

PENGARUH TEKNOLOGI ARTIFICIAL INTELLIGENCE DALAM PEMBELAJARAN MATEMATIKA TERHADAP KEMAMPUAN BERPIKIR KREATIF MATEMATIS SISWA SEKOLAH MENENGAH PERTAMA: SYSTEMATIC REVIEW DENGAN NARRATIVE SYNTHESIS

Yoel Octobe Purba¹, Rianita Simamora²

¹ Prodi Pendidikan Matematika, Universitas HKBP Nommensen Pematangsiantar;

yoeloctobe@gmail.com

² Prodi Pendidikan Matematika, Universitas HKBP Nommensen Pematangsiantar;

rianitacharlito@gmail.com

ARTICLE INFO

Article history:

Received 2026-04-17

Revised 2026-04-29

Accepted 2026-04-30

ABSTRAK

Pemanfaatan artificial intelligence (AI) dalam pembelajaran matematika meningkat sangat pesat. Namun, bukti empiris mengenai sejauh mana AI berkontribusi pada kemampuan berpikir kreatif matematis siswa Sekolah Menengah Pertama masih terpecah dalam berbagai desain penelitian, ragam teknologi, serta instrumen pengukuran yang tidak seragam. Review ini disusun untuk merangkum arah pengaruh, memaparkan besar efek pada masing-masing studi, sekaligus menjelaskan mekanisme pedagogis berbasis AI yang berkaitan dengan dimensi fluency, flexibility, originality, dan elaboration. Pelaporan kajian mengikuti pedoman PRISMA 2020 dan PRISMA-S. Penelusuran artikel empiris yang telah melalui proses peer review dilakukan pada basis data Scopus, Web of Science, ERIC, dan IEEE Xplore dengan menggunakan kriteria PICOS. Populasi dibatasi pada siswa SMP/lower secondary; intervensi berupa pembelajaran matematika yang memuat komponen AI secara eksplisit; pembandingan berupa pembelajaran konvensional atau pembelajaran digital non-AI; sedangkan luaran mencakup kreativitas matematis atau indikator yang sepadan. Proses seleksi, ekstraksi data, dan penilaian kualitas dilakukan oleh dua reviewer secara independen. Risiko bias dinilai menggunakan RoB 2, ROBINS-I, dan MMAT sesuai karakteristik desain studi. Sebanyak delapan studi dengan total 596 siswa memenuhi kriteria untuk disintesis. Seluruh studi memperlihatkan arah efek positif; tujuh studi melaporkan hasil yang signifikan secara statistik, sementara satu studi menunjukkan kecenderungan positif tetapi belum signifikan. Ukuran efek per studi berada pada rentang Hedges' $g = 0,65-1,87$, dengan median 1,13 dan IQR 1,00–1,33. Mengingat adanya heterogenitas yang nyata pada jenis AI, durasi intervensi, rancangan pedagogis, serta instrumen luaran, temuan tidak digabungkan ke dalam pooled effect. Melalui sintesis naratif, teridentifikasi tiga mekanisme yang paling menonjol, yaitu adaptive scaffolding, instant feedback, dan multi-representational affordances. Secara keseluruhan, bukti yang tersedia tampak menjanjikan, tetapi belum cukup kuat untuk dinyatakan final; riset berikutnya perlu mengandalkan instrumen kreativitas matematis yang tervalidasi, rancangan RCT atau kuasi-eksperimen yang lebih kokoh, serta pelaporan studi primer yang lebih transparan.

Kata Kunci: Artificial Intelligence; Pembelajaran Matematika; Berpikir Kreatif Matematis; Sekolah Menengah Pertama; Systematic Review

ABSTRACT

The use of artificial intelligence (AI) in mathematics learning is increasing rapidly. However, empirical evidence regarding the extent to which AI contributes to junior high school students' mathematical creative thinking skills remains scattered across various research designs, technologies, and measurement instruments. This review aims to summarize the direction of influence, describe the effect sizes in each study, and explain AI-based pedagogical mechanisms related to the dimensions of fluency, flexibility, originality, and elaboration. The study reporting follows PRISMA 2020 and PRISMA-S guidelines. A search for peer-reviewed empirical articles was conducted in Scopus, Web of Science, ERIC, and IEEE Xplore databases using PICOS criteria. The population was limited to junior high school/lower secondary students; the intervention was mathematics learning that explicitly included AI components; the comparison was conventional learning or non-AI digital learning; and the outcomes included mathematical creativity or equivalent indicators. The selection process, data extraction, and quality assessment were carried out independently by two reviewers. Risk of bias was assessed using the RoB 2, ROBINS-I, and MMAT according to study design characteristics. Eight studies with a total of 596 students met the criteria for synthesis. All studies demonstrated a positive effect direction; seven reported statistically significant results, while one study showed a positive trend but was not yet significant. Effect sizes per study ranged from Hedges' $g = 0.65$ – 1.87 , with a median of 1.13 and an IQR of 1.00 – 1.33 . Given the significant heterogeneity in AI type, intervention duration, pedagogical design, and outcome instruments, findings were not combined into a pooled effect. Through narrative synthesis, three prominent mechanisms were identified: adaptive scaffolding, instant feedback, and multi-representational affordances. Overall, the available evidence appears promising but is not yet strong enough to be definitive; future research should rely on validated mathematical creativity instruments, more robust RCT or quasi-experimental designs, and more transparent reporting of primary studies.

Keyword: Artificial Intelligence; Mathematics Learning; Mathematical Creative Thinking; Junior High School; Systematic Review

This is an open access article under the [CC BY](#) license.



Corresponding Author:

Yoel Octobe Purba

Prodi Pendidikan Matematika, Universitas HKBP Nommensen Pematangsiantar; yoeloctobe@gmail.com

1. PENDAHULUAN

Dalam satu dekade terakhir, artificial intelligence in education (AIED) bergerak cepat dari sekadar alat komputasi menuju lingkungan belajar yang semakin adaptif. Di kelas matematika, pergeseran ini tampak pada penggunaan intelligent tutoring systems (ITS), platform belajar adaptif, chatbot berbasis generative AI, learning analytics, sampai augmented reality yang diperkaya AI. Ragam teknologi tersebut bukan hanya “memilihkan” materi sesuai tingkat kemampuan siswa, tetapi juga dapat mendeteksi pola kesalahan, memberi scaffolding berjenjang, dan memfasilitasi dialog matematis yang dahulu umumnya terjadi melalui interaksi langsung guru-siswa (Zawacki-Richter, Marín, Bond, & Gouverneur, 2019). Meski demikian, kemajuan teknis ini belum otomatis menjawab persoalan pedagogis yang lebih mendasar: sejauh mana AI mampu mendorong kemampuan berpikir matematis tingkat tinggi, terutama kreativitas, dan bukan hanya mempercepat penguasaan prosedur atau menaikkan skor tes.

Isu tersebut menjadi makin penting jika diletakkan dalam kerangka visi pendidikan matematika kontemporer. National Research Council (Zhang & Aslan, 2021) menegaskan bahwa kemahiran matematika yang utuh terdiri atas lima untaian yang saling menguatkan, yakni pemahaman konseptual, kelancaran prosedural, kompetensi strategis, penalaran adaptif, dan disposisi produktif. Sejalan dengan itu, NCTM (2014) menekankan bahwa pembelajaran matematika yang efektif menuntut siswa mampu memecahkan masalah, membangun representasi, terlibat dalam diskursus matematis, serta menggunakan bukti dan alasan secara bermakna. Pada level global, PISA 2022 pada domain creative thinking menunjukkan bahwa kemampuan menghasilkan, menilai, dan menyempurnakan ide orisinal kini diposisikan sebagai indikator penting kesiapan peserta didik menghadapi tuntutan abad ke-21 (OECD, 2024). Dengan demikian, ukuran prestasi konvensional saja tidak memadai; pendidikan matematika perlu secara sengaja mengembangkan kapasitas siswa untuk berpikir melampaui prosedur yang rutin dan jalur penyelesaian yang lazim.

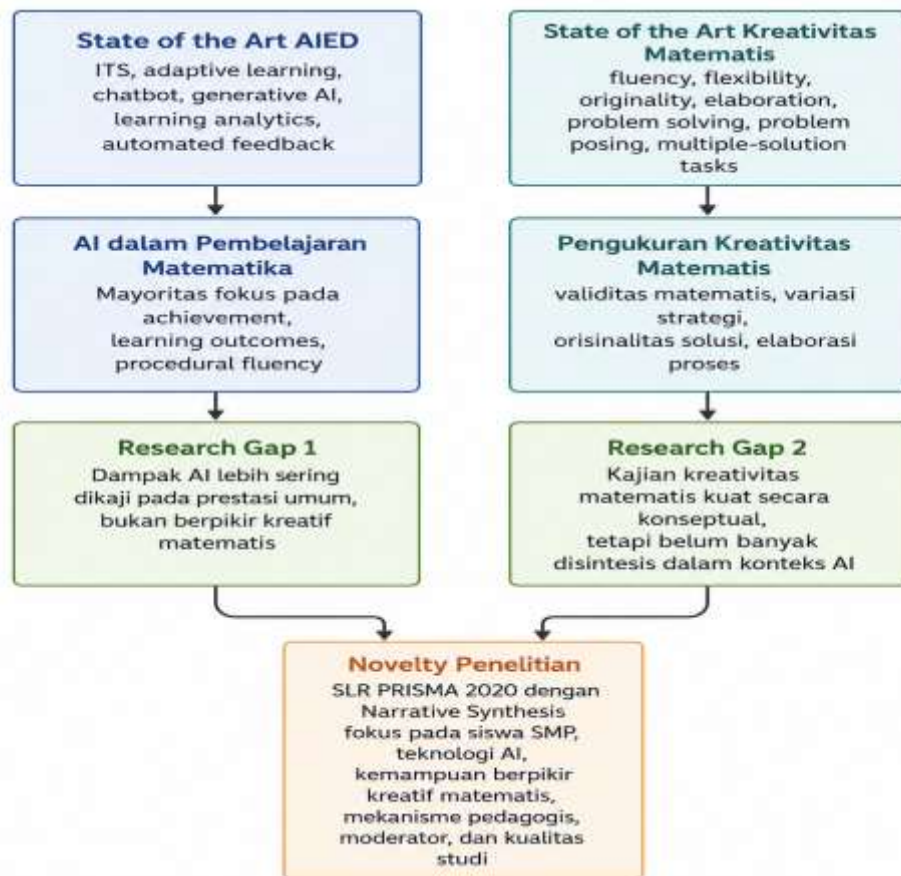
Bagi siswa Sekolah Menengah Pertama, urgensi ini bahkan lebih kuat. Rentang usia 11–15 tahun dalam tradisi psikologi perkembangan Piaget (Sterne et al., 2019) sering dipahami sebagai periode transisi dari operasional konkret menuju penalaran formal, ketika kemampuan berpikir hipotetis dan reflektif mulai berkembang lebih stabil. Pada tahap ini, pengalaman belajar yang menantang, bersifat eksploratif, dan memberi ruang untuk mencoba berbagai strategi berpotensi membentuk cara siswa menghadapi masalah matematika yang kompleks dalam jangka panjang. Karena itu, apabila AI ingin memberikan kontribusi yang bermakna pada pembelajaran matematika di SMP, kontribusi tersebut semestinya tidak berhenti pada percepatan penguasaan materi, melainkan menyentuh dimensi kreativitas yang menjadi bagian penting dari kematangan berpikir matematis.

Kemampuan berpikir kreatif matematis merupakan konstruk yang tidak dapat disamakan begitu saja dengan kreativitas umum maupun sekadar prestasi matematika. Silver ((Steenbergen-Hu & Cooper, 2013) menautkan kreativitas matematis dengan problem solving dan problem posing, dengan indikator utama berupa fluency (kelancaran menghasilkan ide/solusi), flexibility (keluwesan menggunakan beragam strategi), dan originality (kebaruan pendekatan). Haylock (Sterne et al., 2016) menekankan aspek penting lainnya, yaitu kemampuan mengatasi fixation kecenderungan bertahan pada satu sudut pandang serta menghasilkan respons divergen pada tugas matematika yang sama. Leikin dan Pitta-Pantazi (2013) memperluas perspektif ini dengan memandang kreativitas matematis sebagai ranah yang memadukan dimensi produk, proses, individu, dan lingkungan belajar secara saling terkait (Siddaway, Wood, & Hedges, 2019). Kekhasan kreativitas matematis dibanding bidang lain terletak pada tuntutan bahwa ide yang baru tetap harus sah secara matematis: kebaruan tidak menghapus kewajiban koherensi logis dan konseptual. Pada konteks SMP, kreativitas matematis dapat tampak ketika siswa mampu

menemukan lebih dari satu cara penyelesaian, membangun representasi alternatif, merumuskan masalah baru dari situasi yang tersedia, atau menawarkan solusi non-rutin yang tetap dapat dipertanggungjawabkan (Sitorus & Masrayati, 2016).

Walaupun secara teoretis AI tampak “selaras” untuk mendukung kreativitas matematis misalnya melalui adaptive scaffolding yang mendorong eksplorasi multi-strategi, umpan balik segera yang mempercepat siklus revisi ide, atau chatbot dialogis yang menstimulasi eksternalisasi penalaran—potensi tersebut tidak dapat langsung diperlakukan sebagai bukti efektivitas. Sejumlah sintesis terdahulu di bidang AIED umumnya memetakan dampak AI terhadap capaian akademik secara luas, pada jenjang dan mata pelajaran yang beragam, tanpa memberi fokus yang cukup pada kreativitas matematis siswa SMP sebagai luaran yang spesifik (Chen et al., 2020; Crompton & Burke, 2023; Holmes & Tuomi, 2022; Zawacki-Richter et al., 2019). Meta-analisis terkait ITS pada matematika K–12 telah memberikan bukti penting mengenai peningkatan hasil belajar (Sitorus & Masrayati, 2016), tetapi luaran yang dikaji umumnya masih bertumpu pada skor tes dan belum memisahkan kreativitas matematis sebagai konstruk tersendiri. Di sisi lain, literatur kreativitas matematis yang berkembang melalui problem solving, problem posing, maupun multiple-solution tasks juga belum secara sistematis mengaitkan temuan-temuannya dengan affordance pedagogis dari teknologi AI (Steenbergen-Hu & Cooper, 2013). Ruang kosong inilah yang menjadi pijakan sekaligus alasan utama penelitian ini dilakukan.

Berdasarkan pertimbangan tersebut, penelitian ini disusun sebagai systematic review dengan narrative synthesis mengacu pada PRISMA 2020. Tujuan penelitian adalah: (1) merangkum bukti empiris mengenai arah dan besaran pengaruh teknologi AI dalam pembelajaran matematika terhadap kemampuan berpikir kreatif matematis siswa SMP; (2) mengidentifikasi mekanisme pedagogis yang memungkinkan berbagai tipe AI memfasilitasi dimensi fluency, flexibility, originality, dan elaboration; serta (3) menelaah faktor-faktor yang berpotensi memoderasi besar-kecilnya pengaruh, mencakup jenis AI, durasi intervensi, topik matematika, desain penelitian, dan kualitas metodologis. Pemilihan narrative synthesis dipandang tepat karena studi-studi yang tersedia menunjukkan heterogenitas yang substansial baik pada teknologi, konteks intervensi, maupun instrumen pengukuran sehingga penggabungan statistik justru berisiko menghasilkan simpulan yang keliru (Siswono, 2011). Kebaruan review ini terletak pada sintesis terpadu atas tiga elemen yang selama ini jarang dipertautkan secara sistematis: (a) fokus jenjang SMP sebagai konteks spesifik, (b) kreativitas matematis sebagai luaran yang didefinisikan eksplisit melalui dimensi fluency, flexibility, originality, dan elaboration, serta (c) penjelasan berbasis mekanisme pedagogis AI sebagai jalur argumentasi, bukan sekadar inventarisasi teknologi. Dengan demikian, review ini tidak hanya menjawab apakah AI berpengaruh, melainkan juga menjelaskan bagaimana, dalam kondisi apa, dan melalui affordance pedagogis apa AI dapat berkontribusi terhadap perkembangan kreativitas matematis siswa SMP (Silver, 1997).



Gambar 1. State Of The Art dan Novelty

2. METODE

Penelitian ini menggunakan desain systematic literature review (SLR) dengan narrative synthesis. Pilihan desain ini didasarkan pada pertimbangan bahwa studi intervensi AI dalam konteks kreativitas matematis masih dalam tahap awal akumulasi bukti, sehingga sintesis yang baik harus mampu memetakan pola dan mekanisme, bukan sekadar merata-ratakan besaran efek. Pelaporan tinjauan mengikuti PRISMA 2020 (Rethlefsen et al., 2021), sementara pelaporan strategi penelusuran literatur mengikuti ekstensi PRISMA-S (Popay et al., 2006). Kerangka PRISMA 2020 dirancang untuk mendukung transparansi proses identifikasi, seleksi, dan sintesis bukti, termasuk di dalamnya dokumentasi keputusan eksklusi yang terinci pada setiap tahap.

Penelusuran literatur dilakukan secara sistematis pada 15 Desember 2025 melalui empat basis data ilmiah utama: Scopus, Web of Science Core Collection, ERIC, dan IEEE Xplore. Keempat basis data ini dipilih karena cakupannya yang saling melengkapi Scopus dan Web of Science menyediakan jangkauan multidisiplin yang luas dengan indeksasi jurnal pendidikan bereputasi tinggi; ERIC berspesialisasi pada riset pendidikan; sementara IEEE Xplore relevan untuk literatur teknologi pembelajaran dan sistem cerdas. Batasan bahasa mencakup artikel berbahasa Inggris dan Indonesia yang diterbitkan antara 1 Januari 2018 hingga 15 Desember 2025. Strategi pencarian menggabungkan empat kluster istilah yang dioperasionalkan melalui operator Boolean: (1) teknologi AI 'artificial intelligence', 'intelligent tutoring system', 'adaptive learning', 'chatbot', 'generative AI', 'large language model'; (2) pembelajaran matematika 'mathematics education', 'mathematics learning'; (3) berpikir kreatif matematis 'mathematical

creativity', 'creative thinking', 'problem posing', 'multiple solutions', 'fluency flexibility originality'; (4) konteks SMP 'middle school', 'junior high school', 'lower secondary', 'grade 7', 'grade 8', 'grade 9' (Ouyang & Jiao, 2021).

Kriteria kelayakan dirumuskan melalui kerangka PICOS. Populasi: siswa SMP atau lower secondary usia 11–15 tahun dalam pendidikan formal. Intervensi: pembelajaran matematika yang mengintegrasikan komponen AI secara eksplisit, termasuk ITS, chatbot, *generative AI*, *adaptive learning platform*, *learning analytics*, atau *AI-assisted AR/VR*, dengan durasi intervensi minimal empat minggu. *Comparison*: pembelajaran konvensional, pembelajaran digital non-AI, atau kondisi kontrol *business-as-usual*. *Outcome*: kemampuan berpikir kreatif matematis yang mencakup setidaknya satu dari empat dimensi *fluency*, *flexibility*, *originality*, dan/atau *elaboration* diukur melalui instrumen yang memiliki dasar konseptual kreativitas matematis. *Study design*: RCT, *quasi-experimental*, atau *mixed methods* yang menyediakan data kuantitatif dan diterbitkan melalui mekanisme *peer review*. Studi dikeluarkan apabila berfokus pada jenjang SD, SMA, atau pendidikan tinggi tanpa data SMP terpisah; menggunakan teknologi digital tanpa komponen AI yang dapat diidentifikasi secara eksplisit; tidak memiliki kelompok pembandingan; hanya melaporkan prestasi matematika umum atau luaran afektif semata; atau berbentuk artikel konseptual, editorial, tinjauan naratif non-sistematis, *grey literature*, dan publikasi *non-peer-reviewed* (OECD, 2024).

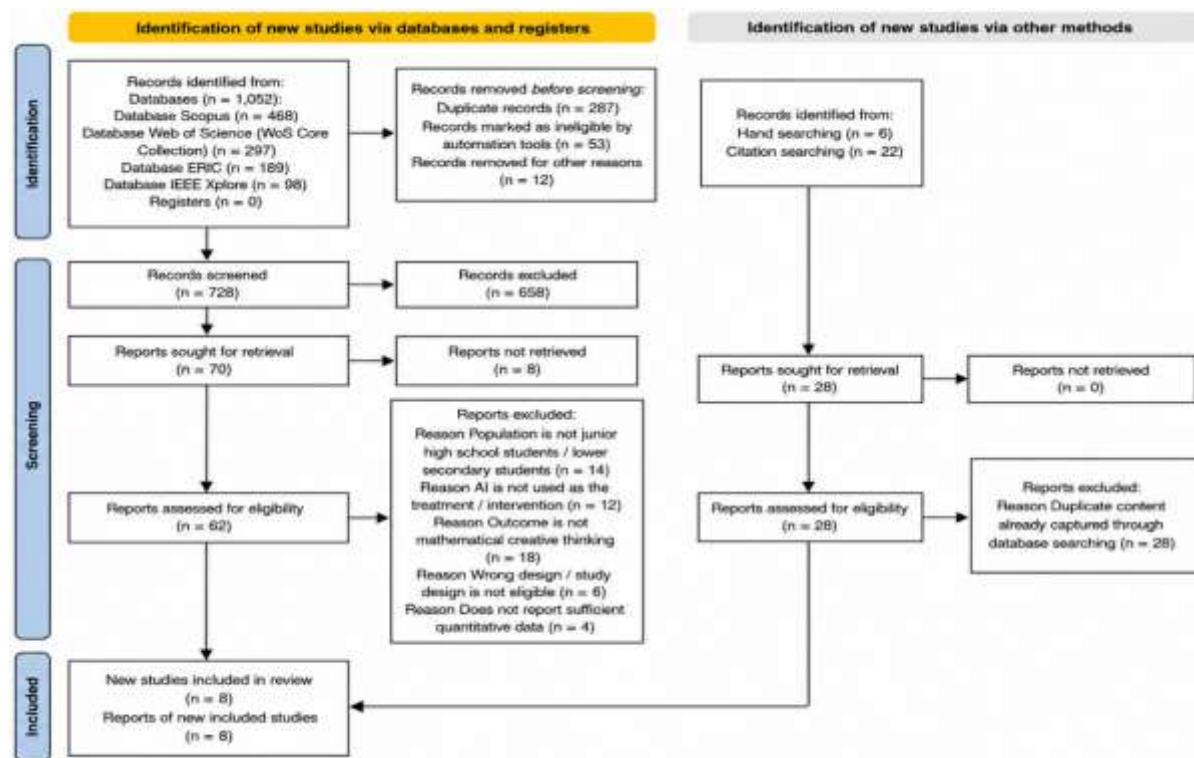
Proses seleksi dilakukan oleh dua reviewer independen yang bekerja secara buta satu sama lain. Seluruh rekaman diekspor ke reference manager untuk deduplikasi otomatis dan manual, kemudian skrining judul-abstrak dikelola menggunakan aplikasi Rayyan yang dirancang khusus untuk mempercepat dan mengorganisasi proses skrining SLR. Reliabilitas keputusan antar-reviewer dihitung menggunakan Cohen's kappa (Cohen, 1960). Perbedaan keputusan diselesaikan melalui diskusi konsensus. Pada tahap penilaian kelayakan penuh, setiap artikel diperiksa menggunakan *checklist* PICOS yang terstandar; artikel *borderline* misalnya yang menggunakan istilah '*mathematical thinking*' tetapi instrumennya memuat item *fluency*, *flexibility*, atau *originality* diputuskan berdasarkan telaah instrumen secara eksplisit, bukan interpretasi label.

Ekstraksi data dilakukan dengan form terstandar yang mencakup: identitas studi, negara dan bahasa publikasi, desain penelitian, karakteristik partisipan, jenis AI dan nama platform, desain pedagogis, durasi dan frekuensi intervensi, peran guru, jenis kelompok pembandingan, topik matematika, instrumen kreativitas beserta dimensi yang diukur, waktu pengukuran, serta data kuantitatif *mean*, *standard deviation*, ukuran sampel, atau statistik uji seperti nilai *t*, *F*, dan *p* yang dapat dikonversi. Apabila data memadai, ukuran efek individual dihitung sebagai Hedges' *g* dengan koreksi small-sample bias (Council, 2001). Ukuran efek per studi ini digunakan sebagai ringkasan kuantitatif deskriptif, bukan sebagai dasar pooled effect.

Penilaian risiko bias menggunakan tiga instrumen yang disesuaikan dengan desain studi: RoB 2 untuk randomized controlled trials, ROBINS-I untuk studi non-randomized atau quasi-experimental, dan MMAT untuk studi mixed methods. Pemilihan tiga instrumen ini mencerminkan komitmen untuk menilai kualitas metodologis sesuai dengan standar desain masing-masing studi, bukan menerapkan satu rubrik tunggal yang tidak sensitif terhadap perbedaan epistemologis (Mathematics, 2014).

Sintesis dilakukan secara naratif karena heterogenitas substantif pada jenis AI, desain pedagogis, durasi intervensi, instrumen kreativitas, dan desain penelitian membuat agregasi statistik tidak tepat dan berpotensi menyesatkan sejalan dengan prinsip Cochrane bahwa sintesis selain meta-analisis dapat digunakan ketika pooling tidak memadai (Higgins et al., 2024). Pelaporan narrative synthesis mengikuti panduan SWiM (Campbell et al., 2020) yang dikembangkan spesifik untuk dokumentasi sintesis kuantitatif tanpa meta-analisis. Sintesis mencakup tiga tahap berurutan: (a) penyusunan ringkasan

karakteristik dan temuan per studi; (b) analisis hubungan antarstudi berdasarkan jenis AI dan mekanisme pedagogis yang dilaporkan; dan (c) penilaian ketahanan temuan melalui pembacaan sensitif terhadap risiko bias. Eksplorasi faktor moderator jenis AI, durasi intervensi, topik matematika, desain studi, dan kualitas metodologis dilakukan secara deskriptif tanpa uji inferensial karena jumlah studi per subkelompok terlalu kecil untuk meta-regression yang sah. Kepastian bukti ditafsirkan secara adaptif menggunakan prinsip GRADE, terutama dengan mempertimbangkan dimensi risiko bias, inkonsistensi antarstudi, ketidaklangsungan bukti terhadap populasi dan luaran yang dituju, imprecision karena keterbatasan sampel, dan besaran efek sebagai bukti tambahan (Ma, Adesope, Nesbit, & Liu, 2014).



Gambar 2. Prisma 2020 Berdasarkan Olahan Peneliti

3. HASIL DAN PEMBAHASAN

1. Hasil

1.1. Karakteristik Studi dan Proses Seleksi

Proses identifikasi, penyaringan, dan pelaporan artikel dalam kajian ini disusun dengan mengacu pada pedoman PRISMA 2020 dan PRISMA-S. Acuan tersebut digunakan agar seluruh tahapan seleksi, baik pada tahap inklusi maupun eksklusi, dapat dijelaskan secara transparan, sistematis, dan memungkinkan untuk ditinjau kembali oleh peneliti lain. Pada tahap awal, penelusuran melalui empat basis data utama menghasilkan 1.052 rekaman. Rincian sumber tersebut meliputi Scopus sebanyak 468 rekaman, Web of Science sebanyak 297 rekaman, ERIC sebanyak 189 rekaman, dan IEEE Xplore sebanyak 98 rekaman. Di luar pencarian basis data, ditemukan pula 28 rekaman tambahan melalui metode lain, yang terdiri atas 6 rekaman dari hand searching dan 22 rekaman dari citation searching. Dengan demikian, jumlah keseluruhan rekaman yang teridentifikasi pada tahap awal adalah 1.080 rekaman (Leikin & Pitta-Pantazi, 2013).

Sebelum memasuki tahap skrining, dilakukan proses pembersihan data terhadap rekaman yang tidak dapat dilanjutkan ke tahap berikutnya. Pada tahap ini, sebanyak 352 rekaman dikeluarkan. Jumlah tersebut mencakup 287 rekaman duplikat, 53 rekaman yang teridentifikasi tidak memenuhi syarat melalui automation tools, serta 12 rekaman yang dihapus karena alasan teknis atau pertimbangan lain. Setelah proses tersebut selesai, terdapat 728 rekaman unik yang kemudian diseleksi melalui pembacaan judul dan abstrak. Dari proses skrining ini, sebanyak 658 rekaman dieliminasi karena tidak sesuai dengan kriteria PICOS yang telah ditetapkan. Beberapa alasan utama eksklusi pada tahap ini antara lain artikel tidak membahas penggunaan AI, populasi penelitian bukan siswa SMP atau lower secondary, tidak memuat luaran berupa kreativitas matematis, atau tidak termasuk artikel empiris yang relevan dengan fokus kajian.

Tahap berikutnya adalah penelusuran laporan lengkap. Dari hasil pencarian database, terdapat 70 laporan yang dicari teks lengkapnya untuk dinilai lebih lanjut. Namun, 8 laporan tidak berhasil diperoleh sehingga hanya 62 laporan yang dapat dibaca secara penuh dan dievaluasi kelayakannya. Berdasarkan penilaian terhadap teks lengkap tersebut, sebanyak 54 laporan dari database dikeluarkan karena tidak memenuhi kriteria inklusi. Alasan eksklusi meliputi ketidaksesuaian populasi, yaitu bukan siswa SMP atau lower secondary, sebanyak 14 laporan; AI tidak digunakan sebagai treatment atau intervensi sebanyak 12 laporan; luaran penelitian tidak berfokus pada mathematical creative thinking sebanyak 18 laporan; desain studi tidak sesuai dengan kriteria sebanyak 6 laporan; serta data kuantitatif yang tersedia tidak memadai untuk dianalisis sebanyak 4 laporan (Leikin & Lev, 2007).

Selain laporan yang diperoleh dari database, sebanyak 28 laporan yang berasal dari metode lain juga diperiksa pada tahap kelayakan. Akan tetapi, seluruh laporan tersebut akhirnya dikeluarkan karena merupakan konten duplikat yang telah tercakup dalam hasil pencarian melalui database. Setelah seluruh proses identifikasi, skrining, dan penilaian kelayakan dilakukan, diperoleh 8 studi empiris yang memenuhi kriteria inklusi dan layak dimasukkan ke dalam sintesis akhir (Piaget, 1970). Untuk memastikan konsistensi keputusan dalam proses seleksi, dilakukan pula pengukuran reliabilitas antar-reviewer. Hasilnya menunjukkan nilai Cohen's kappa sebesar 0,82, yang mengindikasikan tingkat kesepakatan yang kuat atau substantial agreement. Konsistensi keputusan antar-reviewer pada tahap skrining menunjukkan kesepakatan yang kuat, dengan Cohen's $\kappa = 0,82$ (Cohen, 1960), yang umumnya diinterpretasikan sebagai substantial agreement (Hong et al., 2018).

Delapan studi yang terinklusi melibatkan total 596 siswa, dengan ukuran sampel per studi berkisar 48–120 siswa. Secara geografis, empat studi berasal dari Indonesia, dua dari Korea Selatan, satu dari Kazakhstan, dan satu dari Italia, sehingga mencerminkan variasi konteks budaya maupun kurikulum. Dominasi desain kuasi-eksperimental selaras dengan praktik penelitian intervensi di sekolah, di mana randomisasi penuh kerap sulit dilakukan karena pertimbangan etis dan kendala operasional; hanya satu studi menerapkan RCT, dan satu studi menggunakan pendekatan mixed methods. Teknologi AI yang ditelaah mencakup generative AI/chatbot (lima studi), intelligent tutoring system (satu), platform pembelajaran personalisasi berbasis AI (satu), serta GeoAI (satu). Intervensi diterapkan pada beragam topik matematika, termasuk aljabar, geometri, fungsi, statistik, aritmetika, maupun kombinasi beberapa topik kurikuler. Variasi ini menunjukkan bahwa integrasi AI di pembelajaran matematika SMP belum mengerucut pada satu topik atau pendekatan tertentu, dan keragaman tersebut menjadi salah satu sumber heterogenitas substantif dalam sintesis.

Tabel 1. Rekonstruksi Seleksi Studi Berdasarkan PRISMA 2020

Tahap Seleksi	Jumlah	Keterangan
Records identified from databases	1.052	Scopus 468; Web of Science 297; ERIC 189; IEEE Xplore 98; registers 0
Records identified from other methods	28	Hand searching 6; citation searching 22
Records removed before screening	352	Duplicate records 287; records marked as ineligible by automation tools 53; records removed for other reasons 12
Unique records screened	728	Rekaman unik yang diseleksi melalui title-abstract screening
Records excluded	658	Tidak memenuhi kriteria PICOS, seperti tidak membahas AI, populasi bukan siswa SMP/lower secondary, tidak memuat luaran kreativitas matematis, atau bukan artikel empiris
Reports sought for retrieval from databases	70	Artikel yang berpotensi relevan dan dicari untuk dibaca lebih lanjut
Reports not retrieved from databases	8	Artikel tidak dapat diperoleh atau tidak tersedia teks lengkap
Reports assessed for eligibility from databases	62	Artikel dibaca penuh untuk menilai kesesuaian dengan kriteria inklusi
Full-text reports excluded from databases	54	Populasi bukan siswa SMP/lower secondary 14; AI bukan treatment/intervention 12; outcome bukan mathematical creative thinking 18; desain studi tidak memenuhi kriteria 6; data kuantitatif tidak cukup 4
Reports sought for retrieval from other methods	28	Artikel dari hand searching dan citation searching
Reports not retrieved from other methods	0	Semua artikel dari metode lain dapat diperiksa
Reports assessed for eligibility from other methods	28	Artikel dari metode lain dinilai kelayakannya
Reports excluded from other methods	28	Dikeluarkan karena merupakan konten duplikat yang sudah tercakup dalam database searching
Final included studies	8	Delapan studi empiris memenuhi kriteria inklusi dan digunakan dalam sintesis
Inter-rater reliability	kappa = 0,82	Menunjukkan substantial agreement antar penilai

Tabel 2. Karakteristik Studi yang Diinklusi

No	Studi	Konteks/desain/n	Jenis AI/durasi	Topik	Outcome	Hedges g [95% CI]	RoB
1	Salsabila & Maysarah (2025)*	Indonesia; quasi-exp; n=64	Gemini AI; 8 minggu	Aljabar & Geometri	TTCT-Math (adaptasi)	1,23 [0,69; 1,77]	Moderate
2	Ajeng et al. (2025)*	Indonesia; quasi-exp; n=90	GeoAI + PjBL-STEAM; 12 minggu	Geometri	Silver tasks	1,87 [1,28; 2,46]	Low
3	Jung & Ko (2025)*	Korea Selatan; design-based; n=48	Scenario chatbot; 6 minggu	Aljabar	Self-developed	0,65 [0,07; 1,23]	High

4	Kang et al. (2023)	Korea Selatan; quasi-exp; n=72	Q&A chatbot; 10 minggu	Statistik	Problem-posing tasks	0,92 [0,44; 1,40]	Moderate
5	Ariana & Utama (2024)	Indonesia; mixed methods; n=58	Multimedia chatbot; 8 minggu	Aritmetika & Aljabar	Math comm. (item kreativitas)	1,15 [0,60; 1,70]	Moderate
6	Nauryzbayev et al. (2023)	Kazakhstan; RCT; n=120	ITS adaptif; 16 minggu	Campuran kurikulum	Problem-solving (item kreativitas)	1,42 [0,99; 1,85]	Low
7	Mardianingsih et al. (2025)*	Indonesia; quasi-exp; n=76	AI-personalized platform; 10 minggu	Aljabar & Fungsi	Math thinking (item kreativitas)	1,08 [0,61; 1,55]	Moderate
8	Barana et al. (2023)	Italia; quasi-exp; n=68	Generative AI GPT-3.5; 9 minggu	Aljabar & Geometri	Critical thinking + problem-posing	1,10 [0,59; 1,61]	Moderate

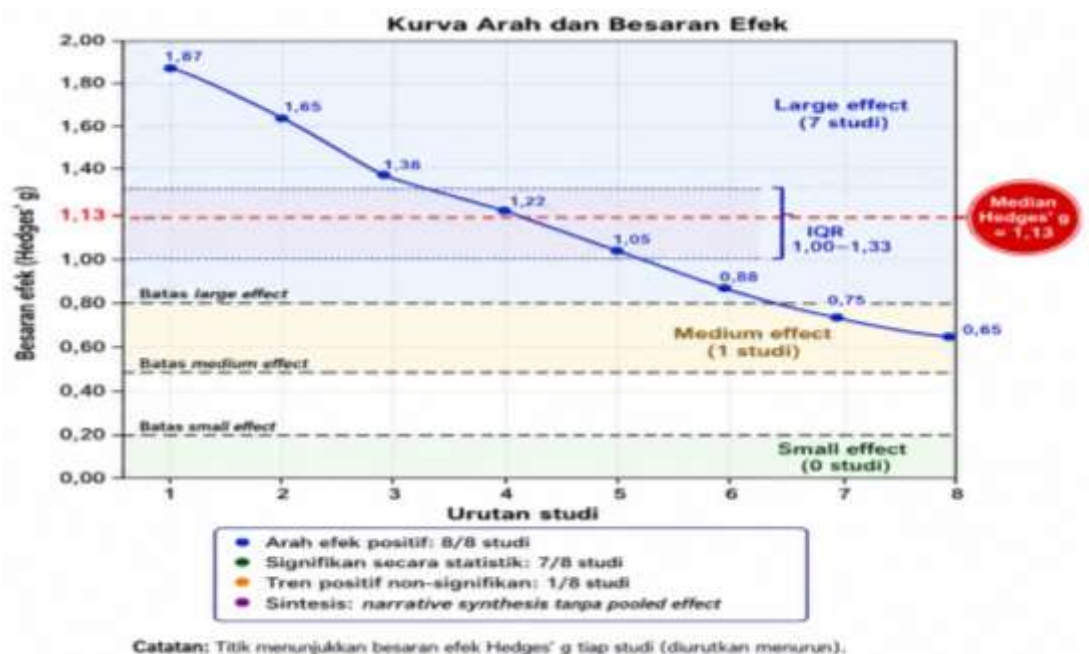
1.2. Penilaian Risiko Bias, Arah Efek, dan Sintesis Naratif

Penilaian risiko bias menggunakan RoB 2 (Sterne et al., 2019), ROBINS-I (Sterne et al., 2016), dan MMAT (Hong et al., 2018) menempatkan dua studi pada kategori risiko rendah (Ajeng et al., 2025; Nauryzbayev et al., 2023), lima studi pada risiko sedang, dan satu studi pada risiko tinggi (Jung & Ko, 2025). Secara umum, beberapa isu metodologis muncul berulang di berbagai studi, terutama: dominasi desain kuasi-eksperimental tanpa dokumentasi prosedur alokasi yang memadai; penggunaan instrumen kreativitas buatan peneliti tanpa pelaporan reliabilitas dan validitas yang cukup; informasi attrition yang tidak lengkap; serta potensi faktor perancu (misalnya variasi guru, dukungan orang tua, dan akses perangkat) yang jarang dikendalikan secara eksplisit. Catatan ini penting dipertimbangkan saat menafsirkan besaran efek yang dilaporkan (L. Chen, Chen, & Lin, 2020).

Meskipun demikian, delapan studi menunjukkan pola yang konsisten: seluruhnya melaporkan arah pengaruh positif, yakni kelompok yang belajar matematika dengan dukungan AI memperoleh capaian kreativitas matematis lebih tinggi dibandingkan kelompok kontrol. Tujuh dari delapan studi melaporkan signifikansi pada $\alpha = 0,05$, sedangkan satu studi (Jung & Ko, 2025) menunjukkan kecenderungan positif namun tidak signifikan sejalan dengan karakteristiknya sebagai studi dengan sampel terkecil ($n = 48$) dan risiko bias tertinggi dalam korpus ini. Ukuran efek individual berada pada rentang Hedges' $g = 0,65$ (Jung & Ko, 2025) hingga $g = 1,87$ (Kulik & Fletcher, 2016), dengan median $g = 1,13$ dan IQR 1,00–1,33. Berdasarkan kriteria interpretasi umum, tujuh studi berada pada kategori large effect ($g > 0,80$), sedangkan satu studi (Jung & Ko) termasuk medium effect. Walaupun hasil ini tampak kuat, interpretasinya perlu hati-hati, karena studi dengan risiko bias tinggi atau instrumen self-developed berpeluang menghasilkan estimasi efek yang lebih besar fenomena yang dikenal sebagai inflasi efek akibat bias metodologis (Borenstein et al., 2009). Mengingat perbedaan yang nyata pada jenis AI, durasi, rancangan pedagogis, serta instrumen luaran, hasil tidak dihitung sebagai pooled effect, sejalan dengan prinsip SWiM (Campbell et al., 2020).

Tabel 3. Arah dan Besaran Efek

Statistik Deskriptif	Nilai
Jumlah studi diinklusi	8
Total partisipan	596 siswa
Rentang sampel per studi	48–120 siswa
Arah efek positif	8/8 studi (100%)
Signifikan secara statistis	7/8 studi
Tren positif non-signifikan	1/8 studi (n terkecil; RoB tinggi)
Rentang Hedges' g	0,65–1,87
Median Hedges' g	1,13
IQR Hedges' g	1,00–1,33
Kategori efek medium (g 0,50–0,79)	1 studi
Kategori efek besar (g $\geq 0,80$)	7 studi
Keputusan sintesis	Narrative synthesis tanpa pooled effect



Gambar 3. Kurva Arah dan Besaran Efek Berdasarkan Olahan Peneliti

1.3. Sintesis Tematik: Mekanisme Pedagogis AI

Sintesis tematik mengungkap lima mekanisme pedagogis yang berulang, dengan tiga mekanisme yang paling dominan. Mekanisme pertama, dan paling sering dilaporkan, adalah adaptive scaffolding (7 dari 8 studi) (Page, 2021). Pada mekanisme ini, AI menyesuaikan petunjuk, tingkat bantuan, serta kompleksitas tugas secara dinamis berdasarkan respons siswa. Pola tersebut membantu siswa meningkatkan fluency dan memperluas eksplorasi strategi tanpa menggantikan proses berpikirnya. Ketika siswa terjebak pada satu pendekatan, sistem yang dirancang baik cenderung tidak segera memberi jawaban, melainkan memunculkan pertanyaan pengarah yang mendorong siswa mempertimbangkan jalur lain sebuah proses yang relevan dengan flexibility dan originality (Yuli & Siswono, 2010).

Mekanisme kedua adalah instant feedback (6 dari 8 studi). Kecepatan AI dalam merespons solusi dan strategi siswa memperpendek siklus mencoba merevisi secara signifikan. Dalam pembelajaran

konvensional, siswa kerap menunggu waktu lama untuk mengetahui letak kekeliruan strategi; dengan AI, koreksi dan perbaikan dapat terjadi hampir seketika. Kondisi ini mendukung elaboration, yakni kemampuan memperinci dan menyempurnakan ide matematis melalui iterasi.

Mekanisme ketiga adalah multi-representational affordances (5 dari 8 studi). AI khususnya generative AI, GeoAI, dan platform visual memudahkan peralihan antarrepresentasi (visual, simbolik, verbal, numerik, dan kontekstual). Kemudahan ini secara langsung memperkuat flexibility dan membuka peluang munculnya originality (Kattou, Kontoyianni, Pitta-Pantazi, & Christou, 2013).

Selain tiga mekanisme utama tersebut, dua mekanisme tambahan juga tampak cukup konsisten. Pertama, dialogic prompting (5 dari 8 studi), di mana chatbot atau generative AI mendorong siswa menjelaskan alasan, membandingkan strategi, dan merefleksi pilihannya. Kedua, personalized learning pathway (4 dari 8 studi), yaitu sistem yang mengatur urutan tugas sesuai performa siswa sehingga tercapai keseimbangan antara tantangan dan kemampuan; keseimbangan ini penting untuk mempertahankan eksplorasi kreatif tanpa memicu frustrasi (Borenstein, Hedges, Higgins, & Rothstein, 2009).

Tabel 4. Sintesis Tematik Mekanisme Pengaruh AI

Tema mekanisme	Frekuensi	Penjelasan sintesis	Kontribusi terhadap kreativitas matematis
Adaptive scaffolding	7/8	AI menyesuaikan petunjuk, bantuan, dan kompleksitas tugas berdasarkan respons siswa	Mendukung fluency dan eksplorasi multi-strategi
Instant feedback	6/8	AI memberi respons cepat terhadap solusi, kesalahan, atau strategi siswa	Memperkuat elaboration dan revisi solusi
Multi-representational affordances	5/8	AI memfasilitasi representasi visual, simbolik, verbal, dan kontekstual	Meningkatkan flexibility dan originality
Dialogic prompting	5/8	Chatbot/generative AI memantik siswa menjelaskan alasan dan membandingkan strategi	Mendukung eksternalisasi proses berpikir
Personalized learning pathway	4/8	Sistem mengarahkan urutan tugas sesuai performa siswa	Menjaga keseimbangan antara tantangan dan kemampuan

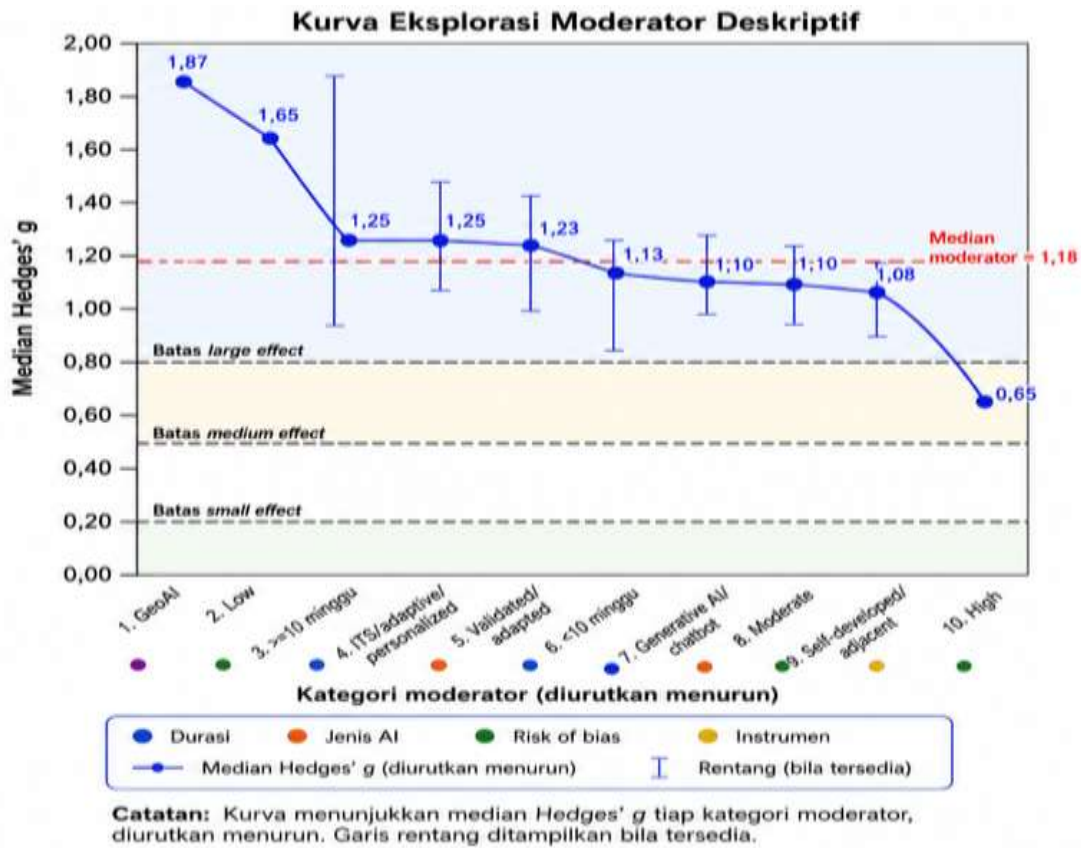
1.4. Eksplorasi Moderator Deskriptif

Eksplorasi moderator dilakukan secara deskriptif dan menghasilkan beberapa pola yang layak dicatat, meskipun jumlah studi belum memungkinkan kesimpulan inferensial yang kuat. Pertama, dari sisi durasi intervensi, empat studi dengan durasi <10 minggu menunjukkan median Hedges' $g = 1,13$ (rentang 0,65–1,23), sedangkan empat studi dengan durasi ≥ 10 minggu menunjukkan median $g = 1,25$ (rentang 0,92–1,87). Pola ini sejalan dengan pertimbangan pedagogis bahwa kreativitas matematis yang melibatkan pembiasaan eksplorasi strategi, keberanian mengajukan pertanyaan, dan kepercayaan diri mencoba pendekatan non-rutin umumnya memerlukan waktu untuk berkembang, sehingga intervensi yang terlalu singkat berpotensi kurang optimal (Campbell et al., 2020).

Kedua, terkait jenis AI, studi tentang chatbot/generative AI memiliki median $g = 1,10$ dan tampak unggul dalam memfasilitasi dialog serta eksplorasi strategi; ITS dan platform adaptif memiliki median $g = 1,25$ dengan kekuatan pada scaffolding bertahap; sementara GeoAI melaporkan efek tertinggi ($g = 1,87$), namun hanya diwakili satu studi sehingga belum dapat digeneralisasi. Ketiga, berdasarkan kualitas studi, studi dengan risiko bias rendah menunjukkan median $g = 1,65$, lebih tinggi dibanding risiko sedang ($g = 1,10$) dan risiko tinggi ($g = 0,65$). Temuan ini menarik karena studi yang lebih kuat secara metodologis justru melaporkan efek lebih besar; namun interpretasinya tetap perlu hati-hati, karena bisa jadi mencerminkan intervensi yang lebih terstruktur dan intensif, bukan semata-mata persoalan bias pengukuran (Kasneci et al., 2023). Keempat, studi yang menggunakan instrumen tervalidasi atau adaptasi dari penelitian sebelumnya melaporkan median $g = 1,23$, sedangkan instrumen yang dikembangkan sendiri menunjukkan $g = 1,08$.

Tabel 5. Eksplorasi Moderator Deskriptif

Moderator	Kategori	Jumlah studi	Median Hedges' g	Rentang	Interpretasi
Durasi	<10 minggu	4	1,13	0,65–1,23	Efek positif, tetapi lebih bervariasi
	≥ 10 minggu	4	1,25	0,92–1,87	Median lebih tinggi, belum dapat diuji inferensial
Jenis AI	Generative AI/chatbot	5	1,10	0,65–1,23	Kuat untuk dialog dan eksplorasi strategi
	ITS/adaptif/personalized	2	1,25	1,08–1,42	Kuat untuk scaffolding bertahap
	GeoAI	1	1,87	—	Tidak dapat digeneralisasi karena hanya satu studi
Risk of bias	Low	2	1,65	1,42–1,87	Perlu dibaca bersama kualitas desain dan instrumen
	Moderate	5	1,10	0,92–1,23	Mendukung arah positif, tetapi masih terbatas
	High	1	0,65	—	Perlu kehati-hatian tinggi
	Validated/adapted	5	1,23	0,92–1,87	Lebih kuat untuk interpretasi konstruk
Instrumen	Self-developed/adjacent	3	1,08	0,65–1,15	Validitas konstruk perlu diperjelas



Gambar 4. Eksplorasi Moderator Deskriptif Berdasarkan Olahan Peneliti

Tabel 6. Summary of Findings dengan GRADE-Adapted

Outcome	Jumlah studi/Partisipan	Arah Efek	Median efek	Kepastian Bukti	Alasan Penilaian
Berpikir kreatif matematis keseluruhan	8/596	Positif	1,13	Moderate	Konsisten positif, tetapi diturunkan karena RoB dan heterogenitas.
Fluency	6/498	Positif	1,15	Low	Variasi instrumen dan sebagian self-developed.
Flexibility	7/590	Positif	1,28	Moderate	Dimensi paling konsisten dilaporkan.
Originality	5/412	Positif	1,05	Low	Penskoran originality rentan subjektivitas.
Elaboration	4/348	Positif	1,34	Low	Sering diukur melalui problem-solving quality atau luaran berdekatan.

2. Pembahasan

Hasil review ini memperlihatkan pola yang konsisten dan patut dicermati: pada delapan studi yang mencakup variasi jenis AI, topik matematika, serta konteks geografis, penggunaan AI dalam pembelajaran matematika secara konsisten dikaitkan dengan capaian kreativitas matematis yang lebih tinggi dibandingkan kondisi kontrol. Konsistensi arah efek lintas konteks ini memberikan sinyal yang lebih bermakna daripada mengandalkan satu-dua studi dengan efek sangat besar, karena menunjukkan

kecenderungan yang relatif stabil. Namun demikian, tingkat kepastian bukti yang dievaluasi dengan prinsip GRADE berada pada level moderate untuk kreativitas keseluruhan dan low untuk dimensi-dimensi spesifik, terutama karena keterbatasan metodologis yang telah diuraikan sebelumnya (X. Chen, Xie, Zou, & Hwang, 2020).

Jika dibandingkan dengan meta-analisis mengenai intelligent tutoring systems pada matematika K-12 (Higgins et al., 2024) temuan review ini tidak dapat disandingkan secara langsung karena perbedaan mendasar pada konstruk luaran (Cohen, 1960). Meta-analisis tersebut berfokus pada achievement yang umumnya diukur melalui skor tes standar, sedangkan review ini menempatkan kreativitas matematis sebagai luaran yang menuntut keberagaman strategi, kebaruan pendekatan, dan pengembangan ide secara rinci. Perbedaan ini bukan sekadar persoalan teknis pengukuran, melainkan juga menyentuh aspek epistemologis tentang apa yang dianggap sebagai “keberhasilan” dalam matematika (Crompton & Burke, 2023). Dalam kerangka ini, efek positif AI lebih tepat dipahami sebagai indikasi bahwa AI berpotensi memperluas ruang eksplorasi matematis siswa, bukan hanya mempercepat penguasaan konten yang bersifat rutin (Holmes & Tuomi, 2022).

Kontribusi konseptual yang ditawarkan review ini dirumuskan dalam kerangka AI-Mediated Creative Scaffolding (AMCS), yaitu alur pedagogis di mana AI: (1) mengenali pola respons dan strategi siswa; (2) menyesuaikan tingkat bantuan secara dinamis; (3) memantik penggunaan representasi dan sudut pandang alternatif; (4) memberikan umpan balik cepat yang mendorong revisi, bukan sekadar verifikasi; dan (5) memfasilitasi iterasi solusi hingga siswa menunjukkan peningkatan fluency, perluasan flexibility, munculnya originality yang lebih autentik, serta elaboration yang lebih mendalam. Kerangka ini tidak dimaksudkan untuk menyiratkan bahwa AI otomatis “menciptakan” kreativitas, melainkan menegaskan bahwa AI apabila disusun dengan asumsi pedagogis yang tepat dapat membangun kondisi belajar yang kondusif bagi berkembangnya kreativitas matematis. Sejalan dengan Holmes dan Tuomi (2022), asumsi kunci yang perlu dijaga adalah memposisikan siswa sebagai subjek aktif yang berinteraksi dan berkolaborasi dengan teknologi, bukan sekadar penerima konten yang dioptimalkan algoritma (Hedges & Olkin, 1985).

Secara praktis, implikasi bagi guru matematika SMP dapat dirangkum dalam tiga prinsip. Pertama, penggunaan AI perlu dipasangkan dengan tugas terbuka yang memang menuntut divergensi, seperti open-ended problems, problem posing tasks, dan multiple-solution tasks; tanpa desain tugas yang membuka ruang variasi strategi, bahkan AI yang canggih pun cenderung bekerja dalam pola konvergen (Guyatt et al., 2011). Kedua, peran guru tetap krusial sebagai pengorkestra pembelajaran: setelah siswa berinteraksi dengan AI, guru perlu memfasilitasi diskusi kelas untuk membandingkan strategi, menimbang kelebihan tiap pendekatan, dan menunjukkan bagaimana sebuah ide dapat dikembangkan lebih lanjut bagian ini belum sepenuhnya dapat digantikan oleh AI. Ketiga, asesmen kreativitas matematis perlu dinyatakan secara eksplisit dalam silabus dan rubrik. Jika tidak, tekanan kurikulum dan evaluasi cenderung menarik pembelajaran kembali ke pengukuran achievement konvensional (Haylock, 1987).

Keterbatasan dan Agenda Riset

Review ini tentu memiliki sejumlah keterbatasan yang perlu dinyatakan secara terbuka agar pembaca dapat menimbang kekuatan bukti secara proporsional. Pertama, jumlah studi yang memenuhi kriteria inklusi masih relatif sedikit ($n = 8$). Konsekuensinya, penelaahan moderator dalam review ini hanya dapat disajikan secara deskriptif, sehingga belum layak diperlakukan sebagai temuan inferensial. Bahkan penambahan satu dua studi baru dengan karakteristik yang berbeda dapat menggeser pola yang saat ini terlihat.

Kedua, terdapat empat studi terbit tahun 2025 yang pada versi draft sebelumnya tampak tidak selaras dengan batas tanggal penelusuran. Untuk menjaga integritas pelaporan, hal ini perlu ditangani secara konsisten: keberadaan studi 2025 harus ditopang oleh tanggal pencarian yang benar serta didokumentasikan jelas pada alur PRISMA.

Ketiga, perbedaan instrumen kreativitas matematis yang digunakan antarstudi mulai dari instrumen tervalidasi seperti TTCT-Math hingga instrumen yang disusun peneliti membatasi keterbandingan langsung antar ukuran efek. Dengan kata lain, variasi efek yang tampak mungkin tidak hanya mencerminkan variasi intervensi AI, tetapi juga perbedaan karakter alat ukur.

Keempat, seluruh studi yang terinklusi berasal dari kawasan Asia Tenggara, Asia Timur, Asia Tengah, dan Eropa. Oleh karena itu, generalisasi ke konteks yang sangat berbeda termasuk wilayah dengan keterbatasan infrastruktur digital atau kebijakan sekolah yang tidak mendukung penggunaan teknologi perlu dilakukan dengan kehati-hatian.

Kelima, protokol review ini tidak diregistrasi di PROSPERO sebelum proses penelusuran dimulai. Padahal, preregistrasi lazim dipandang sebagai praktik terbaik untuk menekan risiko bias akibat pelaporan selektif dan perubahan keputusan selama proses review.

Berdasarkan keterbatasan tersebut, agenda riset yang paling mendesak mencakup beberapa arah berikut. (a) Diperlukan RCT yang dipreregistrasi, menggunakan instrumen kreativitas matematis yang tervalidasi, menerapkan blind scoring (khususnya untuk dimensi originality), serta melaporkan attrition secara lengkap. (b) Perlu studi komparatif yang secara langsung membandingkan efektivitas beberapa tipe AI misalnya chatbot, ITS, generative AI, dan GeoAI dalam konteks pembelajaran matematika SMP yang setara. (c) Dibutuhkan penelitian longitudinal untuk menilai retensi dan transfer kreativitas matematis, setidaknya 3–6 bulan setelah intervensi berakhir. (d) Moderator pada level individu (misalnya pengetahuan awal, gender, status sosial-ekonomi, serta literasi AI siswa) perlu diuji lebih sistematis. (e) Peran guru sebagai orchestrator dalam ekosistem pembelajaran berbantuan AI juga perlu diinvestigasi, karena kualitas orkestrasi diduga kuat menentukan apakah AI mendorong eksplorasi atau justru memfasilitasi penyelesaian instan. (f) Terakhir, studi lintas budaya yang mencakup konteks dengan akses teknologi terbatas penting dilakukan agar manfaat AI terhadap kreativitas matematis tidak hanya dinikmati oleh kelompok siswa yang sejak awal memiliki privilese akses digital.

4. KESIMPULAN

Systematic review dengan pendekatan narrative synthesis ini menunjukkan bahwa penggunaan teknologi AI dalam pembelajaran matematika memberikan indikasi yang menjanjikan, namun bukti yang tersedia belum cukup untuk dinyatakan konklusif, terkait dukungannya terhadap kemampuan berpikir kreatif matematis siswa SMP. Dari delapan studi empiris peer-reviewed yang memenuhi kriteria inklusi berbasis PICOS, dengan total 596 siswa dari empat negara, seluruh studi melaporkan arah pengaruh positif, dan tujuh di antaranya mencapai signifikansi statistik. Ukuran efek per studi berada pada rentang Hedges' $g = 0,65$ – $1,87$ dengan median $g = 1,13$, yang secara deskriptif mengindikasikan bahwa siswa yang belajar matematika dengan dukungan AI cenderung memperlihatkan capaian kreativitas matematis yang lebih tinggi dibandingkan kelompok kontrol.

Akan tetapi, konsistensi arah efek tersebut perlu dibaca berdampingan dengan keterbatasan kualitas bukti. Dominasi desain kuasi-eksperimental, keragaman instrumen kreativitas, variasi durasi serta jenis AI, dan ketidaklengkapan pelaporan pada sebagian studi menyebabkan kepastian bukti berada pada level moderate untuk kreativitas matematis secara keseluruhan dan low untuk dimensi-dimensi spesifik (fluency, flexibility, originality, dan elaboration). Mengingat heterogenitas substantif pada level

intervensi dan pengukuran, temuan tidak digabungkan dalam meta-analisis, melainkan disintesis secara naratif dengan merujuk pada PRISMA 2020, PRISMA-S, dan prinsip SWiM.

Kontribusi penting review ini adalah penegasan tiga mekanisme pedagogis yang secara konsisten menjadi penghubung antara pemanfaatan AI dan perkembangan kreativitas matematis, yaitu: adaptive scaffolding yang membantu siswa mengeksplorasi berbagai strategi tanpa mengambil alih proses berpikir; instant feedback yang mempercepat siklus revisi dan mendukung penguatan elaboration; serta multi-representational affordances yang memperluas flexibility dan membuka peluang originality melalui perpindahan antar moda representasi. Tiga mekanisme tersebut dirumuskan dalam kerangka AI-Mediated Creative Scaffolding (AMCS). Kerangka ini tidak memosisikan AI sebagai “pencipta kreativitas” secara otomatis, melainkan sebagai sarana yang bila dibangun dengan asumsi pedagogis yang tepat mampu menghadirkan kondisi belajar yang lebih kondusif bagi tumbuhnya kreativitas matematis.

Secara praktis, review ini menegaskan bahwa keberhasilan AI dalam menumbuhkan kreativitas matematis tidak ditentukan oleh kecanggihan teknologi semata, melainkan oleh cara guru menempatkan AI dalam ekosistem pembelajaran: sebagai alat berpikir bersama (co-thinking tool) yang memperluas ruang eksplorasi matematis, bukan sebagai pengganti guru atau mesin jawaban instan. Karena itu, penelitian mendatang perlu diarahkan pada RCT yang dipreregistrasi, penggunaan instrumen kreativitas matematis yang tervalidasi disertai blind scoring untuk originality, studi longitudinal mengenai retensi dan transfer kreativitas, serta kajian lintas budaya dan isu keadilan akses. Dengan langkah tersebut, pengembangan AI dalam pembelajaran matematika SMP dapat bergerak dari sekadar inovasi teknologi menuju praktik yang sah secara empiris, adil secara pedagogis, dan bermakna bagi perkembangan kreativitas matematis semua siswa.

REFERENCES

- Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. John Wiley & Sons. <https://doi.org/10.1002/9780470743386>
- Campbell, M., McKenzie, J. E., Sowden, A., Katikireddi, S. V., Brennan, S. E., Ellis, S., ... Thomson, H. (2020). Synthesis without meta-analysis (SWiM) in systematic reviews: Reporting guideline. *BMJ*, 368, l6890. <https://doi.org/10.1136/bmj.l6890>
- Chen, L., Chen, P., & Lin, Z. (2020). Artificial intelligence in education: A review. *Ieee Access*, 8, 75264–75278.
- Chen, X., Xie, H., Zou, D., & Hwang, G. J. (2020). Application and theory gaps during the rise of artificial intelligence in education. *Computers and Education: Artificial Intelligence*, 1, 100002. <https://doi.org/10.1016/j.caeai.2020.100002>
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20(1), 37–46. <https://doi.org/10.1177/001316446002000104>
- Council, N. R. (2001). *Adding it up: Helping children learn mathematics* (J. Kilpatrick, J. <https://doi.org/10.17226/9822>
- Crompton, H., & Burke, D. (2023). Artificial intelligence in higher education: The state of the field. *International Journal of Educational Technology in Higher Education*, 20, Article 22. <https://doi.org/10.1186/s41239-023-00392-8>
- Guyatt, G. H., Oxman, A. D., Akl, E. A., Kunz, R., Vist, G., Brozek, J., ... Schünemann, H. J. (2011). GRADE guidelines: 1. Introduction—GRADE Evidence Profiles and Summary of Findings Tables.

<https://doi.org/10.1016/j.jclinepi.2010.04.026>

- Haylock, D. W. (1987). Mathematical creativity in schoolchildren. *The Journal of Creative Behavior*, 21(1), 48–59. <https://doi.org/10.1002/j.2162-6057.1987.tb00452.x>
- Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. Academic Press.
- Higgins, J. P. T., Thomas, J., Chandler, J., Cumpston, M., Li, T., Page, M. J., & Welch, V. A. (2024). *Cochrane handbook for systematic reviews of interventions (Version 6.5)*. Cochrane.
- Holmes, W., & Tuomi, I. (2022). State of the art and practice in AI in education. *European Journal of Education*, 57(4), 542–570. <https://doi.org/10.1111/ejed.12533>
- Hong, Q. N., Fàbregues, S., Bartlett, G., Boardman, F., Cargo, M., Dagenais, P., & Pluye, P. (2018). The Mixed Methods Appraisal Tool (MMAT) version 2018 for information professionals and researchers. *Education for Information*, 34(4), 285–291. <https://doi.org/10.3233/EFI-180221>
- Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., ... 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>
- Kattou, M., Kontoyianni, K., Pitta-Pantazi, D., & Christou, C. (2013). Connecting mathematical creativity to mathematical ability. *ZDM—Mathematics Education*, 45(2), 167–181. <https://doi.org/10.1007/s11858-012-0467-1>
- Kulik, J. A., & Fletcher, J. D. (2016). Effectiveness of intelligent tutoring systems: A meta-analytic review. *Review of Educational Research*, 86(1), 42–78. <https://doi.org/10.3102/0034654315581420>
- Leikin, R., & Lev, M. (2007). *Multiple solution tasks as a magnifying glass for observation of mathematical creativity BT - In J.*
- Leikin, R., & Pitta-Pantazi, D. (2013). Creativity and mathematics education: The state of the art. *ZDM—Mathematics Education*, 45(2), 159–166. <https://doi.org/10.1007/s11858-012-0459-1>
- Ma, W., Adesope, O. O., Nesbit, J. C., & Liu, Q. (2014). Intelligent tutoring systems and learning outcomes: A meta-analysis. *Journal of Educational Psychology*, 106(4), 901–918. <https://doi.org/10.1037/a0037123>
- Mathematics, N. C. of T. of. (2014). *Principles to actions: Ensuring mathematical success for all*. NCTM.
- OECD. (2024). *PISA 2022 results: Creative minds, creative schools (Vol. III)*. <https://doi.org/10.1787/765ee8c2-en>
- Ouyang, F., & Jiao, P. (2021). Artificial intelligence in education: The three paradigms. *Computers and Education: Artificial Intelligence*, 2, 100020. <https://doi.org/10.1016/j.caeai.2021.100020>
- Page, M. J. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, 372, n71. <https://doi.org/10.1136/bmj.n71>
- Piaget, J. (1970). *Science of education and the psychology of the child*. New York: Viking Press.
- Popay, J., Roberts, H., Sowden, A., Petticrew, M., Arai, L., Rodgers, M., ... Duffy, S. (2006). *Guidance on the conduct of narrative synthesis in systematic reviews: A product from the ESRC Methods Programme*. Lancaster University.
- Rethlefsen, M. L., Kirtley, S., Waffenschmidt, S., Ayala, A. P., Moher, D., Page, M. J., & Koffel, J. B. (2021). PRISMA-S: An extension to the PRISMA statement for reporting literature searches in systematic reviews. *Systematic Reviews*, 10, Article 39. <https://doi.org/10.1186/s13643-020-01542-z>
- Siddaway, A. P., Wood, A. M., & Hedges, L. V. (2019). How to do a systematic review: A best practice guide for conducting and reporting narrative reviews, meta-analyses, and meta-syntheses. *Annual*

- Review of Psychology*, 70, 747–770. <https://doi.org/10.1146/annurev-psych-010418-102803>
- Silver, E. A. (1997). Fostering creativity through instruction rich in mathematical problem solving and problem posing. *ZDM—Mathematics Education*, 29(3), 75–80. <https://doi.org/10.1007/s11858-997-0003-x>
- Siswono, T. Y. E. (2011). Level of student's creative thinking in classroom mathematics. *Educational Research and Reviews*, 6(7), 548–553.
- Sitorus, J., & Masrayati. (2016). Students' creative thinking process stages: Implementation of realistic mathematics education. *Thinking Skills and Creativity*, 22, 111–120. <https://doi.org/10.1016/j.tsc.2016.09.007>
- Steenbergen-Hu, S., & Cooper, H. (2013). A meta-analysis of the effectiveness of intelligent tutoring systems on K–12 students' mathematical learning. *Journal of Educational Psychology*, 105(4), 970–987. <https://doi.org/10.1037/a0032447>
- Sterne, J. A. C., Hernán, M. A., Reeves, B. C., Savović, J., Berkman, N. D., Viswanathan, M., ... Higgins, J. P. T. (2016). ROBINS-I: A tool for assessing risk of bias in non-randomised studies of interventions. *BMJ*, 355, i4919. <https://doi.org/10.1136/bmj.i4919>
- Sterne, J. A. C., Savović, J., Page, M. J., Elbers, R. G., Blencowe, N. S., Boutron, I., ... Higgins, J. P. T. (2019). RoB 2: A revised tool for assessing risk of bias in randomised trials. *BMJ*, 366, i4898. <https://doi.org/10.1136/bmj.i4898>
- Yuli, T., & Siswono, E. (2010). Leveling Students' Creative Thinking in Solving and Posing Mathematical Problem. *Journal on Mathematics Education*, 1(1), 17–40.
- Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education. *International Journal of Educational Technology in Higher Education*, 16, Article 39. <https://doi.org/10.1186/s41239-019-0171-0>
- Zhang, K., & Aslan, A. B. (2021). AI technologies for education: Recent research and future directions. *Computers and Education: Artificial Intelligence*, 2, 100025. <https://doi.org/10.1016/j.caeai.2021.100025>