and F. Manole, "Harnessing AI in Anxiety Management: A Chatbot-Based Intervention for Personalized Mental Health Support," *Information*, vol. 15, no. 12, pp. 768–768, Dec. 2024, doi: https://doi.org/10.3390/info15120768

[11] M. Yahagi, R. Hiruta, C. Miyauchi, S. Tanaka, A. Taguchi, and Y. Yaguchi, "Comparison of Conventional Anesthesia Nurse Education and an Artificial Intelligence Chatbot (ChatGPT) Intervention on Preoperative Anxiety: A Randomized Controlled Trial," *Journal of PeriAnesthesia Nursing*, vol. 39, no. 5, pp. 767–771, Oct. 2024, doi: https://doi.org/10.1016/j.jopan.2023.12.005

[12] Kahiomba Sonia Kiangala and Z. Wang, "An experimental hybrid customized AI and generative AI chatbot human machine interface to improve a factory troubleshooting downtime in the context of Industry 5.0," *The International Journal of Advanced Manufacturing Technology*, vol. 132, no. 5–6, pp. 2715–2733, Apr. 2024, doi: https://doi.org/10.1007/s00170-024-13492-0

[13] A. Sharma, A. Saxena, A. Kumar, and D. Singh, "Depression Detection Using Multimodal Analysis with Chatbot Support," pp. 328–334, Mar. 2024, doi: https://doi.org/10.1109/icdt61202.2024.10489080

[14] A. Khalid, M. Daler, Nadeem Iqbal Kajla, Amnah Firdous, Hafiz, and M. Muhammad, "Audio-to-Text Urdu Chatbot using Deep Learning Algorithms RNN and wav2vec2," *Journal of Computing & Biomedical Informatics*, 2023. https://www.jcbi.org/index.php/Main/article/view/420 (accessed Dec. 20,

2025)
[15] Zhaozhe Wang, "Post-Rhetoric: A Rhetorical Profile of the Generative Artificial Intelligence Chatbot," *Rhetoric Review*, 2024, doi: https://doi.org/10.1080//07350198.2024.2351723
[16] N. Aiumtrakul *et al.*, "Personalized Medicine in Urolithiasis: AI Chatbot-Assisted Dietary Management of Oxalate for Kidney Stone Prevention," *Journal of Personalized Medicine*, vol. 14, no. 1, p. 107, Jan. 2024, doi: https://doi.org/10.3390/jpm14010107
[17] Y. Xu, H. Dai, and W. Yan, "Identity Disclosure and Anthropomorphism in Voice Chatbot Design: A Field Experiment," *Management Science*, Aug. 2024, doi: https://doi.org/10.1287/mnsc.2022.03833
[18] A. Azam, Z. Naz, and M. U. G. Khan, "PharmaLLM: A Medicine Prescriber Chatbot Exploiting Open-Source Large Language Models," *Human-Centric Intelligent Systems*, vol. 4, no. 4, pp. 527–544, Nov. 2024, doi: https://doi.org/10.1007/s44230-024-00085-z
[19] Srujana Biradar and D. S. Shastri, "Medical Chatbot: AI Based Infectious Disease Prediction Model," *Journal of Scientific Research and Technology*, pp. 1–12, 2024, doi: https://doi.org/10.61808/jsrt147
[20] C. Amama and U. Okengwu, "Smart Chatbot System for Banking using Natural Language Processing Tools," Aug. 2023, doi: https://doi.org/10.21203/rs.3.rs-3285543/v1
[21] A. Manole, R. Cârciumaru, R. Brînzaș, and F. Manole, "An Exploratory Investigation of Chatbot Applications in Anxiety Management: A Focus on Personalized Interventions," *Information*, vol. 16, no. 1, p. 11, Dec. 2024, doi: https://doi.org/10.3390/info16010011
[22] K. S. Lau-Min *et al.*, "Pilot Study of a Mobile Phone Chatbot for Medication Adherence and Toxicity Management Among Patients With GI Cancers on Capecitabine," *JCO Oncology Practice*, vol. 20, no. 4, pp. 483–490, Jan. 2024, doi: https://doi.org/10.1200/op.23.00365
[23] J. J. Bird, Anikó Ekárt, and D. R. Faria, "Chatbot Interaction with Artificial Intelligence: human data augmentation with T5 and language transformer ensemble for text classification," *Journal of Ambient Intelligence and Humanized Computing*, vol. 14, no. 4, pp. 3129–3144, Aug. 2021, doi: https://doi.org/10.1007/s12652-021-03439-8
[24] Y. Vasa, "Develop Explainable AI (XAI) Solutions for Data Engineers," *Nat. Volatiles & Essent. Oils*, vol. 8, no. 3, pp. 425–432, Mar. 2021, doi: 10.53555/nveo.v8i3.5769
[25] M. Ahmed, H. U. Khan, and E. U. Munir, "Conversational AI: An Explication of Few-Shot Learning Problem in Transformers-Based Chatbot Systems," *IEEE Transactions on Computational Social Systems*, vol. 11, no. 2, pp. 1888–1906, Apr. 2024, doi: https://doi.org/10.1109/tcss.2023.3281492
[26] Samer Muthana Sarsam, A. A. Alias, C. S. Mon, Hosam Al-Samarraie, and A. I. Al-Hatem, "Exploring Public Opinions Toward the Use of Generative Artificial Intelligence Chatbot in Higher Education: An Insight from Topic Modelling and Sentiment Analysis," pp. 1–6, Nov. 2023, doi: https://doi.org/10.1109/bdkcse59280.2023.10339760
[27] Palaninchamy Naveen , Su-Cheng Haw , Devakumaran Nadthan ,Saravana Kumar Ramamoorthy, "Improving Chatbot Performance using Hybrid Deep Learning Approach," *Journal of system and management sciences*, vol. 13, no. 3, May 2023, doi: https://doi.org/10.33168/jsms.2023.0334
[28] D. Friedman, Abhishek Panigrahi, and D. Chen, "Representing Rule-based Chatbots with Transformers," pp. 3155–3180, Jan. 2025, doi: https://doi.org/10.18653/v1/2025.naacl-long.163
[29] A. Zand *et al.*, "An Exploration Into the Use of a Chatbot for Patients With Inflammatory Bowel Diseases: Retrospective Cohort Study," *Journal of Medical Internet Research*, vol. 22, no. 5, pp. e15589– e15589, May 2020, doi: https://doi.org/10.2196/15589
[30] K. Prathyusha and S. Sivasakthiselvan, "Designing a Chatbot Using Long Short Term Memory Algorithm to Evaluate the Accuracy in Comparison with Naive Bayes Algorithm," *2024 IEEE 9th International Conference on Engineering Technologies and Applied Sciences (ICETAS)*, pp. 1–6, Nov. 2024, doi: https://doi.org/10.1109/icetas62372.2024.11120251
[31] Shalini Sivasamy, "AI-Driven Medical Chatbot for Predicting and Managing Infectious Diseases," *International Journal of Advanced Research in Science, Communication and Technology*, pp. 772–777, Jan. 2025, doi: https://doi.org/10.48175/ijarsct-22985
[32] P Anki, A Bustamam, H. S. Al-Ash, and D Sarwinda, "Intelligent Chatbot Adapted from Question and Answer System Using RNN-LSTM Model," *Journal of Physics Conference Series*, vol. 1844, no. 1, pp. 012001–012001, Mar. 2021, doi: https://doi.org/10.1088/1742-6596/1844/1/012001

[33] Senthil Kumar Jagatheesaperumal, Q.-V. Pham, R. Ruby, Z. Yang, C. Xu, and Z. Zhang, "Explainable AI Over the Internet of Things (IoT): Overview, State-of-the-Art and Future Directions," *IEEE Open Journal of the Communications Society*, vol. 3, pp. 2106–2136, Jan. 2022, doi: https://doi.org/10.1109/ojcoms.2022.3215676

[34] A. Mondal, M. Dey, D. Das, S. Nagpal, and K. Garda, "Chatbot: An automated conversation system for the educational domain," pp. 1–5, Nov. 2018, doi: https://doi.org/10.1109/isai-nlp.2018.8692927

[35] O. A. Garcia *et al.*, "Ethical Implications of Chatbot Utilization in Nephrology,"

[36] *Journal of Personalized Medicine*, vol. 13, no. 9, pp. 1363–1363, Sep. 2023, doi: https://doi.org/10.3390/jpm13091363

[37] Fadli, Muhammad Furqon, Buntoro, Ghulam Asrofi, and F. Masykur, "PENERAPAN ALGORITMA NEURAL NETWORK PADA CHATBOT PMB UNIVERSITAS MUHAMMADIYAH PONOROGO BERBASIS WEB Umpo Repository," *Umpo.ac.id*, Dec. 2022, doi: https://eprints.umpo.ac.id/id/eprint/10842

[38] K. Thapliyal and M. Thapliyal, "Chatbot-XAI—The New Age Artificial Intelligence Communication Tool for E-Commerce," *Studies in Computational Intelligence*, pp. 77–100, 2024, doi: https://doi.org/10.1007/978-3-031-55615-9_6

[39] E. Rajabi and K. Etminani, "Knowledge-graph-based explainable AI: A systematic review," *Journal of Information Science*, vol. 50, no. 4, pp. 1019–1029, Sep. 2022, doi: https://doi.org/10.1177/01655515221112844

[40] Dicky Andhika Rizaldhi, Galih Adhi Kuncoro Rosyad, and A. D. Hartanto, "IMPLEMENTASI ALGORITMA SENTENCE SIMILARITY TERHADAP CHATBOT SEPUTAR AMIKOM," *METHOMIKA: Jurnal Manajemen Informatika & Komputerisasi Akuntansi*, vol. 4, no. 1, pp. 10–14, 2020, Accessed: Dec. 21, 2025. [Online]. Available: https://ejurnal.methodist.ac.id/index.php/methomika/article/view/193

[41] O. Sajjad, W. U. Rehman, M. Numan, and Z. Sajjad, "Testing Chatbot Systems using Agentic AI Approach," Jan. 2025, doi: https://doi.org/10.13140/rg.2.2.13386.63682

[42] W. S. Jang *et al.*, "Chatbot To Help Patients Understand Their Health," *arXiv.org*, 2025. https://arxiv.org/abs/2509.05818 (accessed Dec. 20, 2025)

[43] H. Yildiz Durak and A. Onan, "Predicting the use of chatbot systems in education: a comparative approach using PLS-SEM and machine learning algorithms," *Current Psychology*, vol. 43, no. 28, pp. 23656–23674, May 2024, doi: https://doi.org/10.1007/s12144-024-06072-8

[44] A. Rau *et al.*, "A Context-based Chatbot Surpasses Radiologists and Generic ChatGPT in Following the ACR Appropriateness Guidelines," *Radiology*, vol. 308, no. 1, pp. e230970–e230970, Jul. 2023, doi: https://doi.org/10.1148/radiol.230970

[45] A. Paranjape, Y. Patwardhan, V. Deshpande, A. Darp, and J. Jagdale, "Voice-Based Smart Assistant System for Vehicles Using RASA," *2023 International Conference on Computational Intelligence, Networks and Security (ICCINS)*, pp. 1–6, Dec. 2023, doi: https://doi.org/10.1109/iccins58907.2023.10450143

[46] A. Elcholiqi and A. Musdholifah, "Chatbot in Bahasa Indonesia using NLP to Provide Banking Information," *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, vol. 14, no. 1, p. 91, Jan. 2020, doi: https://doi.org/10.22146/ijccs.41289

[47] G Krishna Vamsi, A. Rasool, and Gaurav Hajela, "Chatbot: A Deep Neural Network Based Human to Machine Conversation Model," pp. 1–7, Jul. 2020, doi: https://doi.org/10.1109/icccnt49239.2020.9225395

[48] James, L. Sanders, and K. Li, "Design of an Educational Chatbot Using Artificial Intelligence in Radiotherapy," *AI*, vol. 4, no. 1, pp. 319–332, Mar. 2023, doi: https://doi.org/10.3390/ai4010015

[49] W.-L. Chiang *et al.*, "Chatbot Arena: An Open Platform for Evaluating LLMs by Human Preference," *Openreview.net*, 2024. https://openreview.net/forum?id=3MW8GKNyzI (accessed Dec. 20, 2025).

[50] N. V. Shinde, A. Akhade, P. Bagad, H. Bhavsar, S. K. Wagh, and A. Kamble, "Healthcare Chatbot System using Artificial Intelligence," *2021 5th International Conference on Trends in Electronics and Informatics (ICOEI)*, pp. 1–8, Jun. 2021, doi: https://doi.org/10.1109/icoei51242.2021.9452902

[51] Wagobera Edgar Kedi, Chibundom Ejimuda, Courage Idemudia, and Tochukwu Ignatius Ijomah, "AI Chatbot integration in SME marketing platforms: Improving customer interaction and service efficiency," *International Journal of Management & Entrepreneurship Research*, vol. 6, no. 7, pp. 2332– 2341, Jul. 2024, doi: https://doi.org/10.51594/ijmer.v6i7.1327

[52] F. Dwitama and A. Rusli, "User stories collection via interactive chatbot to support requirements gathering," *TELKOMNIKA (Telecommunication Computing Electronics and Control)*, vol. 18, no. 2, p. 890, Apr. 2020, doi: https://doi.org/10.12928/telkomnika.v18i2.14866

[53] J. Benzinho *et al.*, "LLM Based Chatbot for Farm-to-Fork Blockchain Traceability Platform," *Applied Sciences*, vol. 14, no. 19, p. 8856, Oct. 2024, doi: https://doi.org/10.3390/app14198856

[54] G. P. Reddy and V. Pavan, "Explainable AI (XAI): Explained," pp. 1–6, Apr. 2023, doi: https://doi.org/10.1109/estream59056.2023.10134984

[55] S. Kayode, "Blockchain and AI Chatbots: A Synergistic Approach to Transparent Supply Chains," *SSRN Electronic Journal*, 2025, doi: https://doi.org/10.2139/ssrn.5151745

[56] A Bold, "Security of Decentralized Fintech: A Blockchain and Artificial Intelligence Concept," *ResearchGate*, Mar.15,2023 https://www.researchgate.net/publication/369239823_Security_of_Decentralized_Fintech_A_Blockchain_and_Artificial_Intelligence_Concept?enrichId=rgreq- c71dd78520f788d2c59e64837673471 6-XXX&enrichSource=Y292ZXJQYWdlOzM2OTIzOTgyMztBUzoxMTQzMTI4MTEyNjkwNDk5OEAxNjc4ODgzMjYzOTUx&el=1_x_2&_esc=publicationCoverPdf (accessed Dec. 20, 2025).

[57] F. Mustakim, Fauziah, and N. Hayati, "Algoritma Artificial Neural Network pada Text-based Chatbot Frequently Asked Question (FAQ) Web Kuliah Universitas Nasional," *Jurnal Teknologi Informasi dan Komunikasi (JTIK)*, vol. 5, no. 4, pp. 438–446, Oct. 2021, doi: https://doi.org/10.35870/jti

[58] A. Mansurova, A. Nugumanova, and Z. Makhambetova, "DEVELOPMENT OF A QUESTION ANSWERING CHATBOT FOR BLOCKCHAIN DOMAIN," *Scientific Journal of Astana IT University*, pp. 27–40, Sep. 2023, doi: https://doi.org/10.37943/15XNDZ6667

[59] R. Fadhilah, M. R. Maulani, W. Resdiana, and D. Hamidin, "INTEGRASI FITUR CHATBOT DALAM APLIKASI EDUKASI KESEHATAN DAN KEBUGARAN MENGGUNAKAN ALGORITMA NEURAL NETWORK," *Jurnal Kecerdasan Buatan dan Teknologi Informasi*, vol. 3, no. 3, pp. 125–135, Aug. 2024, doi: https://doi.org/10.69916/jkbti.v3i3.156

[60] J. Kuai, C. Brantner, M. Karlsson, E. V. Couvering, and S. Romano, "AI chatbot accountability in the age of algorithmic gatekeeping: Comparing generative search engine political information retrieval across five languages," *New Media & Society*, Feb. 2025, doi: https://doi.org/10.1177/14614448251321162

[61] Nika Mozafari, W. H. Weiger, and M. Hammerschmidt, "Resolving the Chatbot Disclosure Dilemma: Leveraging Selective Self-Presentation to Mitigate the Negative Effect of Chatbot Disclosure," *Proceedings of the ... Annual Hawaii International Conference on System Sciences/Proceedings of the Annual Hawaii International Conference on System Sciences*, Jan. 2021, doi: https://doi.org/10.24251/hicss.2021.355

[62] V. Jonatan and A.-A. Igor, "Creation Of A ChatBot Based On Natural Language Proccesing For Whatsapp," *arXiv.org*, 2023. https://arxiv.org/abs/2310.10675 (accessed Dec. 20, 2025).

[63] D. E. Mathew, D. U. Ebem, Anayo Chukwu Ikegwu, P. E. Ukeoma, and Ngozi Fidelia Dibiaezue, "Recent Emerging Techniques in Explainable Artificial Intelligence to Enhance the Interpretable and Understanding of AI Models for Human," *Neural Processing Letters*, vol. 57, no. 1, Feb. 2025, doi: https://doi.org/10.1007/s11063-025-11732-2

[64] Malvin, C. Dylan, and A. Rangkuti, "WhatsApp Chatbot Customer Service Using Natural Language Processing and Support Vector Machine," *International Journal of Emerging Technology and Advanced Engineering Website: www.ijetae.com*, vol. 9001, no. 03, 2008, doi: https://doi.org/10.46338/ijetae0322_15

[65] walaa hassan and ahmed elsayed, "An Interactive Chatbot for College Enquiry," *Journal of Computing and Communication*, vol. 2, no. 1, pp. 20–28, Jan. 2023, doi: https://doi.org/10.21608/jocc.2023.282081

[66] F. E. Roch *et al.*, "Diagnosis, treatment, and prevention of ankle sprains: Comparing free chatbot recommendations with clinical guidelines," *Foot and Ankle Surgery*, vol. 31, no. 4, pp. 329–351, Jun. 2025, doi: https://doi.org/10.1016/j.fas.2024.12.003

[67] M. Erfan Rianto and Ainul Furqon, "Impelementasi AI Chatbot Sebagai Support Assistant Website Universitas Nurul Jadid Menggunakan Algoritma BiLSTM," *Najah: Journal of Research and Community Service*, vol. 2, no. 3, pp. 15–24, 2024, Accessed: Dec. 21, 2025. [Online]. Available: https://kalamnusantara.org/index.php/najah/article/view/62

[68] A. Joukhadar, H. Saghergy, L. Kweider, and N. Ghneim, "Arabic Dialogue Act Recognition for Textual Chatbot Systems." Accessed: Dec. 21, 2025. [Online]. Available: https://aclanthology.org/2019.nsurl-1.7.pdf

[69] M.-H. Hsu, T.-M. Chan, and C.-S. Yu, "Termbot: A Chatbot-Based Crossword Game for Gamified Medical Terminology Learning," *International Journal of Environmental Research and Public Health*, vol. 20, no. 5, p. 4185, Feb. 2023, doi: https://doi.org/10.3390/ijerph20054185

[70] M. Attalariq and Z. K. A. Baizal, "Chatbot-Based Book Recommender System Using Singular Value Decomposition," *Journal of Information System Research (JOSH)*, vol. 4, no. 4, pp. 1293–1301, Jul. 2023, doi: https://doi.org/10.47065/josh.v4i4.3817

[71] J. Fleiß, E. Bäck, and S. Thalmann, "Mitigating algorithm aversion in recruiting: A study on explainable AI for conversational agents," *The DATA BASE for Advances in Information Systems*, vol. [forthcoming], 2023. doi: https://dl.acm.org/doi/abs/10.1145/3645057.3645062

[72] E. Kagan, M. Dada, and B. Hathaway, "AI Chatbots in Customer Service: Adoption Hurdles and Simple Remedies," *SSRN Electronic Journal*, 2022, doi: https://doi.org/10.2139/ssrn.4283285

[73] Ronak Surve, T. Purohit, R. Joseph, and P. Shaikh, "HealthCare Chatbot Using Machine Learning and NLP," pp. 411–416, Dec. 2023, doi: https://doi.org/10.1109/icast59062.2023.10455027

[74] D. Gunawan, Farica Perdana Putri, and Hira Meidia, "Bershca: bringing chatbot into hotel industry in Indonesia," *TELKOMNIKA (Telecommunication Computing Electronics and Control)*, vol. 18, no. 2, pp. 839–839, Mar. 2020, doi: https://doi.org/10.12928/telkomnika.v18i2.14841

[75] T.-J. Ng, K.-W. Ng, and S.-C. Haw, "Lib-Bot: A Smart Librarian-Chatbot Assistant," *International journal of computing and digital system/International Journal of Computing and Digital Systems*, vol. 15, no. 1, pp. 1–11, Jul. 2024, doi: https://doi.org/10.12785/ijcds/160101

[76] "Medical Chatbot (Medibot)," *International Journal of Advanced Trends in Computer Science and Engineering*, vol. 10, no. 3, pp. 2171–2174, Jun. 2021, doi: https://doi.org/10.30534/ijatcse/2021/941032021

[77] Haeruddin Haeruddin, Sabariman Sabariman, and V. Su, "Designing a Chatbot Application Using the Flask Framework and Rule-Based Algorithm," *Jurnal Teknologi Dan Sistem Informasi Bisnis*, vol. 7, no. 1, pp. 133–42, Jan. 2025, doi: https://doi.org/10.47233/jteksis.v7i1.1820

[78] N. Rane, S. Choudhary, and J. Rane, "Explainable Artificial Intelligence (XAI) in healthcare: Interpretable Models for Clinical Decision Support," *SSRN Electronic Journal*, 2023, doi: https://doi.org/10.2139/ssrn.4637897

[79] F. E. Roch *et al.*, "Diagnosis, treatment, and prevention of ankle sprains: Comparing free chatbot recommendations with clinical guidelines," *Foot and Ankle Surgery*, vol. 31, no. 4, pp. 329–351, Dec. 2024, doi: https://doi.org/10.1016/j.fas.2024.12.003

[80] S. Karia, M. Mehta, K. Konar, and Nirali Kabli, "BankBot: Contactless Machine Learning Chatbot for Communication during COVID-19 in Bank," *SSRN Electronic Journal*, Jan. 2020, doi: https://doi.org/10.2139/ssrn.3747951

[81] T. Ahmad, Pranadeep Katari, Ashok, C. S. Ravi, and M. Shaik, "Explainable AI: Interpreting Deep Learning Models for Decision Support," *Advances in Deep Learning Techniques*, vol. 4, no. 1, pp. 80–108, 2024, Accessed: Dec. 22, 2025. [Online]. Available: https://www.thesciencebrigade.org/adlt/article/view/328

[82] D. Ramachandram, H. Joshi, J. Zhu, D. Gandhi, L. Hartman, and A. Raval, "Transparent AI: The Case for Interpretability and Explainability," *arXiv.org*, 2025. https://arxiv.org/abs/2507.23535 (accessed Dec. 20, 2025).

[83] N. Rane, S. Choudhary, and J. Rane, "Explainable Artificial Intelligence (XAI) Approaches for Transparency and Accountability in Financial Decision-Making," *SSRN Electronic Journal*, Jan. 2023, doi: https://doi.org/10.2139/ssrn.4640316

[84] M. Dixit, I. Kansal, V. Khullar, R. Kumar, and S. Kumar, "Analyzing Trustworthiness and Explainability in Artificial Intelligence: A Comprehensive Review," *Recent Advances in Electrical & Electronic Engineering (Formerly Recent Patents on Electrical & Electronic Engineering)*, vol. 18, no. 8, Jul. 2024, doi: https://doi.org/10.2174/0123520965308169240616144800

[85] F. Herrera, "Making Sense of the Unsensible: Reflection, Survey, and Challenges for XAI in Large Language Models Toward Human-Centered AI," *arXiv.org*, 2025. https://arxiv.org/abs/2505.20305 (accessed Dec. 20, 2025).

[86] Madhan Veeramani, P. Karthick, S. Venkateswaran, B. Sriman, Shaik Thasleem Bhanu, and V. S. Devi, "Transparency in Text," pp. 131–160, Mar. 2025, doi: https://doi.org/10.1002/9781394249312.ch7

[87] J. Černevičienė and A. Kabašinskas, "Review of Multi-Criteria Decision-Making Methods in Finance Using Explainable Artificial Intelligence," *Frontiers in Artificial Intelligence*, vol. 5, Mar. 2022, doi: https://doi.org/10.3389/frai.2022.827584

[88] M. Mesinovic, P. Watkinson, and T. Zhu, "Explainable AI for clinical risk prediction: a survey of concepts, methods, and modalities," *arXiv.org*, 2023. https://arxiv.org/abs/2308.08407 (accessed Dec. 20, 2025).

[89] P. Gohel, P. Singh, and M. Mohanty, "Explainable AI: current status and future directions," *arXiv.org*, 2021. doi : https://doi.org/10.48550/arXiv.2107.07045

[90] T. Zhang, T. Chung, A. Dey, and S. W. Bae, "AXAI-CDSS : An Affective Explainable AI-Driven Clinical Decision Support System for Cannabis Use," *arXiv.org*, 2025. doi : https://doi.org/10.48550/arXiv.2503.06463

[91] S. Lopes, M. Mascarenhas, J. Fonseca, and A. F. Leite-Moreira, "Unveiling the Algorithm: The Role of Explainable Artificial Intelligence in Modern Surgery," *Healthcare*, vol. 13, no. 24, pp. 3208–3208, Dec. 2025, doi: https://doi.org/10.3390/healthcare13243208

[92] N. Nobel *et al.*, "Unmasking Banking Fraud: Unleashing the Power of Machine Learning and Explainable AI (XAI) on Imbalanced Data," *Information*, vol. 15, no. 6, pp. 298–298, May 2024, doi: https://doi.org/10.3390/info15060298

[93] R. Karim *et al.*, "Explainable AI for Bioinformatics: Methods, Tools and Applications," *Briefings in Bioinformatics*, Jul. 2023, doi: https://doi.org/10.1093/bib/bbad236

[94] M. Mascarenhas *et al.*, "Explainable AI in Digestive Healthcare and Gastrointestinal Endoscopy," *Journal of Clinical Medicine*, vol. 14, no. 2, pp. 549–549, Jan. 2025, doi: https://doi.org/10.3390/jcm14020549

[95] S. Daram, "Explainable AI in Healthcare: Enhancing Trust, Transparency, and Ethical Compliance in Medical AI Systems," *International Journal of AI, BigData, Computational and Management Studies*, vol. 6, pp. 11–20, 2025, doi: https://doi.org/10.63282/3050-9416.ijaibdcms-v6i2p102

[96] T. Niu, T. Liu, Y. T. Luo, P. C.-I. Pang, S. Huang, and A. Xiang, "Decoding student cognitive abilities: a comparative study of explainable AI algorithms in educational data mining," *Scientific Reports*, vol. 15, no. 1, pp. 26862–26862, Jul. 2025, doi: https://doi.org/10.1038/s41598-025-12514-5

[97] Aditya Mehra , "Unifying Adversarial Robustness and Interpretability in Deep Neural Networks: A Comprehensive Framework for Explainable and Secure Machine Learning Models," *International Research Journal of Modernization in Engineering Technology and Science*, Nov. 2024, doi: https://doi.org/10.56726/irjmets4109

[98] Mohammad N.S. Jahromi, S. M. Muddamsetty, Asta, Anna Murphy Høgenhaug, T. Gammeltoft-Hansen, and T. B. Moeslund, "SIDU-TXT: An XAI algorithm for NLP with a holistic assessment approach," *Natural Language Processing Journal*, vol. 7, pp. 100078–100078, Jun. 2024, doi: https://doi.org/10.1016/j.nlp.2024.100078

[99] H. Tatsat and A. Shater, "Beyond the Black Box: Interpretability of LLMs in Finance," *arXiv.org*, 2025. doi : https://doi.org/10.48550/arXiv.2505.24650

[100] M. Mersha, M. Bitewa, T. Abay, and J. Kalita, "Explainability in Neural Networks for Natural Language Processing Tasks," *arXiv.org*, 2024. doi : https://doi.org/10.48550/arXiv.2412.18036

[101] J. C. L. Chow, "Quantum Computing and Machine Learning in Medical Decision-Making: A Comprehensive Review," *Algorithms*, vol. 18, no. 3, p. 156, Mar. 2025, doi: https://doi.org/10.3390/a18030156

[102] H.-Y. Chen, C. Sharma, S. Sharma, K. Sharma, and G. K. Sethi, "Intellectual Structure of Explainable Artificial Intelligence: a Bibliometric Reference to Research Constituents," Oct. 2023, doi: https://doi.org/10.21203/rs.3.rs-3493299/v1

[103] T. Zhang, M. Zhang, W. Y. Low, X. J. Yang, and B. A. Li, "Conversational Explanations: Discussing Explainable AI with Non-AI Experts," pp. 409–424, Mar. 2025, doi: https://doi.org/10.1145/3708359.3712143

[104] G. He, Nilay Aishwarya, and Ujwal Gadiraju, "Is Conversational XAI All You Need? Human-AI Decision Making With a Conversational XAI Assistant," *arXiv (Cornell University)*, pp. 907–924, Mar. 2025, doi: https://doi.org/10.1145/3708359.3712133

[105] Andi Nurdin, Dhian Satria Yudha Kartika, and A. Rezha, "KLASIFIKASI PENYAKIT DAUN TOMAT DENGAN METODE CONVOLUTIONAL NEURAL NETWORK MENGGUNAKAN ARSITEKTUR INCEPTION-V3," *JURNAL ILMIAH INFORMATIKA*, vol. 12, no. 02, pp. 114–119, Sep. 2024, doi: https://doi.org/10.33884/jif.v12i02.9162

[106] Anindita Pratita, Tri Lathif Mardi Suryanto, Arista Pratama, and Adi Wibowo, "ChatGPT in Education: Investigating Students Online Learning Behaviors," *International Journal of Information and Education Technology*, vol. 15, no. 3, pp. 510–524, 2025, doi: https://doi.org/10.18178/ijiet.2025.15.3.2262

[107] T. L. M. Suryanto, A. P. Wibawa, H. Hariyono, and A. Nafalski, "Evolving Conversations: A Review of Chatbots and Implications in Natural Language Processing for Cultural Heritage Ecosystems," *International Journal of Robotics and Control Systems*, vol. 3, no. 4, pp. 955–1006, Dec. 2023, doi: https://doi.org/10.31763/ijrcs.v3i4.1195

[108] Hu, X., Liu, A., & Dai, Y. (2025). Combining ChatGPT and knowledge graph for explainable machine learning-driven design: a case study. Journal of Engineering Design, 36(7-9), 1479-1501. https://doi.org/10.1080/09544828.2024.2355758

[109] Wang, Q., Anikina, T., Feldhus, N., van Genabith, J., Hennig, L., & Möller, S. (2024). LLMCheckup: Conversational examination of large language models via interpretability tools and self-explanations. In Proceedings of the Third Workshop on Bridging Human–Computer Interaction and Natural Language Processing (3rd HCI+NLP Workshop) (pp. 89-104). Association for Computational Linguistics. https://doi.org/10.18653/v1/2024.hcinlp-1.9

[110] John, B. (2025). The future of AI-powered chatbots: Enhancing user experience while mitigating privacy risks

[111] Guttikonda, D., Indran, D., Narayanan, L., Pasarad, T., & Sandesh, B. J. (2025). Explainable AI: A retrieval-augmented generation based framework for model interpretability. In Proceedings of the 17th International Conference on Agents and Artificial Intelligence (ICAART 2025) - Volume 3 (pp. 948-955). https://doi.org/10.5220/0013241300003890