CHATBOT AI SEBAGAI INSTRUMEN EVALUASI AUTENTIK MAHARAH KITABAH DAN KALAM DALAM PEMBELAJARAN BAHASA ARAB SEBUAH TELAAH LITERATUR

Dwi Noviani[1], Ica Safitri[2], Erma Sapitri[3], Dini Kaspa[4].
Program Studi Pendidikan Bahasa Arab,
Universitas Al-Qu'an Ittifaqiayah, Indralaya, Indonesia[1,2,3,4]
dwi.noviani@uqi.ac.id , icas3611@gmail.com , ermasapitri6@gmail.com,
dinikaspa2@gmail.com

ABSTRACT

Artificial Intelligence (AI)-based chatbots have emerged as innovative tools that can support the evaluation process in Arabic language learning. Arabic language teachers at the secondary school level often face practical constraints, including limited instructional time, large class sizes, and diverse student needs, making it challenging to provide adaptive and personalized feedback. This study aims to examine the role of AI chatbots as authentic assessment instruments for maharah kitabah (writing skills) and maharah kalam (speaking skills). The research employed a narrative literature review method by analyzing 15 nationally accredited scientific articles published between 2016 and 2025, sourced from Google Scholar, Garuda, and SINTA databases. The findings indicate that ChatGPT has the potential to support writing assessment through automated text-based feedback, while Mondly Arabic facilitates speaking assessment through real-time speech recognition technology. The review also discusses issues related to validity, reliability, ethics, and the role of teachers as validators of AI-generated assessment outcomes. The novelty of this study lies in mapping the role of AI chatbots as authentic assessment instruments that integrate writing evaluation, speaking evaluation, self-reflection, and teacher validation into a single conceptual model for secondary-level Arabic language learning. The study concludes that AI is not intended to replace teachers but rather to reduce their workload while expanding students' learning opportunities through a structured and human-centered validation process.

Keywords: *Artificial Intelligence, Chatbot, Authentic Assessment, Arabic Language Learning, Literature Review.*

ABSTRAK

Guru bahasa Arab di tingkat sekolah menengah kerap dihadapkan pada keterbatasan nyata: waktu yang sempit, kelas yang padat, dan kebutuhan setiap siswa yang berbeda-beda membuat pemberian umpan balik yang adaptif dan personal menjadi tantangan tersendiri. Penelitian ini bertujuan mengkaji peran chatbot *Artificial Intelligence* (AI) sebagai instrumen evaluasi autentik untuk *maharah kitabah* dan *maharah kalam*. Metode yang digunakan adalah *narrative*

literature review terhadap 15 artikel ilmiah bereputasi nasional terbitan 2016–2025 dari basis data Google Scholar, Garuda, dan SINTA. Hasil kajian menunjukkan bahwa ChatGPT berpotensi mendukung evaluasi keterampilan menulis melalui umpan balik otomatis berbasis teks, sedangkan Mondly Arabic mendukung evaluasi keterampilan berbicara melalui pengenalan suara secara *real-time*. Kajian ini juga mendiskusikan isu validitas, reliabilitas, etika, dan peran guru sebagai validator hasil evaluasi AI. Kebaruan kajian ini terletak pada pemetaan peran chatbot AI sebagai instrumen evaluasi autentik yang mengintegrasikan evaluasi menulis, berbicara, refleksi mandiri, dan validasi guru dalam satu model konseptual pembelajaran bahasa Arab tingkat menengah. Kajian ini menegaskan bahwa kehadiran AI tidak dirancang untuk menggantikan guru, melainkan untuk meringankan beban guru sekaligus memperluas ruang belajar siswa melalui mekanisme validasi yang terstruktur dan berpusat pada kemanusiaan.

Kata Kunci: *Artificial Intelligence, Chatbot, Evaluasi Autentik, Bahasa Arab, Telaah Literatur.*

A. Pendahuluan

Sistem evaluasi merupakan komponen vital dalam dunia pendidikan yang berfungsi untuk mengukur tingkat pemahaman siswa dan efektivitas metode pengajaran yang diterapkan oleh guru. Di era transisi digital saat ini, sistem evaluasi pendidikan di Indonesia menghadapi tantangan baru seiring dengan pesatnya perkembangan teknologi *Artificial Intelligence* (AI). Kehadiran AI, khususnya teknologi *generative* berbasis teks dan suara, memicu perdebatan mengenai dampak dan pergeseran paradigma dalam mengukur kemampuan akademik siswa. Dalam konteks pembelajaran bahasa asing, khususnya bahasa Arab di tingkat sekolah menengah (SMP/SMA), evaluasi konvensional berbasis kertas seringkali kurang adaptif dalam mengukur kompetensi kebahasaan siswa secara *real-time* dan interaktif.

Pembelajaran bahasa Arab di sekolah menengah masih sering didominasi oleh pendekatan yang berorientasi pada penguasaan kaidah kebahasaan dan aspek teoretis dibandingkan praktik berbahasa secara komunikatif (Bahruddin et al., 2021; Nasaruddin, 2024). Ketika proses evaluasi dilakukan guru sering kali merasa "kewalahan" ketika harus memberikan umpan balik yang cepat dan bersifat personal kepada seluruh peserta didik, terutama pada keterampilan menulis dan berbicara yang menuntut perhatian mendalam

(Kasneci et al., 2023; Kohnke et al., 2023). Akibatnya, motivasi belajar siswa menurun dan evaluasi tidak berjalan secara autentik. Kehadiran platform berbasis AI seperti ChatGPT membuka peluang baru dalam proses evaluasi bahasa karena mampu memberikan umpan balik secara otomatis, sedangkan aplikasi pembelajaran bahasa berbasis pengenalan suara dapat mendukung latihan dan evaluasi keterampilan berbicara secara lebih interaktif (Kasneci et al., 2023; Kohnke et al., 2023; UNESCO, 2023).

Meskipun potensi pemanfaatan AI sebagai instrumen evaluasi mandiri sangat menjanjikan, penerapannya di lapangan tidak luput dari kendala yang kompleks. Beberapa penelitian terdahulu menunjukkan bahwa Implementasi AI dalam pendidikan sangat dipengaruhi oleh kesiapan infrastruktur digital serta kompetensi guru dalam memanfaatkan teknologi secara pedagogis (Holmes et al., 2022; UNESCO, 2023). Selain itu, masih terdapat sejumlah perhatian terkait validitas hasil penilaian, potensi bias algoritma, transparansi sistem, serta perlindungan data peserta didik dalam penggunaan AI untuk pendidikan (UNESCO, 2023; Kasneci et al., 2023; Holmes et al., 2022). Ketergantungan berlebih siswa terhadap AI juga menjadi tantangan pedagogis yang perlu diantisipasi.

Sebagian besar penelitian terdahulu masih menempatkan AI sebagai media pembelajaran atau alat bantu latihan bahasa, bukan sebagai instrumen evaluasi yang terstruktur. Kajian yang secara khusus membahas AI sebagai instrumen evaluasi autentik untuk mengukur *maharah kitabah* dan *maharah kalam* dalam pembelajaran bahasa Arab masih terbatas. Selain itu, belum banyak kajian yang menghubungkan penggunaan AI dengan isu validitas, reliabilitas, etika, dan peran guru sebagai validator hasil evaluasi. Kesenjangan inilah yang menjadi alasan utama penelitian ini ditulis: untuk mengisi ruang kajian yang belum banyak dijamah tersebut sekaligus menawarkan model konseptual evaluasi autentik berbasis AI yang relevan dengan konteks pendidikan bahasa Arab di tingkat menengah.

Berdasarkan uraian latar belakang tersebut, penelitian ini bertujuan untuk menganalisis pemanfaatan *Artificial Intelligence* (AI) sebagai instrumen evaluasi autentik

dalam pembelajaran bahasa Arab pada jenjang sekolah menengah. Secara khusus, penelitian ini mengkaji peran AI dalam mendukung evaluasi *maharah kitabah* dan *maharah kalam*, mengidentifikasi peluang dan tantangan implementasinya dalam konteks pendidikan, mendiskusikan isu validitas, reliabilitas, dan etika penggunaan AI, serta menawarkan model konseptual evaluasi autentik berbasis AI yang dapat dijadikan alternatif dalam pengembangan sistem evaluasi pembelajaran bahasa Arab yang lebih efektif, adaptif, dan sesuai dengan tuntutan pembelajaran abad ke-21.

B. Metode Penelitian

Pengumpulan data dalam penelitian ini dilakukan secara sekunder melalui penelusuran literatur pada berbagai basis data ilmiah, yaitu Google Scholar, Directory of Open Access Journals (DOAJ), Garuda (Garba Rujukan Digital), dan SINTA (Science and Technology Index). Penelusuran dilakukan menggunakan kombinasi kata kunci yang relevan dengan fokus penelitian, antara lain "Artificial Intelligence bahasa Arab", "Chatbot pembelajaran bahasa Arab", "Evaluasi pembelajaran AI", serta "Mondly Arabic atau ChatGPT dalam pendidikan". Artikel yang ditelusuri dibatasi pada publikasi tahun 2019–2025 untuk memastikan keterkaitan dengan perkembangan teknologi Artificial Intelligence (AI) terkini. Namun demikian, beberapa literatur yang diterbitkan sebelum periode tersebut tetap digunakan sebagai landasan teoritis karena memiliki relevansi yang kuat terhadap konsep AI dalam pendidikan dan evaluasi pembelajaran.

Kriteria inklusi yang digunakan dalam penelitian ini meliputi artikel ilmiah yang membahas Artificial Intelligence, chatbot pendidikan, evaluasi pembelajaran, asesmen bahasa, serta pembelajaran bahasa Arab. Seluruh sumber yang digunakan berasal dari jurnal ilmiah yang telah melalui proses peer review dan tersedia dalam bahasa Indonesia maupun bahasa Inggris. Sebaliknya, artikel yang tidak memiliki keterkaitan langsung dengan fokus penelitian atau tidak memberikan informasi yang mendukung tujuan kajian dikeluarkan dari proses analisis.

Berdasarkan hasil penelusuran awal, diperoleh sebanyak 43 artikel yang berpotensi relevan dengan topik penelitian. Selanjutnya dilakukan

proses seleksi awal berdasarkan kesesuaian tema dan fokus kajian, sehingga delapan artikel dikeluarkan dan tersisa 35 artikel yang memenuhi kriteria relevansi. Pada tahap berikutnya, sebanyak 20 artikel dipilih untuk ditelaah secara lebih mendalam melalui pembacaan menyeluruh terhadap isi artikel. Setelah proses evaluasi literatur dilakukan, lima artikel kembali dieliminasi karena tidak memberikan kontribusi yang signifikan terhadap fokus penelitian. Dengan demikian, sebanyak 15 artikel ditetapkan sebagai sumber utama yang dianalisis dalam kajian ini karena memiliki keterkaitan yang paling kuat dengan pemanfaatan Artificial Intelligence sebagai instrumen evaluasi autentik dalam pembelajaran bahasa Arab.

Data yang diperoleh kemudian dianalisis menggunakan pendekatan analisis tematik. Pada tahap awal, peneliti mengidentifikasi berbagai tema utama yang muncul dari literatur terpilih. Tema-tema tersebut kemudian dikelompokkan berdasarkan fokus kajian, meliputi evaluasi maharah kitabah, evaluasi maharah kalam, pemberian umpan balik otomatis, validitas dan reliabilitas penilaian, serta aspek etika dalam penggunaan Artificial Intelligence di bidang pendidikan. Selanjutnya, berbagai temuan tersebut disintesis untuk memperoleh pemahaman yang lebih komprehensif mengenai pemanfaatan chatbot AI dalam proses evaluasi pembelajaran bahasa Arab. Hasil sintesis tersebut digunakan untuk merumuskan model konseptual chatbot AI sebagai instrumen evaluasi autentik yang mengintegrasikan penilaian keterampilan menulis, keterampilan berbicara, refleksi diri peserta didik, serta validasi guru dalam satu kerangka evaluasi yang terpadu.

Berdasarkan proses penelusuran, seleksi, dan analisis literatur yang telah dilakukan, diperoleh lima belas artikel yang dinilai paling relevan dengan fokus penelitian. Seluruh artikel tersebut kemudian dianalisis secara mendalam untuk mengidentifikasi konsep utama, temuan empiris, serta kontribusinya terhadap pengembangan model konseptual chatbot Artificial Intelligence sebagai instrumen evaluasi autentik dalam pembelajaran bahasa Arab pada jenjang sekolah menengah.

Tabel 1. Sintesis Literatur Penelitian Terdahulu

No	Penulis dan Tahun	Fokus Kajian	Temuan Utama	Kontribusi terhadap Penelitian
1	Luckin et al. (2016)	Artificial Intelligence dalam pendidikan	AI berpotensi mendukung personalisasi pembelajaran dan evaluasi peserta didik.	Menjadi landasan teoritis penggunaan AI dalam proses evaluasi pendidikan.
2	Holmes, Bialik, & Fadel (2022)	Artificial Intelligence in Education (AIED)	AI dapat meningkatkan efektivitas pembelajaran, asesmen, dan pengambilan keputusan pendidikan.	Menjelaskan peluang dan tantangan integrasi AI dalam pendidikan.
3	UNESCO (2023)	Pedoman penggunaan Generative AI dalam pendidikan	Penggunaan AI harus memperhatikan etika, privasi data, transparansi, dan peran guru.	Menjadi dasar pembahasan aspek etika dan tata kelola AI.
4	Kasneci et al. (2023)	Pemanfaatan ChatGPT dalam pendidikan	ChatGPT mampu membantu pemberian umpan balik, penilaian, dan pembelajaran adaptif, namun tetap memerlukan pengawasan manusia.	Menjadi dasar analisis penggunaan ChatGPT dalam evaluasi autentik.
5	Kohnke, Moorhouse, & Zou (2023)	Generative AI dan pembelajaran bahasa	AI generatif dapat mendukung pengembangan keterampilan bahasa dan	Menjadi dasar integrasi AI dalam pembelajaran dan evaluasi bahasa.
6	Bahruddin et al. (2021)	Konstruktivisme dalam pembelajaran bahasa Arab	Pembelajaran bahasa Arab perlu berpusat pada aktivitas peserta didik dan pengalaman autentik.	Menjadi landasan pedagogis evaluasi autentik bahasa Arab.
7	Nasaruddin (2024)	Artificial Intelligence dalam pembelajaran bahasa Arab	AI dapat dimanfaatkan untuk mendukung pembelajaran dan evaluasi keterampilan berbahasa Arab.	Menjadi referensi utama terkait AI dalam konteks pendidikan bahasa Arab.
8	Rohim et al. (2024)	Digitalisasi pembelajaran bahasa Arab	Teknologi digital membantu meningkatkan efektivitas proses pembelajaran bahasa Arab.	Mendukung integrasi teknologi dalam evaluasi pembelajaran bahasa Arab.
9	Rohmawaty et al. (2024)	Inovasi pembelajaran bahasa Arab berbasis teknologi	Pemanfaatan teknologi memberikan peluang pembelajaran yang lebih interaktif dan adaptif.	Menjelaskan relevansi penggunaan AI dalam pembelajaran bahasa Arab modern.
10	Muthohharoh et al. (2024)	Media dan teknologi pembelajaran bahasa Arab	Teknologi dapat meningkatkan keterlibatan peserta didik dalam proses belajar.	Mendukung penggunaan AI sebagai sarana evaluasi yang lebih menarik.
11	Farhan et al. (2024)	Transformasi digital dalam pendidikan bahasa Arab	Teknologi digital berkontribusi pada peningkatan kualitas pembelajaran dan evaluasi.	Menjadi dasar pengembangan evaluasi berbasis teknologi.
12	Agustina et al. (2024)	Tren penelitian AI dalam pendidikan	AI semakin banyak digunakan untuk mendukung asesmen dan analisis pembelajaran.	Memperlihatkan perkembangan penelitian AI dalam bidang pendidikan.
13	Supriyanto & Toifah (2024)	Implementasi teknologi dalam evaluasi pembelajaran	Teknologi membantu proses evaluasi menjadi lebih efisien dan terdokumentasi.	Mendukung argumentasi mengenai efektivitas sistem evaluasi digital.
14	Maulana et al. (2024)	Inovasi pembelajaran bahasa berbasis teknologi	Teknologi mendukung pengembangan keterampilan berbahasa secara lebih fleksibel.	Menjadi dasar pembahasan evaluasi maharah kalam dan maharah kitabah.
15	Kholis & Mustofa (2024)	Teknologi dan pembelajaran bahasa Arab	Integrasi teknologi memberikan peluang peningkatan kualitas pembelajaran bahasa Arab.	Mendukung pengembangan model evaluasi autentik berbasis AI.

C.Hasil Penelitian dan Pembahasan

Berdasarkan hasil penelusuran literatur terhadap 15 jurnal ilmiah bereputasi, ditemukan bahwa integrasi *Artificial Intelligence* (AI) membawa potensi perubahan yang signifikan terhadap sistem evaluasi pendidikan bahasa Arab. Evaluasi yang semula berjalan satu arah dan terasa kaku berpotensi berkembang menjadi proses yang lebih hidup—interaktif, adaptif, dan benar-benar berpusat pada kebutuhan peserta didik—melalui pemanfaatan Artificial Intelligence yang bijak. Teknologi AI memungkinkan peserta didik memperoleh umpan balik secara lebih cepat sehingga mendukung proses refleksi dan perbaikan belajar secara berkelanjutan (Luckin et al., 2016; Kasneci et al., 2023; Kohnke et al., 2023).Namun demikian, perlu ditegaskan bahwa temuan dalam kajian ini merupakan potensi berdasarkan literatur yang dikaji, bukan kesimpulan empiris langsung, mengingat penelitian ini berbasis kajian pustaka dan belum didukung oleh data lapangan dari penelitian sendiri.

Analisis Komparasi Fitur AI sebagai Instrumen Evaluasi Bahasa Arab

Dari hasil sintesis literatur, dua platform AI yang paling banyak dikaji dalam konteks evaluasi pembelajaran bahasa Arab adalah ChatGPT dan Mondly Arabic. Tabel 2 berikut menyajikan perbandingan fitur keduanya, termasuk aspek keterbatasan yang perlu diperhatikan.

Tabel 2. Komparasi Fitur ChatGPT dan Mondly Arabic sebagai Instrumen Evaluasi Bahasa Arab

Parameter Evaluasi	ChatGPT (AI Berbasis Teks) (Kasneci et al., 2023; Nasaruddin, 2024)	Mondly Arabic (AI Berbasis Interaktif) (Sakaroni & Azanulhaq, 2024)
Aspek Bahasa yang Diuji	*Maharah Kitabah* (Menulis) dan Pemahaman *Qawaid* (Tata Bahasa)	*Maharah Kalam* (Berbicara) dan *Maharah Istima'* (Mendengar)
Mekanisme Umpan Balik (*Feedback*)	Memberikan koreksi teks yang mendalam, penjelasan aturan nahu/sharaf, serta alternatif redaksi kalimat yang lebih luwes	Memberikan penilaian pelafalan (*speech recognition*) instan, skor interaktif, dan konfirmasi kebenaran aksen
Bentuk Struktur Evaluasi	Fleksibel; dapat diinstruksikan untuk membuat simulasi tes tulis, kuis pilihan ganda, atau permainan kosakata	Terstruktur; menggunakan percakapan berbasis skenario harian (kontekstual)
Keterbatasan Utama	Rentan terhadap halusinasi faktual, belum mampu menilai intonasi dan pelafalan, serta memerlukan verifikasi guru	Keterbatasan dalam menilai variasi dialek Arab; beberapa fitur bersifat berbayar dan bergantung pada koneksi internet

Berdasarkan berbagai penelitian yang telah dikaji, ChatGPT dan Mondly Arabic memiliki fungsi yang berbeda dalam mendukung evaluasi pembelajaran bahasa Arab. ChatGPT lebih banyak dimanfaatkan untuk membantu penilaian keterampilan menulis karena mampu memberikan masukan terkait tata bahasa, penggunaan kosakata, dan penyusunan kalimat. Di sisi lain, Mondly Arabic lebih mendukung penilaian keterampilan berbicara melalui fitur pengenalan suara yang dapat memberikan umpan balik terhadap pelafalan peserta didik. Perbedaan karakteristik tersebut menunjukkan bahwa pemanfaatan AI dalam evaluasi pembelajaran bahasa Arab perlu disesuaikan dengan keterampilan yang hendak diukur. Dengan demikian, penggunaan kedua platform secara bersamaan dapat menjadi alternatif untuk mendukung proses evaluasi yang lebih menyeluruh, meskipun hasil penilaian tetap memerlukan verifikasi dan pertimbangan dari guru.

Landasan Teoretis Integrasi AI dalam Evaluasi Bahasa Arab

Secara konseptual, perubahan model evaluasi ini didukung oleh teori belajar konstruktivisme. Menurut perspektif konstruktivisme, pembelajaran yang efektif memberikan kesempatan kepada peserta didik untuk membangun pemahamannya melalui proses refleksi dan pengalaman belajar yang bermakna (Bahruddin et al., 2021).Melalui *prompting* pada ChatGPT, siswa dituntut berpikir kritis saat membetulkan kalimat bahasa Arab yang keliru. Umpan balik yang diberikan secara langsung oleh sistem AI juga dapat membantu peserta didik mengetahui kesalahan yang dilakukan dan segera melakukan perbaikan. Karakteristik ini sejalan dengan pemanfaatan teknologi digital yang mendukung proses belajar secara lebih aktif dan berkelanjutan (Rohim et al., 2024; Rohmawaty et al., 2024).

Potensi Rekonstruksi Pengujian Maharah Kitabah via Gen-AI

Pemanfaatan ChatGPT berpotensi mendukung perubahan evaluasi maharah kitabah dari yang

semula berfokus pada hasil akhir menjadi proses yang lebih formatif dan berkelanjutan. Melalui umpan balik otomatis, peserta didik dapat mengidentifikasi kesalahan dan melakukan perbaikan secara mandiri (Kasneci et al., 2023; Nasaruddin, 2024). Siswa tingkat SMA dapat mengunggah karangan teks pendek mereka (*insya'*) ke dalam ruang obrolan ChatGPT. Pada maharah kitabah, indikator yang dapat digunakan meliputi ketepatan struktur nahwu, pilihan mufradat, koherensi kalimat, dan kesesuaian konteks. AI dapat membantu mengidentifikasi kesalahan pada struktur kalimat, penggunaan kosakata, dan aspek kebahasaan lainnya sehingga proses revisi dapat dilakukan secara lebih cepat (Nasaruddin, 2024; Kohnke et al., 2023). Namun, pertanyaan kritis yang perlu dijawab adalah: apakah ChatGPT benar-benar mampu menilai tulisan bahasa Arab secara valid?

Kajian menunjukkan bahwa kemampuan ChatGPT dalam menilai aspek morfologi dan sintaksis bahasa Arab yang kompleks masih memiliki keterbatasan. Model ini cenderung lebih andal dalam mendeteksi kesalahan leksikal dan struktural tingkat dasar, tetapi belum mampu secara konsisten menangani kompleksitas morfologis bahasa Arab seperti tashrif dan i'rab. Selain itu, ada risiko bahwa AI hanya mengenali pola jawaban, bukan mengukur kemampuan bahasa yang sesungguhnya. Oleh karena itu, verifikasi dari guru yang kompeten dalam bahasa Arab tetap menjadi keharusan, bukan pilihan.

Potensi Otomatisasi Evaluasi Maharah Kalam

Evaluasi maharah kalam sering menghadapi berbagai kendala, terutama keterbatasan kesempatan praktik berbicara dan pemberian umpan balik secara individual kepada peserta didik. Pemanfaatan aplikasi berbasis AI berpotensi membantu mengatasi kendala tersebut melalui interaksi yang lebih intensif dan pemberian umpan balik secara langsung terhadap performa berbahasa peserta didik (Sakaroni & Azanulhaq, 2024; Maulana et al., 2024). Pada maharah kalam, indikator yang dapat digunakan meliputi pelafalan (makhraj al-huruf), kelancaran (tala'qah), ketepatan kosakata, respons komunikatif, dan keberanian berbicara. Pemanfaatan *chatbot* audio seperti Mondly Arabic membuka ruang yang aman bagi

siswa untuk berlatih: sistem AI bertindak sebagai mitra bicara virtual yang sabar, objektif, dan tidak menghakimi, sehingga siswa yang sebenarnya mampu namun tersandera rasa takut salah dapat berlatih tanpa tekanan sosial. Namun, pertanyaan kritis yang muncul adalah: apakah Mondly Arabic cocok untuk semua aksen dan dialek bahasa Arab? Kajian menunjukkan bahwa sistem pengenalan suara pada Mondly Arabic dirancang berdasarkan bahasa Arab standar modern (Fusha) dan memiliki keterbatasan dalam mengenali variasi dialek Arab serta aksen penutur non-Arab. Hal ini berpotensi menyebabkan penilaian yang kurang adil bagi siswa yang sebelumnya terpapar dialek Arab regional. Selain itu, perlu dipertanyakan pula apakah AI benar-benar menilai kemampuan bahasa siswa atau sekadar mengenali pola jawaban yang sesuai dengan pola pelatihan modelnya. Ini menjadi tantangan serius yang harus dijawab melalui validasi empiris sebelum model evaluasi berbasis AI diterapkan secara luas (Sakaroni & Azanulhaq, 2024).

Realitas Hambatan: Kesenjangan Teknologi, Pedagogi, dan Etika

Sebelum membahas hambatan teknis dan infrastruktur, terdapat dua pertanyaan mendasar yang perlu dijawab secara kritis. Pertama, bagaimana guru dapat memastikan umpan balik AI tidak salah? Guru adalah manusia yang mengenal siswanya—tahu siapa yang sedang berjuang, siapa yang butuh dorongan lebih, siapa yang perlu pendekatan berbeda. Oleh karena itu, guru perlu menjalankan fungsi verifikasi aktif, bukan sekadar menerima hasil evaluasi AI begitu saja. Ini dapat dilakukan melalui pemeriksaan silang hasil penilaian AI dengan rubrik penilaian konvensional, diskusi reflektif bersama siswa, dan pengamatan langsung terhadap proses belajar. Guru juga perlu memiliki literasi AI yang cukup untuk mengenali jenis kesalahan yang rentan dihasilkan oleh model bahasa, seperti halusinasi faktual atau penilaian yang terlalu permisif terhadap kesalahan morfologi bahasa Arab (Holmes et al., 2022; Kasneci et al., 2023). Kedua, apa risiko jika siswa terlalu bergantung pada AI? Ini pertanyaan yang layak direnungkan secara serius. Ketergantungan

berlebih dapat perlahan menggerus kemampuan berpikir mandiri siswa—mereka mungkin menjadi terbiasa mengandalkan koreksi otomatis tanpa benar-benar memahami alasan di balik kesalahan yang mereka buat. Risiko lain yang patut diwaspadai adalah terbentuknya ilusi kompetensi: siswa merasa sudah menguasai bahasa Arab hanya karena mendapat skor tinggi dari mesin, padahal kemampuan komunikatif nyata mereka—kemampuan berbicara dengan manusia sungguhan dalam situasi nyata—belum pernah benar-benar diuji. Untuk mengantisipasi risiko ini, guru perlu merancang tugas-tugas yang mendorong refleksi mendalam, bukan sekadar interaksi dengan AI (UNESCO, 2023).

Implementasi teknologi AI dalam evaluasi pembelajaran bahasa Arab tidak luput dari berbagai hambatan yang perlu mendapat perhatian serius. Pertama, hambatan infrastruktur: terdapat kesenjangan dalam ketersediaan infrastruktur digital yang dapat memengaruhi pemanfaatan teknologi AI dalam proses evaluasi pembelajaran. Keterbatasan akses internet dan perangkat digital masih menjadi tantangan dalam penerapan teknologi pendidikan secara merata (UNESCO, 2023). Selain itu, tidak semua platform berbasis AI dapat diakses secara bebas. Beberapa layanan menyediakan fitur lanjutan yang memerlukan biaya berlangganan sehingga berpotensi menjadi kendala bagi institusi pendidikan yang memiliki keterbatasan sumber daya (UNESCO, 2023).

Kedua, dari aspek validitas dan reliabilitas, hasil evaluasi berbasis AI tidak selalu konsisten dan dapat bervariasi tergantung pada cara siswa merumuskan pertanyaan (*prompting*). Selain itu, sistem AI generatif juga berpotensi menghasilkan informasi yang kurang akurat atau tidak sesuai dengan konteks yang diharapkan. Oleh karena itu, hasil evaluasi yang diberikan oleh AI perlu ditinjau kembali oleh guru agar tetap sesuai dengan tujuan pembelajaran dan karakteristik peserta didik (Kasneci et al., 2023; UNESCO, 2023). Ketiga, dari aspek etika dan privasi, penggunaan platform AI komersial menimbulkan pertanyaan mengenai keamanan data siswa, terutama bagi peserta didik di bawah umur. Selain itu, ketergantungan berlebih terhadap AI dikhawatirkan dapat mengikis kemampuan berpikir mandiri siswa

jika tidak dikelola dengan baik (UNESCO, 2023). Keempat, dari aspek kompetensi guru, Implementasi AI dalam pembelajaran memerlukan kompetensi digital yang memadai agar guru mampu mengintegrasikan teknologi secara efektif dan tetap mempertahankan fungsi pedagogisnya dalam proses pembelajaran maupun evaluasi (Holmes et al., 2022; UNESCO, 2023).

Meskipun chatbot AI mampu memberikan umpan balik secara cepat dan konsisten, hasil evaluasi yang dihasilkan tetap memerlukan verifikasi dari guru. Hal ini disebabkan AI pada dasarnya bekerja berdasarkan pengenalan pola dari data yang tersedia, sehingga belum sepenuhnya mampu menangkap aspek pedagogis, konteks komunikasi, maupun karakteristik kebahasaan bahasa Arab yang bersifat kompleks dan kontekstual. Dalam konteks maharah kitabah dan maharah kalam, penilaian terhadap ketepatan makna, kesesuaian konteks, serta tujuan komunikasi masih membutuhkan pertimbangan profesional dari guru sebagai evaluator utama.

Model Konseptual Evaluasi Autentik Bahasa Arab Berbasis AI

Berdasarkan sintesis terhadap 15 literatur yang telah dianalisis, penelitian ini menawarkan model konseptual evaluasi autentik bahasa Arab berbasis AI yang dapat diterapkan pada jenjang sekolah menengah. Model ini dikembangkan dari integrasi berbagai temuan yang relevan, bukan hanya dari satu penelitian tertentu. Setiap tahap dalam model ini dirancang untuk saling melengkapi dan membentuk satu siklus evaluasi yang utuh—di mana siswa bukan sekadar dinilai, tetapi juga dibimbing untuk tumbuh.

Dalam model ini, AI tidak diposisikan sebagai pengganti guru, melainkan sebagai instrumen pendukung yang membantu proses evaluasi secara lebih efisien. Guru tetap memegang peran utama sebagai perancang pembelajaran, penafsir hasil evaluasi, dan pembimbing pengembangan kompetensi bahasa peserta didik (Luckin et al., 2016; Holmes et al., 2022). Secara operasional, setiap tahap model ini dapat dijelaskan sebagai berikut. Pada Tahap 1 (Pemberian Tugas Autentik), guru merancang tugas kontekstual seperti

menulis paragraf deskriptif dalam bahasa Arab atau melakukan simulasi percakapan harian; instrumen AI yang digunakan adalah ChatGPT untuk tugas menulis dan Mondly Arabic untuk tugas berbicara; indikator keberhasilan adalah kesesuaian tugas dengan konteks komunikatif nyata. Pada Tahap 2 (Evaluasi Berbantuan AI), siswa secara mandiri menggunakan platform AI untuk mendapatkan umpan balik awal; jenis umpan balik mencakup koreksi nahwu dan sharaf untuk kitabah, serta skor pelafalan untuk kalam; guru memantau proses dan memastikan penggunaan AI sesuai panduan etika. Pada Tahap 3 (Refleksi dan Perbaikan Mandiri), siswa menganalisis umpan balik AI, merevisi hasil kerja mereka, dan mendokumentasikan proses perbaikan; peran guru adalah memfasilitasi diskusi refleksi kelompok atas temuan evaluasi AI. Pada Tahap 4 (Validasi Guru), guru memverifikasi hasil evaluasi AI, menambahkan penilaian aspek afektif dan pragmatik yang tidak dapat ditangkap oleh AI, lalu memberikan nilai akhir yang komprehensif; validasi ini adalah komponen yang tidak dapat dihilangkan dari model ini.

Model konseptual yang ditawarkan terdiri atas empat tahapan utama yang disajikan dalam Tabel 3 berikut:

Tabel 3. Model Konseptual Evaluasi Autentik Bahasa Arab Berbasis AI

Tahap	Komponen	Peran Guru	Peran Siswa	Peran AI
1. Pemberian Tugas Autentik	Tugas menulis (insya') atau berbicara (hiwar) kontekstual	Merancang aktivitas berbasis penggunaan bahasa nyata	Menerima instruksi dan memahami kriteria tugas	Menyediakan contoh prompt atau skenario percakapan
2. Evaluasi Berbantuan AI	Analisis otomatis hasil kerja siswa	Memantau proses dan menetapkan panduan penggunaan AI	Menggunakan ChatGPT/Mondly untuk evaluasi mandiri	Memberikan umpan balik otomatis: koreksi qawaid, skor pelafalan
3. Refleksi & Perbaikan Mandiri	Revisi berdasarkan umpan balik AI	Memfasilitasi diskusi refleksi hasil evaluasi	Menganalisis kesalahan dan melakukan revisi mandiri	Memberikan penjelasan aturan bahasa dan alternatif kalimat
4.Validasi Guru	Verifikasidan penilaian akhir	Mengkonfirmasi, menambahkan penilaian afektif, dan memberikan nilai akhir	Menerima umpan balik komprehensif dari guru	Menyediakan data hasil evaluasi sebagai bahan pertimbangan guru

Melalui model ini, evaluasi pembelajaran bahasa Arab tidak hanya berorientasi pada hasil akhir, tetapi juga memperhatikan proses belajar peserta didik. Integrasi AI memungkinkan terciptanya evaluasi yang lebih autentik karena siswa memperoleh kesempatan untuk belajar dari kesalahan dan melakukan perbaikan secara berkelanjutan. Mekanisme validasi guru pada tahap keempat menjadi bagian yang sangat penting untuk memastikan bahwa hasil evaluasi yang dihasilkan AI tetap akurat dan sesuai dengan tujuan pembelajaran. Peran guru tetap tak tergantikan: menafsirkan hasil evaluasi dengan empati, memberi umpan balik yang kontekstual dan penuh pertimbangan manusiawi, serta memperhatikan aspek pedagogis dan afektif yang tidak pernah bisa

sepenuhnya diakomodasi oleh sistem otomatis (Holmes et al., 2022; UNESCO, 2023). Oleh karena itu, pemanfaatan AI sebagai instrumen evaluasi autentik berpotensi menjadi salah satu inovasi penting dalam mendukung pembelajaran bahasa Arab yang relevan dengan tuntutan pendidikan abad ke-21 (UNESCO, 2023).

D. Kesimpulan

Berdasarkan hasil telaah terhadap 15 literatur yang relevan, pemanfaatan *Artificial Intelligence* (AI) menunjukkan potensi yang cukup besar untuk mendukung pelaksanaan evaluasi autentik dalam pembelajaran bahasa Arab di tingkat sekolah menengah. Kehadiran teknologi AI memungkinkan proses evaluasi berlangsung secara lebih adaptif, interaktif, dan berkelanjutan melalui penyediaan umpan balik yang cepat serta kesempatan bagi peserta didik untuk melakukan refleksi dan perbaikan secara mandiri. Dalam konteks *maharah kitabah*, AI berpotensi dimanfaatkan untuk membantu mengidentifikasi kesalahan kebahasaan dan memberikan alternatif penggunaan kosakata. Pada *maharah kalam*, teknologi pengenalan suara berpotensi mendukung penilaian aspek pelafalan secara lebih objektif.

Perlu ditekankan bahwa klaim-klaim mengenai efektivitas AI dalam kajian ini disampaikan sebagai potensi berdasarkan literatur yang dikaji, bukan sebagai kesimpulan empiris langsung. Efektivitas pemanfaatan AI dalam evaluasi pembelajaran bahasa Arab dipengaruhi oleh berbagai faktor, seperti kesiapan infrastruktur digital, ketersediaan akses teknologi, reliabilitas sistem AI, potensi bias algoritma, isu privasi data, serta kompetensi literasi digital guru. Oleh karena itu, AI perlu diposisikan sebagai instrumen pendukung yang memperkuat proses evaluasi, bukan sebagai pengganti peran guru yang memiliki tanggung jawab dalam membimbing, mengarahkan, dan menilai perkembangan peserta didik secara komprehensif.

Selain itu, penelitian ini menawarkan model konseptual evaluasi autentik bahasa Arab berbasis AI yang mengintegrasikan empat tahapan: pemberian tugas autentik, evaluasi berbantuan AI, refleksi dan perbaikan mandiri, serta validasi guru. Model tersebut menunjukkan bahwa pemanfaatan AI

akan lebih efektif apabila digunakan dalam satu siklus evaluasi yang melibatkan peran aktif guru sebagai validator akhir.

Penelitian selanjutnya disarankan untuk melakukan pengujian empiris melalui studi lapangan atau penelitian eksperimen guna mengukur secara langsung efektivitas dan keterbatasan penerapan model evaluasi berbasis AI dalam meningkatkan kompetensi bahasa Arab peserta didik. Kajian lanjutan juga perlu memberi perhatian serius pada aspek validitas dan reliabilitas penilaian AI, isu etika dan perlindungan data peserta didik, serta potensi bias algoritma dalam konteks bahasa Arab yang kaya variasi dialek dan kompleksitas morfologis. Pada akhirnya, teknologi yang baik adalah teknologi yang melayani manusia bukan sebaliknya

DAFTAR PUSTAKA

Agustina, R., Zaim, M., Thahar, H. E., & Afroka, M. (2024). Penggunaan teknologi artificial intelligence (AI) dalam pembelajaran bahasa di madrasah: Systematic literature review. *Jurnal Pendidikan Agama Islam Al-Thariqah*, 9(1), 110–122. https://doi.org/10.25299/althariqah.v9i1.16741

Azanulhaq, & Sakaroni, R. (2024). Persepsi siswa terhadap integrasi teknologi AI TalkPal dalam pembelajaran bahasa Arab. *EDUKA: Jurnal Pendidikan, Hukum dan Bisnis*, 9(2), 90–105. https://doi.org/10.32493/eduka.v9i2.39538

Bahruddin, U., Amrullah, A. M. K., & Audina, N. A. (2021). Konstruktivisme dalam perkuliahan maharah kalam menggunakan media Instagram: Implementasi, problematika dan persepsi mahasiswa di Indonesia. *Arabiyatuna: Jurnal Bahasa Arab*, 5(1), 127–140. https://doi.org/10.29240/jba.v5i1.2396

Chapelle, C. A., & Chung, Y. R. (2010). The promise of NLP and speech processing technologies in language assessment. *Language Testing*, 27(3), 301–315. https://doi.org/10.1177/0265532210364194

Dikli, S. (2006). An overview of automated scoring of essays. *Journal of Technology, Learning, and Assessment*, 5(1), 1–35.

Farhan, A., Kamal, M., & Sarip, M. (2024). Pengembangan instrumen evaluasi berbasis digital untuk pembelajaran bahasa Arab di sekolah menengah. *Arabiyatuna: Jurnal Bahasa Arab*, 8(2), 487–508. https://doi.org/10.29240/jba.v8i2.9858

Holmes, W., Bialik, M., & Fadel, C. (2022). *Artificial intelligence in education: Promises and implications for teaching and learning*. Center for Curriculum Redesign.

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

Kholis, N., & Mustofa, M. A. (2024). Development of competency-based Arabic language curriculum in traditional Islamic boarding schools. *Arabiyatuna: Jurnal Bahasa Arab*, 8(2), 827–848. https://doi.org/10.29240/jba.v8i2.10819

Kohnke, L., Moorhouse, B. L., & Zou, D. (2023). ChatGPT for language teaching and learning. *RELC Journal*, 54(2), 537–550. https://doi.org/10.1177/00336882231162868

Luckin, R., Holmes, W., Griffiths, M., & Forcier, L. B. (2016). *Intelligence unleashed: An argument for AI in education*. Pearson.

Maulana, M. R., Ridha, Z., & Ahmed, B. M. B. (2024). Language environment on speaking skills training in Islamic middle school. *Arabiyatuna: Jurnal Bahasa Arab*, 8(2), 617–640. https://doi.org/10.29240/jba.v8i2.10962

Muthohharoh, N., Mardliyah, M. K., Rifqi, A. F., Muslih, Hamdun, K., Hamidah, A., Ahmad, N. C., Islamiyah, N., Qudsiyah, N., Rochmawati, L., & Charisma, M. R. (2024). Evaluasi pembelajaran bahasa Arab berbasis AI Opexam. *Edukatif: Jurnal Ilmu Pendidikan*, 6(4), 3001–3010. https://doi.org/10.31004/edukatif.v6i4.7057

Nasaruddin, N. (2024). Using ChatGPT in teaching Arabic as a foreign language. *Arabiyatuna: Jurnal Bahasa Arab*, 8(1), 1–24. https://doi.org/10.29240/jba.v8i1.9413

Rohim, A. F., Maslamah, M., & Qosim, M. N. (2024). Penggunaan kecerdasan buatan untuk pembelajaran bahasa Arab yang adaptif dan terpersonalisasi. *MODELING: Jurnal Program Studi PGMI*, 11(2), 371–379. https://doi.org/10.69896/modeling.v11i2.2385

Rohmawaty, E. N., Hilmi, D., Salimul Uqba, M. S., & Saleh, U. S. (2024). Peran *Artificial Intelligence* (AI) dalam pembelajaran bahasa Arab mahasiswa pascasarjana UIN Maulana Malik Ibrahim Malang. *Khatulistiwa: Jurnal Pendidikan dan Sosial Humaniora*, 4(3), 316–328. https://doi.org/10.55606/khatulistiwa.v4i3.4023

Supriyanto, S., & Toifah, N. (2024). Efektifitas penggunaan artificial intelligence (AI) dalam pembelajaran bahasa Arab di era Society 5.0: Systematic literature review. *Lugawiyyat: Jurnal Bahasa Arab dan Linguistik*, 6(2), 1–15. https://doi.org/10.18860/lg.v6i2.29713

UNESCO. (2023). *Guidance for generative AI in education and research*. UNESCO. https://doi.org/10.54675/ZJTE2084