



Development and Preliminary Evaluation of a Dayak Paser Culture-Contextualized Mathematics Chatbot for Fraction Learning among Elementary School Students in Balikpapan

Pengembangan dan Evaluasi Awal Chatbot Matematika Berkonteks Budaya Dayak Paser untuk Pembelajaran Pecahan Siswa Sekolah Dasar di Balikpapan

Junaldelvia Tandungan¹, Irma Soleha², Helga Apriana Renleeuw³

Politeknik Ilmu Pelayaran Balikpapan, Indonesia

junaldelviatandungan@pip-balikpapan.ac.id¹; irmasoleha@pip-balikpapan.ac.id²; helgaarenleeuw@pip-balikpapan.ac.id³

Correspondence author Email: junaldelviatandungan@pip-balikpapan.ac.id

Paper received: Juli-2026; Accepted: Agustus-2026; Publish: Agustus-2026

Abstract

Fraction concepts are challenging for many elementary school students because they require coordination among symbolic, visual, and contextual representations. This study developed and conducted a preliminary evaluation of a large-language-model-based mathematics chatbot that incorporates selected Dayak Paser cultural contexts for fraction learning in Balikpapan. The study used a Research and Development approach organized with the ADDIE framework and a one-group pretest-posttest evaluation involving 60 fourth-grade students. Evidence comprised expert product validation, student responses, pretest-posttest scores, the System Usability Scale (SUS), a descriptive review of 120 chatbot answers, and latency records from 100 interactions. The product-validation dimensions were rated 92%-94%, with an overall score of 93.4%. Positive student-response percentages ranged from 92% to 95%, and the SUS score was 87.6. Mean test scores were 55.3 before and 84.7 after use, a within-group difference of 29.4 points. The normalized gain calculated from the reported group means was 0.66 (moderate). The reported paired-samples result was $t(59)=18.93$, $p<.001$; the corresponding paired standardized change was $d_z=2.44$ and the reconstructed 95% confidence interval for the mean difference was 26.29-32.51. These findings describe a substantial pre-post change but do not establish that the chatbot caused it because the design did not include a comparison group, randomization, a retention test, or a documented intervention dose. The study therefore supports product feasibility and warrants a controlled follow-up, while making no empirical claim regarding adaptive learning, increased motivation, cultural preservation, or retrieval-augmented generation

Keywords: artificial intelligence; chatbot; ethnomathematics; Dayak Paser; fractions; elementary education; preliminary evaluation.

Abstrak

Konsep pecahan menantang bagi banyak siswa sekolah dasar karena menuntut koordinasi representasi simbolik, visual, dan kontekstual. Penelitian ini mengembangkan dan mengevaluasi secara awal chatbot matematika berbasis Large Language Model yang memuat konteks budaya Dayak Paser terpilih untuk pembelajaran pecahan di Balikpapan. Penelitian menggunakan pendekatan Research and Development yang diorganisasikan melalui kerangka ADDIE serta evaluasi satu kelompok pretest-posttest terhadap 60 siswa kelas IV. Bukti penelitian mencakup validasi produk oleh ahli, respons siswa, skor pretest-posttest, System Usability Scale (SUS), telaah deskriptif terhadap 120 jawaban chatbot, dan catatan latensi dari 100 interaksi. Dimensi validasi produk memperoleh skor 92%-94% dengan rata-rata 93,4%. Persentase respons positif siswa berada pada rentang 92%-



95%, sedangkan skor SUS mencapai 87,6. Rata-rata skor tes berubah dari 55,3 sebelum penggunaan menjadi 84,7 setelah penggunaan, atau berbeda 29,4 poin dalam kelompok yang sama. N-Gain yang dihitung dari rata-rata kelompok adalah 0,66 (kategori sedang). Hasil uji berpasangan yang dilaporkan ialah $t(59)=18,93$, $p<0,001$; besaran perubahan terstandar berpasangan yang bersesuaian adalah $d_z=2,44$ dengan interval kepercayaan 95% hasil rekonstruksi 26,29-32,51 untuk selisih rata-rata. Temuan tersebut menggambarkan perubahan pre-post yang besar, tetapi tidak membuktikan bahwa chatbot menjadi penyebabnya karena penelitian tidak menggunakan kelompok pembanding, randomisasi, tes retensi, maupun dokumentasi dosis intervensi. Karena itu, penelitian mendukung kelayakan awal produk dan kebutuhan studi terkontrol berikutnya, tanpa mengajukan klaim empiris mengenai pembelajaran adaptif, peningkatan motivasi, pelestarian budaya, atau penerapan Retrieval-Augmented Generation

Kata kunci: kecerdasan buatan; chatbot; etnomatematika; Dayak Paser; pecahan; sekolah dasar; evaluasi awal.

Copyright and License

Authors retain copyright and grant the journal right of first publication with the work simultaneously licensed under a Creative Commons Attribution 4.0 International License that allows others to share the work with an acknowledgment of the work's authorship and initial publication in this journal.



1. Pendahuluan

Pemahaman pecahan pada jenjang sekolah dasar tidak hanya menuntut kemampuan menghitung, tetapi juga kemampuan menghubungkan bagian-keseluruhan, perbandingan, representasi visual, dan simbol matematika. Media percakapan berbasis kecerdasan buatan berpotensi memberi kesempatan kepada siswa untuk bertanya dan menerima penjelasan secara langsung. Namun, potensi tersebut tidak dengan sendirinya membuktikan efektivitas pembelajaran. Large Language Model (LLM) dapat menghasilkan respons yang lancar tetapi tidak selalu akurat, sehingga desain pedagogis, pembatasan konten, evaluasi terkontrol, dan perlindungan peserta didik harus dilaporkan secara jelas (Kasneci et al., 2023; UNESCO, 2023).

Literatur yang lebih luas menunjukkan bahwa chatbot pendidikan dapat mendukung interaksi, akses informasi, dan pengalaman belajar, tetapi manfaatnya tetap bergantung pada desain serta konteks implementasi (Okonkwo & Ade-Ibijola, 2021). Tinjauan AIED juga menempatkan peran AI pada pembelajaran, pengajaran, asesmen, dan administrasi sekaligus mencatat tantangan implementasi yang perlu dikendalikan (Chiu et al., 2023). Pada generative



AI, Tlili et al. (2023) menekankan bahwa pemanfaatan chatbot perlu menyeimbangkan peluang pembelajaran dengan risiko dan pengawasan manusia.

Etnomatematika menawarkan cara untuk menghubungkan ide matematika dengan praktik, artefak, dan pengalaman budaya. Dalam penelitian ini, konteks Dayak Paser dimasukkan melalui contoh pembagian hasil, pola anyaman, permainan tradisional, ornamen rumah, dan narasi lokal yang dapat merepresentasikan konsep bagian-keseluruhan. Penggunaan konteks tersebut diposisikan sebagai keputusan desain pembelajaran, bukan sebagai bukti bahwa siswa memiliki identitas Dayak Paser atau bahwa produk telah menghasilkan pelestarian budaya. Kedua luaran tersebut tidak diukur dalam penelitian ini.

Bukti mutakhir menunjukkan bahwa kesimpulan mengenai manfaat tutor AI sangat dipengaruhi oleh rancangan evaluasi. Tinjauan sistematis Létourneau et al. (2025) menemukan hasil yang umumnya positif pada pendidikan K-12, tetapi menekankan perlunya intervensi yang lebih panjang, sampel yang lebih beragam, kontrol eksperimen yang lebih baik, dan perhatian terhadap etika. Kestin et al. (2025) dapat menyampaikan klaim komparatif karena menggunakan uji acak terkontrol dengan kondisi pembandingan dan pencatatan waktu belajar. Standar tersebut berbeda dari desain satu kelompok dalam penelitian ini. Karena itu, bahasa kausal seperti “chatbot meningkatkan”, “terbukti efektif”, atau “memberikan pengaruh” tidak digunakan dalam versi revisi ini.

Penelitian ini bertujuan: (1) mendeskripsikan pengembangan chatbot matematika yang menggunakan konteks Dayak Paser untuk materi pecahan; (2) menilai kelayakan awal produk melalui validasi, respons siswa, dan usability; (3) mendeskripsikan perubahan skor pemahaman pecahan sebelum dan sesudah penggunaan pada kelompok yang sama; serta (4) menyajikan evaluasi teknis awal yang dapat ditelusuri dari data yang tersedia. Kontribusi penelitian dibatasi pada integrasi LLM, konten pecahan, dan konteks budaya lokal dalam satu prototipe pada satu konteks pendidikan dasar di Balikpapan.



1.1 State of the Art dan Posisi Penelitian

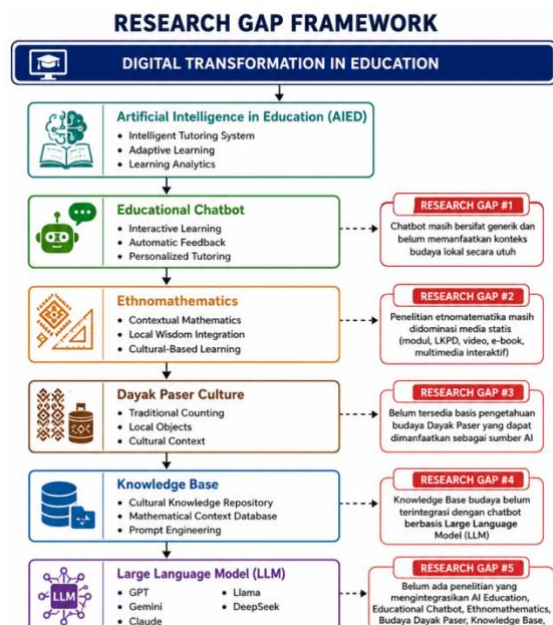
Studi	Kontribusi yang relevan	Posisi terhadap penelitian ini
Rosa & Orey (2020, 2022)	Etnomatematika dan pendidikan matematika relevan secara budaya.	Tidak mengevaluasi chatbot LLM pada konteks Dayak Paser.
Kasneci et al. (2023)	Peluang dan risiko LLM untuk pendidikan; menekankan pedagogi dan tata kelola.	Tidak berfokus pada pecahan SD atau konteks budaya lokal Indonesia.
Lo (2023)	Tinjauan cepat dampak ChatGPT dalam pendidikan.	Bukti lintas konteks; bukan evaluasi terkontrol chatbot etnomatematika.
Létourneau et al. (2025)	Tinjauan 28 studi ITS K-12; menekankan kontrol, durasi, keragaman sampel, dan etika.	Menjadi acuan pembatasan interpretasi desain satu kelompok.
Kestin et al. (2025)	RCT tutor AI pada pendidikan tinggi dengan kondisi pembandingan dan waktu belajar terdokumentasi.	Populasi dan desain berbeda; digunakan sebagai benchmark klaim kausal.
Thesen & Park (2025)	RAG dengan materi kurasi, arsitektur, penggunaan, dan trade-off reliabilitas-cakupan yang dilaporkan.	RAG tidak terdokumentasi secara replikabel pada penelitian ini.
Okonkwo & Ade-Ibijola (2021)	Tinjauan sistematis 53 studi aplikasi chatbot dalam pendidikan; memetakan manfaat dan tantangan.	Mendukung rasional chatbot pendidikan, tetapi tidak spesifik pada pecahan SD, LLM, atau Dayak Paser.
Chiu et al. (2023)	Tinjauan sistematis AIED; mengidentifikasi peran AI dalam pembelajaran dan tantangan implementasi.	Menjadi acuan desain dan kehati-hatian evaluasi; tidak menguji prototipe etnomatematika.
Tlili et al. (2023)	Studi kasus ChatGPT dalam pendidikan; menyoroti potensi, risiko, dan kolaborasi manusia-AI.	Relevan untuk tata kelola generative AI; tidak mengevaluasi chatbot pecahan pada siswa SD.
Penelitian ini	Prototipe LLM untuk pecahan dengan konteks Dayak Paser dan evaluasi awal pada 60 siswa.	Tidak menguji kausalitas, adaptivitas, motivasi, hasil budaya, atau retensi.

1.2 Research Gap dan Kebaruan yang Terukur

Kesenjangan yang ditangani ialah masih terbatasnya laporan pengembangan chatbot LLM untuk pembelajaran pecahan sekolah dasar yang memanfaatkan konteks Dayak Paser dan sekaligus menyajikan validasi produk, pengalaman pengguna, perubahan pre-post, serta evaluasi teknis awal. Kebaruan tidak dinyatakan sebagai “menutup seluruh research gap”. Kebaruan yang dapat dipertanggungjawabkan adalah konfigurasi desain pada satu konteks



sekolah dasar di Balikpapan: LLM, materi pecahan, dan konten budaya Dayak Paser terpilih dipadukan dalam sebuah prototipe percakapan. Klaim mengenai personalisasi atau adaptivitas tidak diajukan karena mekanisme learner model, perubahan respons berbasis performa, dan evaluasi adaptivitas tidak tersedia.



Gambar 1 Kesenjangan dan Posisi Penelitian

Berdasarkan literatur yang ditelaah, terdapat ruang penelitian pada pengembangan chatbot berbasis *Large Language Model* (LLM) yang secara khusus menggabungkan materi pecahan sekolah dasar dengan konteks budaya lokal. Kesenjangan tersebut dikelompokkan ke dalam empat aspek, yaitu integrasi konseptual, desain pedagogis, evaluasi produk, dan implementasi LLM pada pendidikan dasar.

Pertama, kesenjangan integrasi konseptual. Kajian mengenai *Artificial Intelligence in Education* (AIED), chatbot berbasis LLM, dan etnomatematika umumnya berkembang dalam fokus yang berbeda. Literatur AIED dan LLM banyak membahas potensi, risiko, serta penggunaan sistem percakapan dalam pembelajaran, sedangkan etnomatematika menekankan hubungan antara konsep matematika dan konteks budaya. Dalam literatur yang ditelaah, laporan mengenai prototipe yang memadukan LLM, pembelajaran pecahan sekolah dasar, dan



konteks Dayak Paser masih terbatas. Penelitian ini menempatkan ketiga unsur tersebut dalam satu desain prototipe pembelajaran.

Kedua, kesenjangan desain pedagogis. Penggunaan konteks budaya dalam media pembelajaran tidak selalu disertai fasilitas dialog yang memungkinkan siswa mengajukan pertanyaan dan memperoleh penjelasan secara langsung. Penelitian ini merespons kebutuhan tersebut melalui antarmuka percakapan yang menyajikan contoh dan latihan pecahan menggunakan konteks Dayak Paser terpilih. Dialog dan umpan balik diposisikan sebagai bentuk *scaffolding* dalam desain pembelajaran, bukan sebagai bukti bahwa sistem telah melakukan personalisasi atau adaptasi berdasarkan karakteristik dan performa setiap siswa.

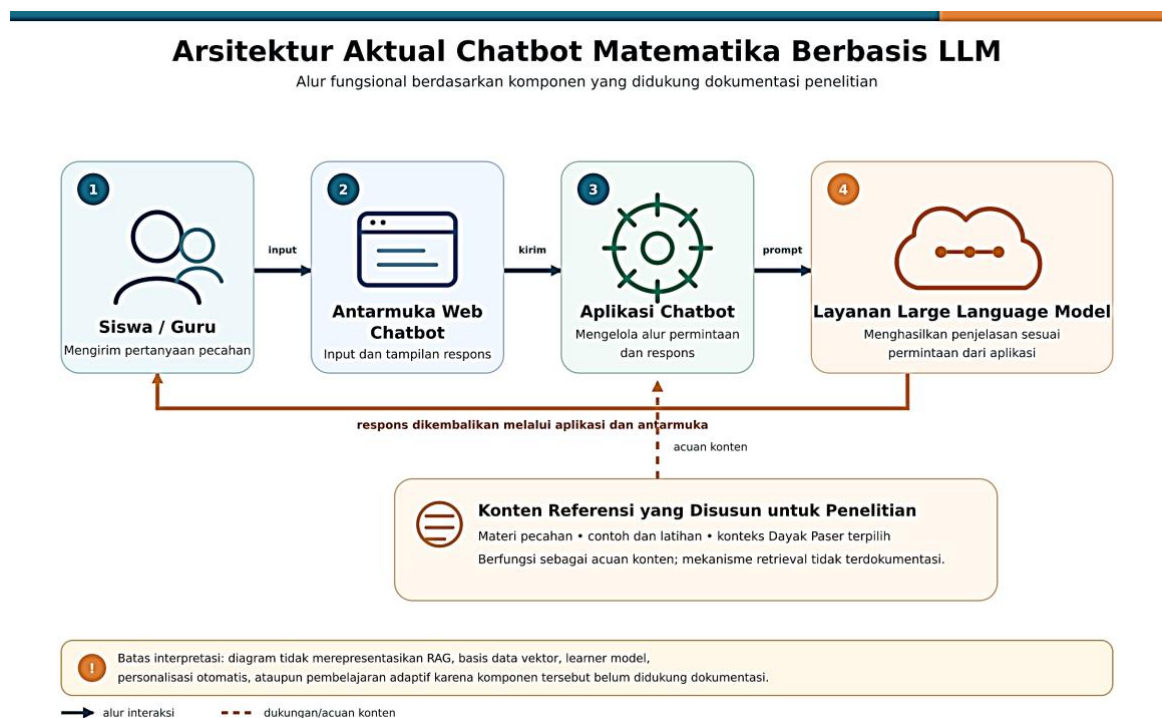
Ketiga, kesenjangan evaluasi produk. Evaluasi chatbot pendidikan memerlukan bukti yang mencakup kelayakan produk, pengalaman pengguna, perubahan hasil belajar, dan kinerja teknis awal. Penelitian ini menyajikan validasi produk, respons siswa, *System Usability Scale* (SUS), perbandingan skor *pretest-posttest* pada kelompok yang sama, telaah deskriptif terhadap jawaban chatbot, serta catatan latensi interaksi. Namun, karena evaluasi menggunakan desain satu kelompok tanpa kelompok pembanding, hasil penelitian hanya menunjukkan perubahan skor sebelum dan setelah penggunaan. Temuan tersebut tidak membuktikan bahwa perubahan terjadi semata-mata karena chatbot.

Keempat, kesenjangan implementasi dan pelaporan LLM pada pendidikan dasar. Penerapan LLM untuk pembelajaran pecahan sekolah dasar dengan konteks budaya lokal masih memerlukan pengembangan dan evaluasi lebih lanjut. Selain itu, implementasi LLM perlu disertai dokumentasi mengenai model, *prompt*, parameter generasi, sumber konten, mekanisme keselamatan, dan prosedur evaluasi respons agar dapat direplikasi. Penelitian ini menghasilkan prototipe berbasis LLM dengan konten referensi yang disusun untuk pembelajaran pecahan. Akan tetapi, karena dokumentasi teknis yang tersedia belum menjelaskan mekanisme *retrieval* secara memadai, penelitian ini tidak mengklaim telah menerapkan *Retrieval-Augmented Generation* (RAG).



Berdasarkan keempat aspek tersebut, penelitian ini mengembangkan dan melakukan evaluasi awal terhadap chatbot matematika berbasis LLM yang memadukan materi pecahan dengan konteks budaya Dayak Paser terpilih. Kontribusi penelitian dibatasi pada konfigurasi desain prototipe dan bukti kelayakan awal pada satu konteks pendidikan dasar di Balikpapan. Penelitian ini tidak mengajukan klaim mengenai efektivitas kausal, pembelajaran adaptif, peningkatan motivasi, autentisitas budaya yang telah divalidasi komunitas, maupun pelestarian budaya.

Produk yang dikembangkan berupa chatbot berbasis web yang menggunakan *Large Language Model* untuk menghasilkan respons mengenai materi pecahan. Sistem menghubungkan antarmuka percakapan siswa dengan layanan aplikasi, model bahasa, dan konten referensi yang disusun untuk penelitian. Alur fungsional sistem yang dapat diidentifikasi berdasarkan dokumentasi penelitian ditunjukkan pada Gambar 2



Gambar 2. Arsitektur Aktual Chatbot Matematika Berbasis LLM



Arsitektur tersebut menggambarkan aliran interaksi dari pengguna melalui antarmuka chatbot menuju aplikasi yang meneruskan pertanyaan kepada layanan LLM. Respons dihasilkan dengan merujuk pada materi pecahan dan konteks Dayak Paser terpilih yang disusun untuk penelitian, kemudian ditampilkan kembali kepada pengguna. Dokumentasi yang tersedia belum menjelaskan identitas dan versi model, *system prompt*, parameter generasi, mekanisme *embedding*, *vector database*, maupun proses *retrieval*. Oleh karena itu, arsitektur ini tidak menyatakan adanya personalisasi, *learner model*, pembelajaran adaptif, ataupun penerapan *Retrieval-Augmented Generation* (RAG).

2. Landasan Konseptual dan Pemetaan Operasional

Versi awal menggunakan lima teori secara paralel, tetapi tidak seluruhnya dioperasionalkan atau diukur. Versi revisi membatasi kerangka pada dua prinsip desain: konstruktivisme sebagai dasar pemberian kesempatan eksplorasi dan scaffolding melalui dialog (Piaget, 1972; Vygotsky, 1978), serta situated learning/etnomatematika sebagai dasar penggunaan konteks lokal (Lave & Wenger, 1991; Rosa & Orey, 2020, 2022). Teori-teori tersebut berfungsi sebagai rasional desain, bukan variabel yang diuji.

Prinsip	Keputusan desain	Bukti yang tersedia	Batas inferensi
Konstruktivisme dan scaffolding	Dialog tanya-jawab, contoh bertahap, umpan balik segera.	Respons siswa, SUS, perubahan skor pre-post.	Tidak membuktikan personalisasi, adaptivitas, atau reduksi miskonsepsi individual.
Situated learning dan etnomatematika	Contoh pecahan memakai konteks Dayak Paser terpilih.	Validasi materi dan respons siswa terhadap kemudahan/ketertarikan.	Tidak mengukur identitas budaya, pengetahuan budaya, autentisitas komunitas, atau pelestarian budaya.

3. Metode

3.1 Desain Penelitian

Penelitian menggunakan Research and Development dengan kerangka ADDIE (Analysis, Design, Development, Implementation, Evaluation) (Branch, 2009). Evaluasi hasil belajar menggunakan desain satu kelompok pretest-posttest. Desain tersebut dapat menunjukkan apakah skor setelah penggunaan berbeda dari skor sebelum penggunaan pada peserta yang sama, tetapi tidak dapat mengisolasi kontribusi chatbot dari maturasi, pengulangan tes, fasilitasi guru, latihan tambahan, novelty effect, atau faktor pembelajaran lain.



3.2 Tahapan ADDIE dan Bukti Pelaksanaan

Tahap	Kegiatan yang terdokumentasi	Bukti/keluaran dan batas pelaporan
Analysis	Observasi pembelajaran, wawancara guru kelas IV, telaah Kurikulum Merdeka, dan identifikasi kebutuhan materi pecahan serta kesiapan digital.	Daftar kebutuhan pembelajaran dan konteks budaya yang akan digunakan.
Design	Pemetaan materi pecahan, skenario dialog, contoh budaya terpilih, alur antarmuka, serta rancangan basis konten dan arsitektur LLM.	Spesifikasi konseptual prototipe. Catatan keputusan desain rinci tidak tersedia.
Development	Pembuatan prototipe web chatbot dan penilaian akhir pada dimensi materi, media, dan bahasa.	Prototipe dan skor validasi. Log revisi iteratif tidak terdokumentasi.
Implementation	Pretest, penggunaan chatbot dengan guru sebagai fasilitator, observasi, posttest, respons siswa, dan SUS pada 60 siswa.	Data kelompok tunggal. Jumlah pertemuan dan durasi sesi tidak terdokumentasi.
Evaluation	Ringkasan validasi, respons siswa, pre-post, SUS, telaah 120 respons, dan 100 catatan latensi.	Bukti kelayakan awal; tidak ada retensi, kelompok pembandingan, atau audit adaptivitas.

3.3 Sistem yang Dievaluasi

Produk berupa chatbot berbasis web yang menggunakan LLM untuk menghasilkan respons mengenai materi pecahan dengan merujuk pada konten yang disusun untuk penelitian. Dokumen sumber menyebut basis pengetahuan, tetapi tidak menyediakan identitas penyedia dan versi model, temperature, seed, system prompt, aturan keselamatan, metode chunking, embedding, retrieval, pembaruan konten, atau bukti perbandingan dengan dan tanpa retrieval. Karena itu, versi revisi tidak menyatakan bahwa Retrieval-Augmented Generation (RAG) telah diterapkan. RAG diposisikan sebagai arah pengembangan yang memerlukan dokumentasi arsitektur dan evaluasi khusus, sebagaimana dicontohkan oleh Thesen dan Park (2025).

3.4 Peserta dan Konteks Implementasi

Peserta terdiri atas 60 siswa kelas IV sekolah dasar di Kota Balikpapan yang dipilih secara purposif berdasarkan penerapan Kurikulum Merdeka dan ketersediaan fasilitas pembelajaran digital. Guru berperan sebagai fasilitator selama penggunaan. Catatan penelitian yang tersedia tidak memuat jumlah sekolah dan kelas, pengalaman awal siswa menggunakan chatbot, pembelajaran lain yang berlangsung bersamaan, jumlah pertemuan, durasi per sesi, total time-on-task, atau distribusi akses. Afiliasi budaya peserta dengan komunitas Dayak Paser juga tidak diukur; karena itu, kedekatan budaya siswa tidak diasumsikan.



3.5 Instrumen dan Sumber Data

Sumber	Konstruk/aspek	Cakupan terdokumentasi	Batas interpretasi
Validasi produk	Materi, media, dan bahasa	Tiga dimensi; skor akhir tersedia	Jumlah butir dan spesialisasi validator tidak konsisten pada dokumen awal; tidak diinferensikan.
Respons siswa	Menarik, mudah digunakan, menyenangkan, membantu memahami pecahan	60 siswa; empat persentase aspek	Tidak diperlakukan sebagai skala motivasi tervalidasi.
Tes pemahaman	Skor pecahan sebelum dan sesudah	60 pasangan skor; skala 0-100	Data mentah dan blueprint item tidak tersedia pada berkas revisi.
SUS	Usability	10 butir; 60 siswa	Dilaporkan sebagai skor agregat.
Telaah respons chatbot	Ketepatan deskriptif jawaban	120 pertanyaan; tiga penilai	Blueprint, data per penilai, kappa, dan adjudication tidak tersedia.
Log latensi	Waktu menghasilkan respons	100 interaksi	Kondisi jaringan/perangkat tidak dikontrol secara rinci.

3.6 Analisis Data

Skor validasi dan respons siswa disajikan sebagai persentase. Rata-rata pretest dan posttest digunakan untuk menghitung selisih kelompok dan N-Gain kelompok: $(84,7-55,3)/(100-55,3)=0,66$. Nilai ini menggantikan 0,73 karena angka 0,73 tidak dapat ditelusuri ke distribusi N-Gain individual dalam data yang tersedia.

Hasil uji t berpasangan yang tercatat pada naskah sumber adalah $t(59)=18,93$, $p<0,001$. Konsistensi internal diperiksa melalui identitas matematis statistik ringkasan. Berdasarkan selisih 29,4 dan $t=18,93$, standard error selisih adalah 1,553; interval kepercayaan 95% ($df=59$) direkonstruksi menjadi 26,29-32,51. Besaran perubahan berpasangan dihitung dengan Cohen's $dz=t/\sqrt{n}=2,44$. Angka tersebut adalah besaran perubahan dalam kelompok, bukan estimasi efek kausal chatbot. Uji Levene dihapus karena tidak relevan untuk pasangan pengamatan yang sama. Uji normalitas pretest dan posttest secara terpisah juga dihapus karena asumsi uji t berpasangan berkaitan dengan distribusi skor selisih, yang tidak dapat dihitung ulang tanpa data mentah.



4. Hasil

4.1 Hasil Pengembangan Produk

Tahapan ADDIE menghasilkan prototipe chatbot web untuk materi pecahan yang memuat contoh kontekstual Dayak Paser terpilih. Fitur yang dapat dinyatakan dari dokumentasi ialah antarmuka percakapan, penyajian contoh dan latihan pecahan, serta konten referensi yang disusun untuk mendukung jawaban. Dokumentasi tidak memadai untuk menyatakan adanya learner model, personalisasi berbasis profil, adaptasi berdasarkan performa sebelumnya, atau pipeline RAG.

4.2 Validasi Produk

Dimensi	Skor	Interpretasi terbatas
Materi	94,0%	Sangat valid menurut kriteria yang digunakan
Media	92,0%	Sangat valid menurut kriteria yang digunakan
Bahasa	94,0%	Sangat valid menurut kriteria yang digunakan
Rata-rata	93,4%	Ringkasan tiga dimensi

Skor tersebut menunjukkan penilaian positif terhadap produk pada tiga dimensi. Penilaian ini tidak sama dengan validasi autentisitas budaya oleh pemangku pengetahuan Dayak Paser dan tidak menguji perilaku AI secara komprehensif.

4.3 Respons Siswa dan Usability

Aspek	Hasil
Menarik	95%
Mudah digunakan	92%
Menyenangkan	94%
Membantu memahami pecahan	93%
System Usability Scale	87,6

Respons siswa berada pada rentang 92%-95% untuk empat aspek, sedangkan skor SUS adalah 87,6. Data ini mendukung penerimaan dan usability awal. Karena tidak terdapat skala motivasi tervalidasi atau kondisi pembanding, hasil tersebut tidak ditafsirkan sebagai peningkatan motivasi.



4.4 Perubahan Skor Pemahaman Pecahan

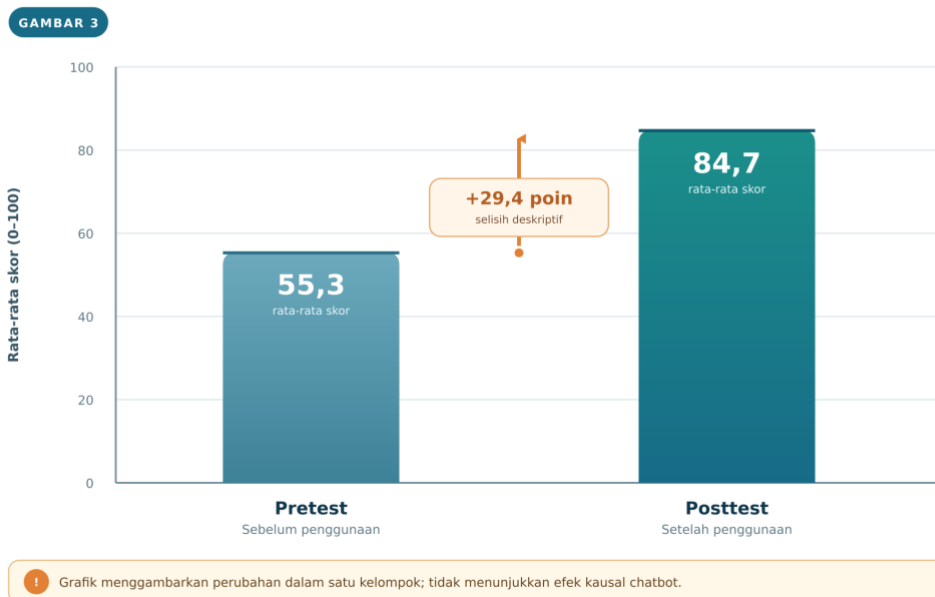
Statistik	Nilai	Status angka
Jumlah pasangan	60	Dilaporkan
Mean pretest	55,3	Dilaporkan
Mean posttest	84,7	Dilaporkan
Selisih mean	29,4	Dihitung dari mean final
SD pretest	8,72	Dilaporkan
SD posttest	6,43	Dilaporkan
N-Gain kelompok	0,66 (sedang)	Dihitung dari mean final
Paired t-test	$t(59)=18,93; p<0,001$	Dilaporkan
95% CI selisih	26,29-32,51	Direkonstruksi dari mean dan t
Cohen's dz	2,44	Dihitung sebagai $t/\sqrt{60}$

Rata-rata posttest lebih tinggi 29,4 poin daripada rata-rata pretest pada kelompok yang sama. Perbedaan tersebut signifikan pada ringkasan uji yang tersedia dan memiliki besaran perubahan berpasangan yang besar. Namun, hasil ini tetap merupakan perubahan pre-post dalam satu kelompok. Tanpa kelompok pembanding, hasil tidak menunjukkan seberapa besar perubahan yang secara khusus dapat diatribusikan kepada chatbot.



Perbandingan Rata-Rata Skor Pretest dan Posttest

Pemahaman pecahan sebelum dan setelah penggunaan chatbot pada kelompok yang sama (n = 60)



Gambar 3. Perbandingan Rata-Rata Skor Pretest dan Posttest

Rata-rata skor pemahaman pecahan pada kelompok yang sama (n = 60) berubah dari 55,3 pada pretest menjadi 84,7 pada posttest, dengan selisih deskriptif 29,4 poin. Grafik ini menunjukkan perubahan skor sebelum dan setelah penggunaan chatbot, bukan efek kausal chatbot.

4.5 Evaluasi Teknis Awal

Komponen	Hasil	Batas interpretasi
Telaah jawaban	113 dari 120 dinilai tepat; proporsi 94,2%	Proporsi deskriptif, bukan benchmark accuracy tervalidasi.
Latensi	Mean 2,18 detik; rentang 1,42-3,84; 100 interaksi	Deskriptif pada jaringan sekolah.
Precision/Recall/F1	Tidak dilaporkan pada versi revisi	TP, FP, dan FN tidak didefinisikan untuk tugas yang jelas.
Konsistensi generatif	Tidak dilaporkan pada versi revisi	Versi model, temperature, prompt, seed, dan retrieval tidak terdokumentasi.



Sebanyak 113 dari 120 jawaban dinilai tepat oleh tiga penilai. Karena blueprint pertanyaan, data penilaian per ahli, kesepakatan antarpemilai, dan prosedur adjudication tidak tersedia, angka 94,2% diperlakukan sebagai proporsi ketepatan deskriptif. Precision, recall, dan F1-score dihapus karena tidak ada definisi operasional true positive, false positive, dan false negative. Consistency rate juga dihapus karena kondisi generasi tidak dapat direplikasi.

5. Pembahasan

5.1 Kontribusi Pengembangan

Kontribusi yang paling dapat dipertanggungjawabkan adalah keberhasilan membangun prototipe percakapan yang menyatukan materi pecahan dan konteks Dayak Paser terpilih dalam satu antarmuka. Skor validasi 92%-94%, respons siswa 92%-95%, dan SUS 87,6 menunjukkan bahwa produk diterima secara positif pada evaluasi awal. Temuan ini mendukung kelayakan untuk pengembangan lanjutan, tetapi belum cukup untuk menyatakan bahwa produk bersifat adaptif. Dalam penelitian AI pendidikan, adaptivitas mensyaratkan bukti bahwa sistem mengubah isi, urutan, tingkat kesulitan, atau dukungan berdasarkan data pembelajar; mekanisme tersebut tidak diuji di sini.

5.2 Interpretasi Perubahan Pre-Post

Perubahan mean 29,4 poin, N-Gain kelompok 0,66, serta hasil t berpasangan yang signifikan menunjukkan bahwa skor setelah penggunaan lebih tinggi daripada skor sebelum penggunaan. Temuan tersebut relevan sebagai sinyal awal untuk penelitian lanjutan. Akan tetapi, desain satu kelompok tidak menyediakan counterfactual. Peningkatan dapat berkaitan dengan kombinasi paparan chatbot, penjelasan guru, latihan berulang, familiaritas terhadap tes, pembelajaran lain, atau novelty effect.

Perbedaan standar bukti tampak ketika hasil ini dibandingkan dengan Kestin et al. (2025), yang menggunakan kondisi pembandingan acak dan mencatat time-on-task sehingga dapat menguji kontribusi relatif tutor AI. Tinjauan Létourneau et al. (2025) juga menunjukkan bahwa hasil positif tutor AI pada K-12 perlu dibaca bersama durasi, desain pembandingan, keragaman peserta, dan kualitas kontrol. Dengan demikian, versi revisi menggunakan istilah



“perubahan pre-post” dan “evaluasi awal”, bukan “efek”, “pengaruh”, atau “terbukti meningkatkan”.

5.3 Respons Pengguna Tidak Sama dengan Motivasi

Persentase ketertarikan, kemudahan, kesenangan, dan persepsi membantu memahami pecahan menggambarkan respons positif setelah penggunaan. Namun, empat persentase tersebut bukan bukti peningkatan motivasi karena tidak terdapat definisi konstruk, skala motivasi tervalidasi, skor sebelum penggunaan, atau kelompok pembandingan. Kesimpulan dibatasi pada penerimaan dan usability awal.

5.4 Kualitas Respons AI dan Reprodusibilitas

Proporsi 113/120 jawaban yang dinilai tepat memberi indikasi awal bahwa sebagian besar respons pada set pertanyaan yang digunakan dapat diterima. Nilai tersebut belum dapat diperlakukan sebagai benchmark yang replikabel karena domain, kesulitan, komposisi item budaya dan nonbudaya, penggunaan item selama pengembangan, serta kesepakatan penilai tidak tersedia. Pelaporan yang lebih kuat memerlukan blueprint set uji, blind rating, Fleiss' kappa atau statistik kesepakatan lain, prosedur adjudication, dan pemisahan set pengembangan dari set evaluasi.

Klaim bahwa basis pengetahuan mengurangi hallucination juga dihapus. Studi RAG yang terdokumentasi, seperti Thesen dan Park (2025), menjelaskan pemuatan sumber, pemotongan teks, embedding, vector database, retrieval, prompt, serta pola penggunaan. Dokumentasi setara tidak tersedia pada penelitian ini. Penelitian berikutnya perlu membekukan versi model dan mencatat penyedia, tanggal akses, temperature, seed, system prompt, aturan keselamatan, sumber korpus, chunking, embedding, top-k retrieval, serta prosedur pembaruan.

5.5 Kedudukan Konteks Budaya

Konteks Dayak Paser berfungsi sebagai sumber contoh pembelajaran, sejalan dengan situated learning dan etnomatematika. Namun, relevansi geografis tidak otomatis membuktikan keterhubungan budaya peserta. Penelitian tidak mengukur afiliasi budaya siswa, pengetahuan budaya, identitas, atau dampak pada komunitas. Selain itu, dokumentasi tidak menunjukkan validasi isi oleh pemangku pengetahuan Dayak Paser. Karena itu, penelitian



hanya menyimpulkan bahwa konten budaya telah diintegrasikan dalam desain, bukan bahwa konten tersebut telah terverifikasi secara komunitas atau menghasilkan pelestarian budaya.

5.6 Keterbatasan

Keterbatasan utama penelitian meliputi desain satu kelompok tanpa kontrol dan randomisasi; jumlah sekolah/kelas yang tidak terdokumentasi; dosis intervensi dan time-on-task yang tidak tersedia; tidak adanya tes retensi; potensi pengaruh guru, pembelajaran lain, pengulangan tes, dan novelty effect; data mentah yang tidak tersedia untuk audit normalitas skor selisih dan distribusi N-Gain individual; ketidakjelasan jumlah butir serta spesialisasi validator pada catatan awal; tidak adanya validator budaya dari komunitas Dayak Paser; tidak diukurnya motivasi, adaptivitas, pengetahuan budaya, identitas, dan pelestarian budaya; tidak adanya detail model, prompt, parameter generasi, keselamatan, dan retrieval; benchmark respons tanpa blueprint dan kesepakatan antarpemangku; serta dokumentasi etika, consent, assent, privasi, dan pengelolaan data AI yang belum tersedia. Keterbatasan tersebut membatasi generalisasi dan menempatkan hasil sebagai evaluasi awal.

6. Kesimpulan

Penelitian menghasilkan prototipe chatbot matematika berbasis LLM yang mengintegrasikan materi pecahan dengan konteks Dayak Paser terpilih. Produk memperoleh skor validasi rata-rata 93,4%, respons siswa 92%-95%, dan skor SUS 87,6. Pada satu kelompok berjumlah 60 siswa, rata-rata skor berubah dari 55,3 menjadi 84,7; N-Gain kelompok yang konsisten dengan mean final adalah 0,66. Ringkasan uji berpasangan menunjukkan perbedaan statistik, tetapi desain tidak mendukung atribusi kausal kepada chatbot. Evaluasi teknis awal menunjukkan 113 dari 120 jawaban dinilai tepat dan latensi rata-rata 2,18 detik, dengan keterbatasan benchmark serta reproduksibilitas. Dengan demikian, bukti yang tersedia mendukung kelayakan awal produk dan studi lanjutan yang terkontrol, bukan klaim efektivitas kausal, adaptivitas, peningkatan motivasi, penerapan RAG, atau pelestarian budaya.

Penelitian berikutnya perlu menggunakan kelompok pembandingan atau desain crossover, mencatat dosis dan time-on-task, melakukan tes retensi, memvalidasi konten



bersama pemangku pengetahuan Dayak Paser, melengkapi persetujuan etik dan perlindungan data anak, membekukan konfigurasi LLM, serta mempublikasikan protokol evaluasi respons dan kesepakatan antarpenilai.

DEKLARASI PENGGUNAAN AI GENERATIF

Selama penyusunan naskah, penulis menggunakan alat bantu AI, termasuk ChatGPT, Grammarly, dan QuillBot, untuk mendukung penyuntingan bahasa. Penulis bertanggung jawab untuk memverifikasi seluruh isi, statistik, sitasi, kepatuhan etik, dan kesesuaian dengan kebijakan jurnal tujuan.

7. Daftar Pustaka

- Branch, R. M. (2009). *Instructional design: The ADDIE approach*. Springer.
<https://doi.org/10.1007/978-0-387-09506-6>
- Chiu, T. K. F., Xia, Q., Zhou, X., Chai, C. S., & Cheng, M. (2023). Systematic literature review on opportunities, challenges, and future research recommendations of artificial intelligence in education. *Computers and Education: Artificial Intelligence*, 4, 100118.
<https://doi.org/10.1016/j.caeai.2022.100118>
- Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeiffer, F., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., ... Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences*, 103, 102274.
<https://doi.org/10.1016/j.lindif.2023.102274>
- Kestin, G., Miller, K., Klales, A., Milbourne, T., & Ponti, G. (2025). AI tutoring outperforms in-class active learning: An RCT introducing a novel research-based design in an authentic educational setting. *Scientific Reports*, 15, 17458. <https://doi.org/10.1038/s41598-025-97652-6>
- Létourneau, A., Deslandes Martineau, M., Charland, P., Karran, J. A., Boasen, J., & Léger, P. M. (2025). A systematic review of AI-driven intelligent tutoring systems (ITS) in K-12 education. *npj Science of Learning*, 10, 29. <https://doi.org/10.1038/s41539-025-00320-7>
- Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. Cambridge University Press.
- Lo, C. K. (2023). What is the impact of ChatGPT on education? A rapid review of the literature. *Interactive Learning Environments*. Advance online publication.
<https://doi.org/10.1080/10494820.2023.2253852>
- Okonkwo, C. W., & Ade-Ibijola, A. (2021). Chatbots applications in education: A systematic review. *Computers and Education: Artificial Intelligence*, 2, 100033.
<https://doi.org/10.1016/j.caeai.2021.100033>



-
- Piaget, J. (1972). *The psychology of the child*. Basic Books.
- Rosa, M., & Orey, D. C. (2020). *Ethnomathematics and its diverse approaches for mathematics education*. Springer. <https://doi.org/10.1007/978-3-030-37697-2>
- Rosa, M., & Orey, D. C. (2022). Ethnomathematics and culturally relevant mathematics education: An ethnomodelling perspective. *Journal of Mathematics and Culture*, 16(1), 1-24.
- Thesen, T., & Park, S. H. (2025). A generative AI teaching assistant for personalized learning in medical education. *npj Digital Medicine*, 8, 627. <https://doi.org/10.1038/s41746-025-02022-1>
- Tlili, A., Shehata, B., Adarkwah, M. A., Bozkurt, A., Hickey, D. T., Huang, R., & Agyemang, B. (2023). What if the devil is my guardian angel: ChatGPT as a case study of using chatbots in education. *Smart Learning Environments*, 10, 15. <https://doi.org/10.1186/s40561-023-00237-x>
- UNESCO. (2023). *Guidance for generative AI in education and research*. UNESCO. <https://unesdoc.unesco.org/ark:/48223/pf0000386693>
- Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.