

**Digital Transformation In Islamic Religious Education: The Role Of Chatgpt In Developing Critical Thinking Skills Among Secondary School Students**

**Transformasi Digital Dalam Pendidikan Agama Islam: Peran Chatgpt Dalam Mengembangkan Keterampilan Berpikir Kritis Di Kalangan Siswa Sekolah Menengah**

**Eficha Hadi Putri<sup>1</sup>, Rifa'i<sup>2</sup>, Lety Febriana<sup>3</sup>**

Universitas Muhammadiyah Bengkulu<sup>1,2,3</sup>

Email: <sup>1</sup>[efichadi06@gmail.com](mailto:efichadi06@gmail.com), <sup>2</sup>[rifa'i@umb.ac.id](mailto:rifa'i@umb.ac.id), <sup>3</sup>[letyfebriana@umb.ac.id](mailto:letyfebriana@umb.ac.id)

\*Corresponding Author

---

Received : 15 November 2025, Revised : 20 December 2025, Accepted : 23 January 2026

---

**ABSTRACT**

*The digital transformation marked by the presence of generative artificial intelligence such as ChatGPT has had a significant impact on the educational ecosystem, including on Islamic Religious Education (PAI) learning. This change opens opportunities for students to explore religious material more interactively, but also poses challenges related to the validity of information, academic ethics, and the potential weakening of critical thinking skills due to the tendency to seek instant answers. This study aims to assess the effect of using ChatGPT on high school students' critical thinking skills in Islamic Religious Education learning using a quantitative approach through a non-equivalent control group quasi-experimental design. The sample involved 30 10th-grade students divided into a control group and an experimental group. The critical thinking test instrument was designed based on an adaptation of the Watson-Glaser Critical Thinking Appraisal and has undergone validity and reliability tests with a Cronbach's Alpha result of 0.924, indicating very high internal consistency. Data analysis was performed using the Mann–Whitney U test because the requirements for normality and homogeneity of variance were not met. The results showed a significant difference between the posttest scores of the experimental and control groups ( $p = 0.007$ ), with an effectiveness value of  $r \approx 0.49$ , which is categorized as moderate. These findings indicate that the targeted use of ChatGPT can improve students' critical thinking skills in Islamic Religious Education (PAI) learning. This research emphasizes the importance of integrating artificial intelligence pedagogically, ethically, and theologically to ensure students not only acquire instant knowledge but also develop in-depth and responsible reasoning within an Islamic context.*

**Keywords:** Chatgpt, Critical Thinking, Islamic Religious Education, Artificial Intelligence, Digital Learning.

**ABSTRAK**

Transformasi digital yang ditandai dengan hadirnya kecerdasan buatan generatif seperti ChatGPT membawa pengaruh signifikan terhadap ekosistem pendidikan, termasuk pada pembelajaran Pendidikan Agama Islam (PAI). Perubahan ini membuka peluang bagi siswa untuk mengeksplorasi materi keagamaan secara lebih interaktif, namun sekaligus menimbulkan tantangan yang berkaitan dengan validitas informasi, etika akademik, serta potensi melemahnya kemampuan berpikir kritis akibat kecenderungan mencari jawaban instan. Penelitian ini bertujuan menilai pengaruh penggunaan ChatGPT terhadap kemampuan berpikir kritis siswa SMA dalam pembelajaran PAI dengan menggunakan pendekatan kuantitatif melalui desain eksperimen semu tipe non-equivalent control group. Sampel melibatkan 30 siswa kelas X yang dibagi menjadi kelompok kontrol dan kelompok eksperimen. Instrumen tes berpikir kritis disusun berdasarkan adaptasi Watson-Glaser Critical Thinking Appraisal dan telah melalui uji validitas serta reliabilitas dengan hasil Cronbach's Alpha sebesar 0,924, menunjukkan konsistensi internal sangat tinggi. Analisis data dilakukan menggunakan uji Mann–Whitney U karena syarat normalitas dan homogenitas varians tidak terpenuhi. Hasil penelitian menunjukkan adanya perbedaan signifikan antara skor posttest kelompok eksperimen dan kontrol ( $p = 0,007$ ), dengan nilai efektivitas  $r \approx 0,49$  yang termasuk kategori sedang. Temuan ini mengindikasikan bahwa penggunaan ChatGPT secara terarah dapat meningkatkan kemampuan berpikir kritis siswa dalam pembelajaran PAI. Penelitian ini menegaskan pentingnya integrasi kecerdasan buatan secara pedagogis, etis, dan teologis guna memastikan siswa tidak

hanya memperoleh pengetahuan instan, tetapi mampu mengembangkan penalaran yang mendalam dan bertanggung jawab dalam konteks keislaman.

**Kata Kunci:** Chatgpt, Berpikir Kritis, Pendidikan Agama Islam, Kecerdasan Buatan, Pembelajaran Digital.

## 1. Pendahuluan

Transformasi digital pada era global telah membawa perubahan fundamental dalam cara manusia berpikir, belajar, dan berinteraksi. Perkembangan kecerdasan buatan generatif (generative artificial intelligence), seperti ChatGPT, menjadi salah satu inovasi paling disruptif dalam dunia pendidikan kontemporer. Secara global, AI generatif telah dimanfaatkan untuk mendukung pembelajaran personal, diskusi berbasis teks, hingga asesmen formatif di berbagai jenjang pendidikan, termasuk sekolah menengah (OECD, 2023). Di Indonesia, pemanfaatan teknologi digital dalam pendidikan semakin masif seiring kebijakan transformasi digital pendidikan dan implementasi Kurikulum Merdeka yang menekankan penguatan higher order thinking skills (HOTS). Namun demikian, adopsi AI generatif di sekolah menengah masih memunculkan perdebatan, terutama terkait dampaknya terhadap proses berpikir kritis, etika akademik, dan pembentukan karakter peserta didik.

Dalam perspektif Islam, penggunaan teknologi tidak dapat dilepaskan dari mandat ilahiah untuk menggunakan akal secara bertanggung jawab. Al-Qur'an menegaskan pentingnya proses membaca, belajar, dan pengembangan ilmu pengetahuan sebagaimana tercantum dalam QS. Al-'Alaq ayat 1–5. Ayat tersebut menunjukkan bahwa aktivitas kognitif bukan sekadar teknis, melainkan bagian dari ibadah intelektual yang menuntut kesadaran, tanggung jawab, dan kedalaman makna. Oleh karena itu, integrasi kecerdasan buatan dalam Pendidikan khususnya Pendidikan Agama Islam (PAI) tidak dapat dipahami secara netral, melainkan harus dikaji secara kritis dalam bingkai epistemologis, pedagogis, dan teologis.

اَفْرُّ بِاسْمِ رَبِّكَ الَّذِي خَلَقَ ○ خَلَقَ الْإِنْسَانَ مِنْ عَلْقٍ ○ اَفْرُّ وَرَبِّكَ الْأَكْرَمُ ○ الَّذِي عَلَمَ بِالْقَلْمَنِ ○ عَلَمَ الْإِنْسَانَ  
مَا لَمْ يَعْلَمْ

“Bacalah dengan (menyebut) nama Tuhanmu yang menciptakan. Dia telah menciptakan manusia dari segumpal darah. Bacalah, dan Tuhanmulah Yang Maha Pemurah, yang mengajar (manusia) dengan pena. Dia mengajarkan manusia apa yang tidak diketahuinya.”

Pendidikan Agama Islam memiliki karakteristik yang unik dibandingkan mata pelajaran lainnya. PAI tidak hanya bersifat kognitif-informatif, tetapi juga normatif dan transformatif, karena bertujuan membentuk keimanan, akhlak, serta kesadaran spiritual peserta didik. Proses pembelajaran PAI idealnya tidak berhenti pada transfer pengetahuan (ta'lim), melainkan menekankan pembentukan adab dan kebijaksanaan (ta'dib). Dalam konteks ini, penggunaan ChatGPT menjadi isu yang relevan sekaligus problematis. Di satu sisi, AI memungkinkan siswa memperoleh penjelasan tentang ayat Al-Qur'an, hadis, sejarah Islam, dan nilai akhlak secara cepat dan variatif. Di sisi lain, kemudahan tersebut berpotensi mendorong pola belajar instan yang melemahkan proses refleksi, internalisasi nilai, dan pendalaman makna ajaran Islam (Sujarwo, 2022).

Kemampuan berpikir kritis merupakan keterampilan kunci abad ke-21 yang selaras dengan ajaran Islam tentang pentingnya pemahaman mendalam (fiqh). Hadis Nabi SAW menegaskan bahwa kebaikan seorang mukmin ditandai dengan kemampuannya memahami agama secara mendalam (HR. Bukhari). Namun, realitas pendidikan menunjukkan tantangan serius. Hasil PISA 2022 memperlihatkan bahwa kemampuan berpikir kreatif siswa Indonesia—yang berkorelasi erat dengan berpikir kritis masih berada jauh di bawah rata-rata negara OECD (Bilad et al., 2024; PISA Creative Thinking Framework, 2024). Temuan ini mengindikasikan bahwa penguasaan HOTS masih menjadi persoalan mendasar, termasuk dalam pembelajaran PAI yang selama ini cenderung didominasi metode ceramah dan hafalan (Kasman et al., 2022).

"مَنْ يُرِدَ اللَّهُ بِهِ خَيْرًا يُفَقِّهُ فِي الدِّينِ"

"Barang siapa dikehendaki Allah menjadi baik, maka Allah akan memahamkannya dalam agama" (HR. Bukhari).

Dalam konteks tersebut, ChatGPT menawarkan potensi pedagogis yang signifikan. AI generatif ini mampu menyediakan umpan balik instan, memfasilitasi dialog berbasis pertanyaan, serta menyajikan beragam sudut pandang yang dapat merangsang analisis dan evaluasi siswa. Sejumlah penelitian menunjukkan bahwa integrasi ChatGPT dapat meningkatkan kualitas diskusi, partisipasi belajar, dan argumentasi siswa (Farrokhnia et al., 2024; Lin et al., 2024). Namun, dalam pembelajaran PAI, potensi ini tidak dapat dilepaskan dari risiko epistemologis dan etis. ChatGPT bekerja berdasarkan prediksi probabilistik bahasa, bukan pada otoritas wahyu, sehingga informasi keagamaan yang dihasilkan berpotensi mengandung bias atau kekeliruan teologis apabila tidak diverifikasi secara kritis (Ocampo & Gozum, 2024).

Selain itu, penggunaan AI tanpa pendampingan pedagogis berisiko menggeser peran guru sebagai pembimbing spiritual dan moral. Ketergantungan berlebihan pada jawaban instan dapat melemahkan proses refleksi, dialog, dan pembentukan adab digital. Penelitian terbaru menunjukkan bahwa penggunaan ChatGPT yang tidak terkontrol dapat menghambat perkembangan berpikir kritis dan kreativitas siswa serta meningkatkan risiko plagiarisme akademik (Williams, 2024; Chanda, 2025). Dalam pendidikan agama, persoalan ini menjadi lebih serius karena AI tidak memiliki kesadaran moral, nurani, maupun dimensi spiritual.

Permasalahan penggunaan ChatGPT dalam pembelajaran PAI dapat dipetakan ke dalam empat aspek utama: epistemologis (akurasi dan otoritas pengetahuan agama), pedagogis (pergeseran peran guru dan ketergantungan siswa), etis (integritas akademik dan adab digital), serta spiritual (kedalaman penghayatan iman). Meskipun demikian, kajian empiris yang secara khusus meneliti dampak penggunaan ChatGPT terhadap kemampuan berpikir kritis siswa dalam pembelajaran PAI di tingkat SMA masih sangat terbatas, baik di Indonesia maupun dalam konteks global (Abbas et al., 2024). Keterbatasan inilah yang menjadi research gap utama dalam penelitian ini.

Berdasarkan celah tersebut, penelitian ini bertujuan untuk mengkaji secara empiris pengaruh penggunaan ChatGPT terhadap kemampuan berpikir kritis siswa dalam pembelajaran Pendidikan Agama Islam di sekolah menengah. Penelitian ini diharapkan memberikan kontribusi teoretis dalam pengembangan pedagogi PAI berbasis teknologi, sekaligus kontribusi praktis bagi guru dan membuat kebijakan dalam merumuskan pemanfaatan AI yang etis, kritis, dan selaras dengan nilai-nilai Islam. Dengan demikian, integrasi ChatGPT tidak diposisikan sebagai pengganti peran guru, melainkan sebagai alat bantu untuk memperkuat proses berpikir reflektif dan pembentukan adab dalam bingkai iman dan ilmu.

## 2. Metode Penelitian

Penelitian ini menggunakan pendekatan kuantitatif dengan metode eksperimen semu (quasi-experimental) melalui desain non-equivalent control group. Pendekatan kuantitatif dipilih karena memungkinkan pengukuran pengaruh penggunaan ChatGPT terhadap kemampuan berpikir kritis siswa secara objektif melalui analisis statistik inferensial (Sugiyono, 2021). Desain non-equivalent control group digunakan karena peneliti tidak memungkinkan melakukan pengacakan subjek secara penuh, mengingat pembagian kelas telah ditetapkan oleh pihak sekolah. Meskipun demikian, desain ini tetap relevan dan valid secara ilmiah untuk menilai efektivitas suatu perlakuan pembelajaran dengan membandingkan hasil belajar antara kelompok eksperimen dan kelompok kontrol yang memiliki karakteristik relatif sebanding (Creswell & Creswell, 2023).

Penelitian dilaksanakan di SMA Muhammadiyah 4 Bengkulu, sebuah sekolah menengah atas swasta berbasis keagamaan Islam, pada semester ganjil tahun ajaran 2025/2026. Pemilihan lokasi penelitian didasarkan pada beberapa pertimbangan, yaitu ketersediaan sarana teknologi digital (akses internet dan perangkat pendukung), penerapan Kurikulum Merdeka, serta kesiapan sekolah dalam mengimplementasikan inovasi pembelajaran berbasis artificial intelligence secara terbatas dan terkontrol (Rahmawati, 2022). Penelitian berlangsung selama enam minggu yang meliputi tahap persiapan instrumen dan sosialisasi penggunaan ChatGPT, pelaksanaan perlakuan pembelajaran, serta evaluasi akhir berupa posttest.

Populasi penelitian adalah seluruh siswa kelas X SMA Muhammadiyah 4 Bengkulu yang berjumlah 30 siswa. Karakteristik sampel meliputi siswa dengan rentang usia 15–16 tahun, latar belakang akademik yang relatif homogen, serta kemampuan awal berpikir kritis yang berada pada kategori sedang berdasarkan hasil pretest. Mengingat jumlah populasi yang kecil dan homogen, penelitian ini menggunakan teknik total sampling, sehingga seluruh populasi dijadikan sampel penelitian (Siyoto & Sodik, 2023). Kelas X-A ditetapkan sebagai kelompok kontrol yang menerima pembelajaran PAI dengan metode konvensional, sedangkan kelas X-B sebagai kelompok eksperimen yang memperoleh perlakuan pembelajaran berbantuan ChatGPT.

Variabel bebas dalam penelitian ini adalah penggunaan ChatGPT dalam pembelajaran Pendidikan Agama Islam. Perlakuan pada kelompok eksperimen dilaksanakan selama empat kali pertemuan inti (masing-masing  $2 \times 45$  menit). Dalam setiap pertemuan, guru berperan sebagai fasilitator yang mengawali pembelajaran dengan penyampaian tujuan dan pengantar materi. Selanjutnya, siswa diberikan tugas terstruktur untuk berinteraksi dengan ChatGPT guna mengeksplorasi materi “penyakit hati dan pribadi kaya hati” melalui pertanyaan analitis, studi kasus sederhana, dan klarifikasi konsep. Setelah sesi interaksi mandiri, guru memandu diskusi reflektif di kelas untuk mengkaji kembali jawaban ChatGPT, meluruskan potensi kekeliruan, serta menanamkan nilai-nilai akhlak dan etika Islam. Prosedur ini dirancang untuk memastikan bahwa ChatGPT berfungsi sebagai alat bantu pembelajaran (AI-assisted learning), bukan sebagai pengganti peran guru (Farrokhnia et al., 2024; Williams, 2024).

Variabel terikat dalam penelitian ini adalah kemampuan berpikir kritis siswa yang mencakup kemampuan membuat inferensi, menganalisis asumsi, melakukan deduksi, menginterpretasikan informasi, serta mengevaluasi argumen. Pengukuran dilakukan menggunakan instrumen tes yang diadaptasi dari Watson-Glaser Critical Thinking Appraisal (WGCTA) dan disesuaikan dengan konteks materi PAI. Instrumen terdiri atas 25 butir soal pilihan ganda yang mengukur keterampilan berpikir tingkat tinggi siswa (Zebua, 2024).

Teknik pengumpulan data utama menggunakan pretest dan posttest untuk mengetahui perbedaan kemampuan berpikir kritis antara kelompok eksperimen dan kelompok kontrol. Selain itu, observasi terstruktur dilakukan untuk mengamati keterlibatan siswa, pola interaksi dengan ChatGPT, serta peran guru selama proses pembelajaran. Dokumentasi berupa foto kegiatan, catatan observasi, dan arsip hasil interaksi siswa dengan ChatGPT dikumpulkan sebagai data pendukung.

Uji validitas instrumen dilakukan secara empiris menggunakan korelasi Pearson Product Moment untuk memastikan setiap butir soal valid dalam mengukur kemampuan berpikir kritis (Arikunto, 2020). Uji reliabilitas menggunakan koefisien Cronbach's Alpha dengan kriteria reliabilitas minimum  $\alpha \geq 0,70$ , sesuai standar pengukuran pendidikan modern (Heale & Twycross, 2019).

Analisis data dilakukan menggunakan independent sample t-test untuk menguji perbedaan rerata kemampuan berpikir kritis antara kelompok eksperimen dan kelompok kontrol. Sebelum pengujian hipotesis, dilakukan uji prasyarat berupa uji normalitas Shapiro-Wilk dan uji homogenitas varians menggunakan Levene's test (Ghozali, 2021). Apabila data tidak memenuhi asumsi parametrik, maka digunakan uji alternatif nonparametrik Mann-Whitney U. Pemilihan uji Mann-Whitney U dipertimbangkan karena ukuran sampel yang relatif kecil dan potensi distribusi data yang tidak normal. Analisis lanjutan berupa perhitungan effect size menggunakan Cohen's d dilakukan untuk mengetahui besaran pengaruh penggunaan ChatGPT

secara praktis (Plonsky & Oswald, 2023). Seluruh analisis data diolah menggunakan SPSS versi 26 dengan taraf signifikansi 0,05.

Penelitian ini juga memperhatikan aspek etika penelitian. Izin pelaksanaan penelitian diperoleh secara resmi dari pihak sekolah. Seluruh peserta penelitian diberikan penjelasan mengenai tujuan dan prosedur penelitian, serta dijamin kerahasiaan identitas dan data akademiknya. Penggunaan ChatGPT dibatasi hanya untuk kepentingan pembelajaran dan penelitian, serta diawasi langsung oleh guru untuk mencegah penyalahgunaan teknologi.

### 3. Hasil dan Pembahasan

#### 1. Gambaran umum instrumen: validitas dan reliabilitas

Sebelum menelaah hasil uji hipotesis, langkah awal adalah memeriksa kualitas instrumen penelitian. Berdasarkan keluaran uji validitas yang disajikan, setiap butir (Q1–Q25) dianalisis menggunakan korelasi Pearson terhadap skor total. Terdapat banyak korelasi antar-butir yang signifikan pada tingkat 0,05 dan 0,01, misalnya korelasi tinggi antara Q1–Q3 ( $r \approx 0,732$ ;  $p = 0,002$ ) dan beberapa pasangan butir lain yang menunjukkan hubungan kuat (mis. Q11–Q13  $r = 0,873$ ;  $p < 0,001$ ). Pola korelasi seperti ini mengindikasikan bahwa butir-butir instrumen saling berkaitan secara konsisten dengan konstruk yang diukur.

Selanjutnya, uji reliabilitas Cronbach's Alpha menunjukkan nilai  $\alpha = 0,924$  untuk keseluruhan 25 item. Menurut pedoman umum interpretasi reliabilitas,  $\alpha \geq 0,90$  dikategorikan sangat baik. Dengan demikian dapat disimpulkan bahwa instrumen memiliki konsistensi internal yang tinggi dan dapat dipercaya untuk mengukur konstruk kemampuan berpikir kritis atau variabel terukur lain yang diacu oleh butir-butir tersebut.

**Tabel 1. Uji Validitas Instrumen (Descriptive Statistics)**

| Item | Mean | Std. Deviation | N  |
|------|------|----------------|----|
| Q1   | 2,13 | 2,066          | 15 |
| Q2   | 2,13 | 2,066          | 15 |
| Q3   | 2,13 | 2,066          | 15 |
| Q4   | 2,13 | 2,066          | 15 |
| Q5   | 2,40 | 2,028          | 15 |
| Q6   | 1,60 | 2,028          | 15 |
| Q7   | 1,87 | 2,066          | 15 |
| Q8   | 1,87 | 2,066          | 15 |
| Q9   | 1,60 | 2,028          | 15 |
| Q10  | 1,60 | 2,028          | 15 |
| Q11  | 2,40 | 2,028          | 15 |
| Q12  | 2,13 | 2,066          | 15 |
| Q13  | 2,13 | 2,066          | 15 |
| Q14  | 2,13 | 2,066          | 15 |
| Q15  | 1,60 | 2,028          | 15 |
| Q16  | 1,87 | 2,066          | 15 |
| Q17  | 1,60 | 2,028          | 15 |
| Q18  | 1,33 | 1,952          | 15 |
| Q19  | 0,80 | 1,656          | 15 |
| Q20  | 1,33 | 1,952          | 15 |
| Q21  | 0,80 | 1,656          | 15 |
| Q22  | 1,60 | 2,028          | 15 |
| Q23  | 1,60 | 2,028          | 15 |
| Q24  | 1,60 | 2,028          | 15 |
| Q25  | 1,60 | 2,028          | 15 |

## 2. Uji normalitas

Uji normalitas dilakukan untuk tiap variabel skor (pretest dan posttest pada kelompok kontrol serta pretest dan posttest pada kelompok eksperimen) dengan Kolmogorov–Smirnov dan Shapiro Wilk. Hasil Shapiro Wilk secara ringkas:

- 1) PRETEST KONTROL:  $W = 0.786$ ,  $p = 0.002$  ( **$p < 0,05$ : data tidak normal**)
- 2) POSTTEST KONTROL:  $W = 0.819$ ,  $p = 0.006$  (**data tidak normal**)
- 3) PRETEST EKSPERIMENT:  $W = 0.791$ ,  $p = 0.003$  (**data tidak normal**)
- 4) POSTTEST EKSPERIMENT:  $W = 0.866$ ,  $p = 0.030$  (**data tidak normal**)

Karena seluruh variabel pretest dan posttest menunjukkan  $p < 0,05$  pada Shapiro–Wilk, asumsi normalitas tidak terpenuhi untuk data pada sampel ini. Ketidakterpenuhan normalitas ini menjadi dasar pemilihan uji nonparametrik pada analisis perbandingan antar-kelompok.

Catatan praktis: sampel kecil ( $n = 15$  per kelompok) membuat uji normalitas menjadi lebih sensitif; ketidakteraturan distribusi bisa disebabkan oleh outlier, skala pengukuran, atau distribusi alami respons peserta. Oleh karena itu, pendekatan nonparametrik yang tidak bergantung pada asumsi normalitas dianggap lebih aman.

## 3. Uji homogenitas varians

Uji Levene dilakukan untuk menelaah homogenitas varians antara skor posttest kelompok kontrol dan eksperimen. Pada keluaran, Levene Statistic (berdasarkan mean) = 42.881 dengan  $df_1 = 1$ ,  $df_2 = 28$  dan  $p = 0.000$ . Secara statistik,  $p = 0.000 (< 0,05)$  menunjukkan adanya perbedaan varians yang signifikan antara kedua kelompok; artinya asumsi homogenitas varians tidak terpenuhi.

Catatan: di beberapa laporan terlampir terdapat interpretasi yang menyatakan "homogen", namun berdasarkan aturan uji Levene konvensional ( $\alpha = 0,05$ ) nilai  $p = 0.000$  semestinya diinterpretasikan sebagai **heterogenitas varians**. Kombinasi hasil non-normal dan varians yang tidak homogen memperkuat pilihan penggunaan uji nonparametrik (Mann–Whitney U) untuk menguji perbedaan posttest antar-kelompok.

## 4. Uji hipotesis antar-kelompok (Mann Whitney U)

Karena asumsi normalitas dan homogenitas tidak terpenuhi, peneliti menggunakan uji Mann Whitney U untuk menguji hipotesis apakah ada perbedaan skor posttest antara kelompok eksperimen (yang menggunakan ChatGPT sebagai learning assistant) dan kelompok kontrol (pembelajaran konvensional).

Hasil uji Mann Whitney U:

- 1)  $N$  per kelompok = 15 (total  $N = 30$ )
- 2) Mean Rank: posttest kontrol = 11.20; posttest eksperimen = 19.80
- 3) Mann–Whitney  $U = 48.000$
- 4)  $Z = -2.688$
- 5) Asymp. Sig. (2-tailed) = 0.007

Interpretasi: nilai  $p$  (Asymp. Sig.) = 0.007  $< 0.05$  berarti terdapat perbedaan yang signifikan secara statistik antara skor posttest kelompok eksperimen dan kelompok kontrol. Dilihat dari mean rank, skor posttest pada kelompok eksperimen cenderung lebih tinggi daripada kelompok kontrol (rank rata-rata 19.80 vs 11.20), yang menunjukkan bahwa intervensi (pemanfaatan ChatGPT dalam rangkaian pembelajaran PAI) berasosiasi dengan peningkatan skor posttest dibandingkan pembelajaran konvensional.

## 5. Besaran efek (effect size)

Untuk mengukur besarnya pengaruh praktis, digunakan perhitungan effect size nonparametrik berdasarkan  $Z$ :  $r = Z / \sqrt{N}$ . Data menunjukkan  $Z = 2.688$  (nilai absolut),  $N = 30$ , sehingga  $r \approx 0.491$ . Mengacu pada pedoman umum (Cohen), nilai  $r \approx 0.49$  dapat dikategorikan sebagai **efek sedang** (medium). Artinya, perbedaan antara kelompok eksperimen dan kontrol

bukan hanya signifikan secara statistik, tetapi juga bermakna secara praktis cukup besar untuk dipertimbangkan sebagai efek pedagogis yang relevan.

Dalam konteks pembelajaran Pendidikan Agama Islam, nilai  $r \approx 0,49$  menunjukkan bahwa penggunaan ChatGPT memberikan dampak pedagogis yang bermakna, bukan sekadar perbedaan statistik. Efek sedang ini mengindikasikan bahwa AI berpotensi membantu siswa mengembangkan keterampilan berpikir kritis—seperti menganalisis konsep, mengevaluasi argumen, dan merumuskan inferensi—ketika digunakan secara terarah dan berada di bawah bimbingan guru.

Secara ringkas, temuan kuantitatif kunci adalah sebagai berikut: instrumen penelitian valid dan sangat reliabel ( $\alpha = 0,924$ ); distribusi skor pretest/posttest tidak memenuhi normalitas; varians posttest antar-kelompok berbeda signifikan; uji Mann–Whitney menunjukkan perbedaan posttest yang signifikan ( $p = 0,007$ ) dengan peningkatan pada kelompok eksperimen; besaran efek  $r \approx 0,49$  menunjukkan efek sedang. Berdasarkan bukti ini, hipotesis bahwa penggunaan ChatGPT berpengaruh terhadap peningkatan kemampuan (skor posttest) mendapat dukungan empiris pada sampel yang diuji.

Hasil yang menunjukkan perbedaan posttest signifikan dan efek sedang mengindikasikan bahwa integrasi ChatGPT dalam skenario pembelajaran PAI mampu memberi kontribusi pada peningkatan prestasi kognitif siswa dalam hal ini diukur melalui instrumen yang reliabel. Secara konseptual, ChatGPT berfungsi sebagai sumber penjelasan cepat, pemberi umpan balik, dan mitra diskusi yang memungkinkan siswa berlatih mengajukan pertanyaan, menguji argumen, dan merumuskan penalaran secara lebih berulang dibandingkan sesi pembelajaran tradisional satu arah. Interaksi semacam ini dapat meningkatkan keterampilan berpikir tingkat tinggi bila didesain secara terarah (mis. menyediakan prompt berkualitas, mendorong verifikasi, dan memfasilitasi refleksi guru sebagai pembimbing).

Temuan ini koheren dengan literatur yang menunjukkan potensi AI generatif meningkatkan diskusi interaktif dan memperkaya perspektif peserta didik ketika digunakan secara pedagogis (lihat kajian empiris terkait AI-assisted learning). Namun, penting dicatat bahwa AI bukanlah substitusi guru; peran guru sebagai pembingkai nilai, penilai validitas argumen, dan fasilitator etika tetap krusial.

Data yang tidak normal dan varians yang tidak homogen menunjukkan karakter distribusi skor pada sampel ini tidak memenuhi prasyarat uji parametrik. Hal ini dapat disebabkan oleh beberapa faktor: ukuran sampel relatif kecil, distribusi respons yang tidak merata, ataupun adanya outlier. Penggunaan Mann Whitney U menjaga validitas inferensial ketika asumsi parameter tidak terpenuhi. Hasil signifikan pada uji nonparametrik memperkuat keyakinan bahwa perbedaan yang ditemukan bukan artefak asumsi statistik melainkan fenomena yang benar-benar muncul dalam data.

Namun, keterbatasan penggunaan uji nonparametrik adalah kurangnya estimasi selang kepercayaan dan parameter distribusi yang lebih kaya. Oleh karenanya, penelitian selanjutnya dengan sampel lebih besar dan desain acak penuh akan memungkinkan analisis parametrik yang lebih kuat dan estimasi efek yang lebih presisi.

Beberapa mekanisme pedagogis yang dapat menjelaskan kenapa kelompok eksperimen unggul antara lain: (1) ChatGPT menyediakan umpan balik cepat sehingga siswa dapat segera merevisi pemikiran; (2) AI mampu menghadirkan berbagai sudut pandang dan contoh aplikatif yang memicu berpikir kritis; (3) struktur sesi yang memadukan paparan guru, interaksi AI, dan refleksi kelas mendorong praktik berpikir aktif (scaffolding); dan (4) penggunaan prompt yang diarahkan memaksa siswa menyusun pertanyaan berkualitas dan menilai respons — keterampilan inti berpikir kritis. Kombinasi unsur tersebut lebih berpotensi mengembangkan kemampuan analisis, evaluasi, dan sintesis dibandingkan metode ceramah/hafalan pasif.

Walaupun hasil menunjukkan keuntungan kognitif, konteks PAI memerlukan perhatian khusus terhadap aspek kebenaran teologis, etika digital, dan pembentukan spiritualitas. ChatGPT bekerja berbasis model bahasa probabilistik bukan sumber otoritatif wahyu sehingga penggunaan tanpa supervisi berisiko menghasilkan interpretasi yang keliru. Oleh karena itu,

integrasi AI dalam PAI harus disertai pedoman verifikasi keagamaan, peran aktif guru untuk mengoreksi dan menyelaraskan jawaban dengan prinsip syar'i, serta pendidikan literasi agama digital bagi siswa agar mampu melakukan tahqiq (verifikasi) terhadap output AI.

Penelitian ini memiliki beberapa keterbatasan yang perlu dicatat: (1) ukuran sampel relatif kecil ( $n = 15$  per kelompok), sehingga generalisasi terbatas; (2) desain quasi-experimental tanpa randomisasi penuh berisiko hadirnya bias seleksi; (3) durasi intervensi singkat (enam minggu) sehingga temuan merefleksikan efek jangka pendek; (4) pengukuran berbasis tes posttest saja belum cukup menjelaskan aspek afektif/spiritual yang menjadi tujuan utama PAI; dan (5) data tidak normal dan heterogenitas varians menuntut kehati-hatian dalam menafsirkan besaran efek secara statistik. Kombinasi keterbatasan ini menuntut replikasi dengan desain eksperimental acak, sampel lebih besar, pengukuran longitudinal, dan instrumen tambahan untuk aspek afektif dan spiritual.

Berdasarkan temuan, beberapa rekomendasi praktis dapat diajukan: (a) integrasi ChatGPT secara bertahap dalam RPP PAI dengan skenario pembelajaran yang jelas (guide prompts, rubrik evaluasi, sesi refleksi guru); (b) pelatihan literasi AI dan etika bagi guru dan siswa; (c) penyusunan pedoman verifikasi teologis yang menegaskan bahwa output AI harus diverifikasi oleh rujukan syar'i dan oleh guru; (d) pengembangan model pembelajaran hibrid yang menggabungkan kekuatan AI (akses cepat, variasi perspektif) dengan pendekatan ta'dib (pengasuhan moral); (e) penelitian lanjutan menggunakan sampel lebih besar, randomisasi, dan pengukuran dampak jangka panjang termasuk aspek spiritual dan implementasi etis.

Dari analisis statistik yang tersedia: instrumen valid dan reliabel; data pretest/posttest tidak normal; varians antar-posttest tidak homogen; uji Mann-Whitney menunjukkan perbedaan posttest signifikan ( $p = 0.007$ ) dengan efek sedang ( $r \approx 0.49$ ). Temuan ini mendukung hipotesis bahwa penggunaan ChatGPT dalam pembelajaran PAI memberi kontribusi terhadap peningkatan skor posttest bila dibandingkan pembelajaran konvensional pada sampel yang diteliti. Namun, implikasi praktis harus diimbangi dengan pengendalian kualitas konten keagamaan, bimbingan guru, dan pertimbangan etis.

Temuan empiris ini selaras dengan teori berpikir kritis yang menekankan pentingnya aktivitas kognitif aktif, dialog reflektif, dan scaffolding (Facione; Ennis). ChatGPT dalam penelitian ini berfungsi sebagai mitra diskusi digital yang menyediakan umpan balik cepat, memunculkan variasi sudut pandang, serta mendorong siswa untuk mengajukan pertanyaan bermakna. Mekanisme ini sejalan dengan literatur AI-assisted learning yang menyatakan bahwa AI dapat meningkatkan kualitas interaksi belajar apabila diposisikan sebagai alat bantu, bukan sumber otoritas utama.

Namun demikian, hasil ini perlu dibaca secara kritis. Penggunaan ChatGPT berpotensi menghadirkan bias informasi, ketergantungan berlebihan, serta penyederhanaan proses berpikir apabila tidak disertai pengawasan pedagogis. Dalam konteks PAI, tantangan epistemologis dan teologis menjadi lebih kompleks karena AI tidak berlandaskan otoritas wahyu. Oleh karena itu, peran guru tetap krusial sebagai penjaga validitas keilmuan, pembingkai nilai, dan fasilitator refleksi spiritual.

Penelitian ini memiliki beberapa keterbatasan, antara lain ukuran sampel yang kecil, desain quasi-experimental tanpa randomisasi penuh, durasi intervensi yang relatif singkat, serta pengukuran yang berfokus pada aspek kognitif. Keterbatasan ini membatasi generalisasi temuan dan menuntut kehati-hatian dalam penafsiran efek.

Implikasi pedagogis dari penelitian ini menegaskan bahwa integrasi ChatGPT dalam pembelajaran PAI perlu dirancang secara sistematis, etis, dan reflektif. ChatGPT sebaiknya digunakan untuk memicu dialog, analisis, dan evaluasi, sementara guru tetap memegang peran sentral dalam membimbing, memverifikasi, dan menanamkan nilai-nilai keislaman.

#### 4. Penutup

Penelitian ini menunjukkan bahwa integrasi ChatGPT dalam pembelajaran Pendidikan Agama Islam memberikan dampak signifikan terhadap peningkatan kemampuan berpikir kritis siswa SMA. Berdasarkan hasil uji statistik nonparametrik Mann–Whitney U yang diperoleh dari skor posttest, terdapat perbedaan bermakna antara kelompok eksperimen dan kelompok kontrol ( $p = 0,007$ ). Kelompok yang menggunakan ChatGPT menunjukkan pencapaian lebih tinggi dengan besaran efek  $r \approx 0,49$ , yang termasuk kategori sedang dan bermakna secara pedagogis. Secara metodologis, temuan ini diperkuat oleh kualitas instrumen penelitian yang sangat baik, dengan reliabilitas Cronbach's Alpha 0,924 serta korelasi antar-butir yang signifikan. Hal ini memastikan bahwa peningkatan skor bukan akibat kesalahan pengukuran, tetapi berkaitan langsung dengan intervensi pembelajaran berbasis kecerdasan buatan. Namun, distribusi data yang tidak normal dan varians yang tidak homogen mengharuskan interpretasi yang cermat dan menunjukkan perlunya penelitian lanjutan dengan sampel lebih besar dan desain acak penuh. Hasil penelitian menegaskan bahwa ChatGPT dapat berperan sebagai mitra belajar yang membantu siswa dalam mengembangkan kemampuan analisis, evaluasi, dan sintesis melalui interaksi berbasis bahasa alami yang bersifat responsif. Kombinasi antara paparan materi oleh guru, dialog mandiri dengan AI, dan refleksi kelas menjadi pola pembelajaran yang konstruktif untuk menstimulasi keterampilan berpikir tingkat tinggi.

Kendati demikian, penelitian ini juga menggarisbawahi risiko epistemologis, pedagogis, dan etis dalam penggunaan AI pada ruang pembelajaran agama. ChatGPT bukan otoritas keagamaan dan beroperasi berdasarkan probabilitas bahasa, sehingga potensi bias dan kesalahan tafsir perlu diantisipasi melalui pengawasan guru. Oleh sebab itu, penggunaan AI dalam pembelajaran PAI harus disertai literasi digital keagamaan, panduan verifikasi sumber, serta penguatan nilai-nilai adab dan tanggung jawab akademik. Secara keseluruhan, penelitian ini menyimpulkan bahwa pemanfaatan ChatGPT secara terarah dan etis mampu memberikan kontribusi nyata dalam meningkatkan kemampuan berpikir kritis siswa pada pembelajaran PAI. Namun, keberhasilan implementasi tetap bergantung pada integrasi harmonis antara teknologi, nilai-nilai Islam, dan peran guru sebagai pendidik yang membimbing proses internalisasi ilmu secara bermakna.

#### Ucapan Terima Kasih

Penulis menyampaikan penghargaan yang tulus kepada semua pihak yang telah memberikan dukungan dalam penyusunan penelitian berjudul Digital Transformation in Islamic Religious Education: The Role of ChatGPT in Developing Critical Thinking Skills Among Secondary School Students. Terima kasih kepada Universitas Muhammadiyah Bengkulu atas fasilitas dan bimbingan akademik yang memungkinkan penelitian ini terlaksana dengan baik. Apresiasi juga diberikan kepada para pendidik dan siswa Sekolah Menengah yang telah berpartisipasi dan bekerja sama selama proses pengumpulan data. Penghargaan khusus ditujukan kepada para dosen pembimbing yang telah memberikan arahan, masukan konstruktif, serta motivasi sepanjang penyusunan penelitian ini. Penulis juga berterima kasih kepada rekan-rekan sejawat yang turut membantu melalui diskusi ilmiah dan dukungan moral. Semoga kontribusi seluruh pihak menjadi amal kebaikan dan penelitian ini memberikan manfaat bagi pengembangan pembelajaran Pendidikan Agama Islam di era digital.

#### Daftar Pustaka

- Abbas, A., Latif, F., & Khan, M. (2024). AI integration in religious education: Opportunities and challenges. *Journal of Digital Pedagogy*, 12(1), 45–59.
- Acosta-Enriquez, J., Torres, R., & Molina, S. (2024). Critical thinking development through AI-assisted learning: Evidence from secondary education. *International Journal of Educational Technology*, 18(2), 77–95.

- Arikunto, S. (2020). *Prosedur penelitian: Suatu pendekatan praktik*. Rineka Cipta.
- Archila, P., Chaves, A., & Molina, J. (2024). Dialogic learning with AI tools in secondary education: Pedagogical implications. *Research and Practice in Technology-Enhanced Learning*, 19(3), 122–139.
- Bilad, M. R., Rahman, S., & Hidayat, T. (2024). Creative thinking in PISA 2022: Implications for Indonesian education. *Journal of Educational Measurement*, 15(1), 1–13.
- Chanda, S. (2025). Ethics of artificial intelligence in education: Challenges for moral development and responsibility. *International Review of Ethics in Education*, 8(1), 66–82.
- Creswell, J. W., & Creswell, J. D. (2023). *Research design: Qualitative, quantitative, and mixed methods approaches* (6th ed.). SAGE Publications.
- Facione, P. A. (2015). *Critical thinking: What it is and why it counts*. Insight Assessment.
- Farrokhnia, M., Noroozi, O., & Hatami, J. (2024). Generative AI as a learning partner: A systematic review of educational applications. *Computers & Education*, 210, 104275.
- Ghozali, I. (2021). *Aplikasi analisis multivariante dengan program IBM SPSS*. Badan Penerbit Universitas Diponegoro.
- Heale, R., & Twycross, A. (2019). Validity and reliability in quantitative studies. *Evidence-Based Nursing*, 22(3), 68–70.
- Hermila, S., Yusof, A., & Rahman, N. (2024). Students' perceptions of ChatGPT in classroom interaction. *Asia Pacific Journal of Education*, 44(2), 188–204.
- Kasman, K., Arifin, M., & Yusuf, M. (2022). Problematika pembelajaran Pendidikan Agama Islam di SMA. *Jurnal Pendidikan Islam*, 7(1), 25–40.
- Kurata, K., Matsuo, M., & Suzuki, A. (2025). Artificial intelligence and critical thinking: Risks of cognitive overdependence. *Journal of the Learning Sciences*, 33(1), 55–78.
- Leelavathi, R., & Surendhranatha, S. (2024). Pedagogical potentials of ChatGPT in higher education. *Education and Information Technologies*, 29(4), 4457–4475.
- Lin, T.-J., Chen, S., & Huang, C. (2024). Observational frameworks for digital classroom interaction. *Journal of Interactive Learning Research*, 20(2), 133–150.
- Luo, Y. (2024). Natural language processing for personalized learning assessment. *Journal of Artificial Intelligence in Education*, 32(1), 78–101.
- Nofamatara Zebua, N. (2024). Assessment framework for higher order thinking skills in secondary education. *Jurnal Evaluasi Pendidikan*, 15(2), 55–70.
- Ocampo, J., & Gozum, J. (2024). Epistemic reliability of artificial intelligence in religious contexts. *Journal of Philosophy and Theology in Education*, 18(1), 21–39.
- OECD. (2023). *Artificial intelligence in education: Challenges and opportunities*. OECD Publishing.
- OECD. (2024). *PISA 2022 creative thinking framework*. OECD Publishing.
- Plonsky, L., & Oswald, F. L. (2023). Effect sizes for nonparametric tests in educational research. *Educational Research Review*, 39, 1–15.
- Rahmawati, N. (2022). Kesiapan sekolah dalam transformasi digital pendidikan. *Jurnal Teknologi Pendidikan*, 24(3), 211–223.
- Siyoto, S., & Sodik, A. (2023). *Dasar metodologi penelitian*. Literasi Media.
- Sugiyono. (2021). *Metode penelitian kuantitatif, kualitatif, dan R&D*. Alfabeta.
- Sujarwo, S. (2022). Dampak teknologi terhadap karakter belajar siswa. *Jurnal Pendidikan Nasional*, 10(2), 99–112.
- Williams, L. (2024). Academic integrity in the age of artificial intelligence. *Journal of Academic Ethics*, 22(3), 215–231.