AL-MATLA
Journal of Arabic Language Education

Al Matla : Journal of Arabic Language Education
e-ISSN: 3125-0505
https://doi.org/00.0000/ Al Matla.x0x0.000
Volume 1, Issue. 2, 2026, Page 38-53
journal.staimh.ac.id

KECERDASAN BUATAN GENERATIF DALAM PEMBELAJARAN BAHASA ARAB: SYSTEMATIC LITERATURE REVIEW TENTANG TREN, MANFAAT, TANTANGAN, DAN ARAH PENELITIAN

Muslihin[*1], Rohman Pranoto[*2], Cahyudi Prima[*3]

[123] Sekolah Tinggi Agama Islam Minhajul Haq Purwakarta Indonesia
e-mail: muslihin@staimh.ac.id, rohmanpranoto1@staimh.ac.id, cahyudiprima@staimh.ac.id

Submitted: 08-07-2026	Revised : 22-08-2026	Accepted: 27-09-2026	Published: 31-10-2026

Abstract *The landscape of foreign language education, particularly Arabic language instruction, is currently undergoing a fundamental transformation driven by the emergence of Generative Artificial Intelligence (GenAI) and Large Language Models (LLMs). This Systematic Literature Review evaluates recent developments in the integration of GenAI into the Arabic language learning ecosystem, encompassing trends in the evolution of language models, pedagogical effectiveness, the utilization of automated assessment systems, as well as the linguistic and cultural challenges involved. Through a thematic synthesis of empirical and theoretical literature published between 2019 and 2026, this review confirms that GenAI significantly enhances learner engagement, instructional personalization, and language skill acquisition, particularly in vocabulary mastery and reading comprehension. These improvements are driven by the adoption of general-purpose models such as ChatGPT, as well as leading Arabic-centric models such as JAIS and ALLaM. Nevertheless, this innovative integration encounters the inherent structural complexity of the Arabic language. Generative models frequently experience technical limitations in handling syntactic rules (nahwu), morphology (shorof), diacritic disambiguation, and regional dialect variations, as evidenced by performance gaps in linguistic benchmarks such as AraLingBench and AlGhafa. Furthermore, this review discusses rapid advances in trait-based Automated Essay Scoring (AES) supported by precise prompt engineering. Based on these findings, the study concludes that although GenAI has significant potential to bring positive disruption, its successful implementation depends heavily on pedagogical readiness, digital literacy, the application of Human-in-the-Loop (HITL) collaboration, and the development of ethical frameworks aligned with the epistemology of Islamic higher education.*

Keywords: *Generative Artificial Intelligence; Arabic Language Learning; Systematic Literature Review; ChatGPT; Educational Technology.*

Abstrak Lanskap pendidikan bahasa asing, khususnya pengajaran bahasa Arab, saat ini sedang mengalami transformasi fundamental yang digerakkan oleh kemunculan teknologi Kecerdasan Buatan Generatif (*Generative Artificial Intelligence*/GenAI) dan *Large Language Models* (LLM). Tinjauan literatur sistematis (*Systematic Literature Review*) ini mengevaluasi perkembangan terkini dari integrasi GenAI dalam ekosistem pembelajaran bahasa Arab, mencakup tren evolusi teknologi model bahasa, efektivitas pedagogis, pemanfaatan sistem penilaian otomatis, serta tantangan linguistik dan kultural yang menyertainya. Melalui sintesis tematik terhadap literatur empiris dan teoretis yang diterbitkan antara tahun 2019 hingga 2026, tinjauan ini mengonfirmasi bahwa GenAI secara signifikan meningkatkan keterlibatan pembelajar, personalisasi instruksional, dan akuisisi keterampilan berbahasa, khususnya dalam penguasaan kosakata dan pemahaman bacaan. Peningkatan ini didorong oleh adopsi model generik seperti ChatGPT maupun model spesifik bahasa Arab (*Arabic-centric*) terkemuka seperti JAIS dan ALLaM. Namun demikian, integrasi inovatif ini berbenturan dengan kompleksitas struktural inheren dari bahasa Arab. Model generatif sering kali mengalami kegagalan teknis dalam menangani kaidah sintaksis (*nahwu*), morfologi (*shorof*), disambiguasi tanda diakritik, dan variasi dialek regional, yang dibuktikan oleh kesenjangan performa empiris dalam *benchmark* linguistik seperti AraLingBench dan AlGhafa. Lebih jauh, tinjauan ini mendiskusikan kemajuan pesat dalam Sistem Penilaian Esai Otomatis (*Automated Essay Scoring*/AES) berbasis sifat (*trait-based*) yang didukung oleh rekayasa perintah (*prompt engineering*) yang presisi. Berdasarkan temuan-temuan tersebut, studi ini menyimpulkan bahwa meskipun GenAI memiliki potensi disrupsi yang positif, keberhasilannya sangat bergantung pada kesiapan pedagogis, literasi digital, penerapan kolaborasi *Human-in-the-Loop*

© 2026 by the authors; This is an Open Access article distributed under the terms of the Creative Commons Attribution-ShareAlike 4.0 International License. (https://creativecommons.org/licenses/by-sa/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

(HITL), serta pengembangan kerangka etis yang selaras dengan epistemologi pendidikan tinggi Islam.

Kata Kunci: Kecerdasan Buatan Generatif; Pembelajaran Bahasa Arab; Systematic Literature Review; ChatGPT; Teknologi Pendidikan.

1. PENDAHULUAN

Perkembangan pesat Kecerdasan Buatan Generatif (*Generative Artificial Intelligence*/GenAI) telah membawa perubahan besar pada berbagai aspek pendidikan global, terutama dalam bidang pengajaran dan pembelajaran bahasa asing (Alkaabi & Almaamari, 2025). Dalam beberapa tahun terakhir, teknologi berbasis Kecerdasan Buatan telah berkembang jauh melampaui fungsi awalnya sebagai perangkat analisis sederhana atau pemeriksa ejaan. Perkembangannya kini menghasilkan agen percakapan interaktif, evaluator formatif dengan kapasitas tinggi, serta generator konten dinamis yang mampu mereplikasi bentuk interaksi menyerupai komunikasi manusia (Madjid, 2022). Dalam pembelajaran bahasa Arab, pemanfaatan GenAI menghadirkan paradigma baru yang secara teoritis memungkinkan pembelajaran menjadi lebih personal, pemberian umpan balik berlangsung secara cepat, dan keterlibatan peserta didik dapat ditingkatkan melebihi batas yang umumnya ditemukan dalam pembelajaran konvensional di kelas (Sahrir et al., 2025).

Bahasa Arab memiliki karakteristik linguistik, historis, dan sosial-budaya yang khas. Sebagai salah satu bahasa dengan jumlah penutur global yang besar, bahasa Arab memiliki karakter diglosia yang kuat. Kondisi tersebut terlihat dari perbedaan struktural dan leksikal antara *Modern Standard Arabic* (MSA) atau *Fusha*, yang digunakan dalam lingkungan akademik, komunikasi formal, literatur, serta teks keagamaan, dengan berbagai dialek regional (*Ammiyah*) yang lebih banyak digunakan dalam komunikasi sehari-hari (Attruk & Mustapha, 2026). Selain persoalan sosiolinguistik tersebut, bahasa Arab juga mempunyai sistem morfologi non-konkatenatif yang kompleks melalui pola akar kata (*root-and-pattern system*). Penggunaan tanda diakritik (*harakat*) juga berperan penting dalam menentukan fungsi sintaksis dan ketepatan makna suatu kata (Nelken & Shieber, 2025). Karakteristik tersebut menyebabkan pengembangan pemrosesan bahasa alami (*Natural Language Processing*/NLP) dan model generatif untuk bahasa Arab menghadapi tantangan yang relatif lebih kompleks dibandingkan sejumlah bahasa Indo-Eropa.

Kehadiran *Large Language Models* (LLM) komersial seperti ChatGPT yang dikembangkan oleh OpenAI, serta berbagai model *open-source* yang dirancang dengan orientasi terhadap bahasa Arab seperti JAIS dari G42 dan ALLaM dari *Saudi Data and AI Authority* (SDAIA), turut mempercepat pemanfaatan AI dalam pembelajaran bahasa Arab (Umbar et al., 2025). Teknologi tersebut mulai digunakan dalam berbagai aktivitas pendidikan, seperti membantu menjelaskan konsep gramatikal yang kompleks dalam *nahwu* dan *shorof*, mendukung latihan percakapan, menerjemahkan teks dari berbagai dialek, hingga membantu proses penilaian tulisan akademik (Qader & Yalcin Arslan, 2019). Namun, berbagai manfaat tersebut juga disertai persoalan akademik dan pedagogis. Ketergantungan yang terlalu besar terhadap sistem otomatis dapat menimbulkan kesalahan pemahaman, khususnya ketika LLM menghasilkan "halusinasi"

linguistik, tidak mampu memahami nuansa idiomatik secara tepat, atau menghasilkan kekeliruan dalam mereproduksi teks keagamaan (Samaan et al., 2023).

Pembelajaran bahasa Arab, terutama pada lembaga pendidikan tinggi Islam di kawasan Asia Tenggara seperti Indonesia dan Malaysia maupun di kawasan Timur Tengah, memiliki orientasi yang tidak hanya bersifat komunikatif. Lembaga pendidikan juga dituntut menghasilkan peserta didik yang mampu memahami literatur klasik dan teks keagamaan dengan tingkat ketepatan filologis yang tinggi (Sahrir et al., 2025). Karena itu, penerapan GenAI dalam pembelajaran bahasa Arab tidak cukup dipahami sebagai pemanfaatan inovasi teknologi semata. Implementasinya juga memerlukan perhatian terhadap etika, integritas akademik, serta sensitivitas terhadap aspek teologis. Adanya perbedaan antara potensi besar GenAI dan kesiapan pedagogis pengguna menjadi alasan penting untuk melakukan sintesis yang lebih komprehensif terhadap hasil penelitian yang telah tersedia.

Tinjauan Literatur Sistematis (*Systematic Literature Review*/SLR) ini disusun untuk mengintegrasikan, menilai, dan mensintesis penelitian empiris maupun teoretis terkini mengenai pemanfaatan Kecerdasan Buatan Generatif dalam pembelajaran bahasa Arab. Secara khusus, kajian ini membahas perkembangan teknologi model bahasa, efektivitas penggunaannya terhadap berbagai keterampilan berbahasa, keterbatasan linguistik dan potensi bias budaya pada LLM, serta arah pengembangan penelitian berikutnya, khususnya yang berkaitan dengan teknik *prompt engineering* dan sistem evaluasi otomatis.

2. METODE PENELITIAN

Penelitian ini menggunakan desain *Systematic Literature Review* (SLR) dengan mengacu pada kerangka *Preferred Reporting Items for Systematic Reviews and Meta-Analyses* (PRISMA) 2020 yang telah diperbarui (Page et al., 2021). Penggunaan PRISMA dimaksudkan untuk memastikan bahwa proses identifikasi, penyaringan, pemilihan, dan ekstraksi data dari penelitian yang relevan dengan GenAI dalam pembelajaran bahasa Arab berlangsung secara transparan, sistematis, dan dapat direproduksi (Umbar et al., 2025).

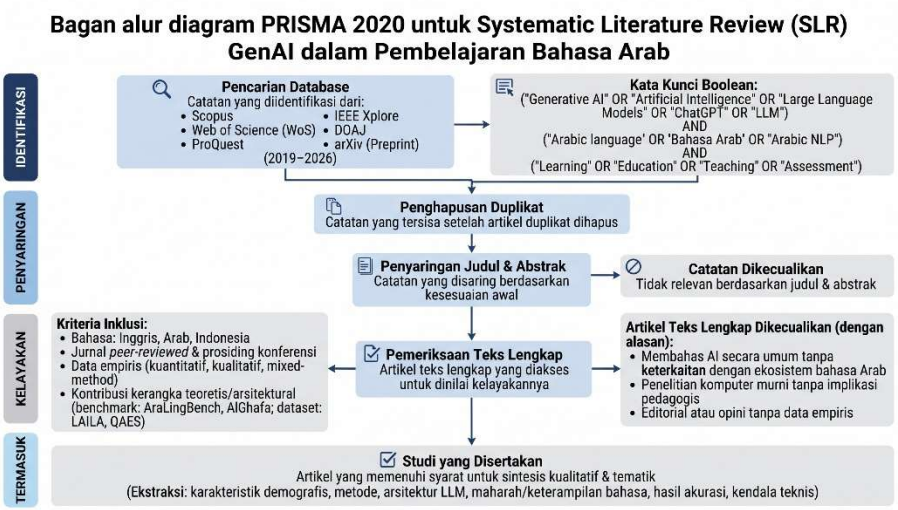
Strategi pencarian literatur dirancang untuk menjangkau penelitian dalam cakupan yang luas, baik yang berhubungan dengan inovasi teknologi maupun penerapannya dalam konteks pembelajaran. Proses pencarian dilakukan pada sejumlah pangkalan data akademik utama, yaitu Scopus, Web of Science (WoS), ProQuest, IEEE Xplore, serta sumber akses terbuka seperti *Directory of Open Access Journals* (DOAJ) dan repositori *preprint* seperti arXiv (Qiao & Zhao, 2023). Penyertaan repositori *preprint* dilakukan karena bidang NLP dan LLM berkembang dengan sangat cepat sehingga hasil penelitian mengenai arsitektur dan model bahasa sering kali tersedia sebelum melalui proses *peer-review* secara konvensional.

String pencarian Boolean disusun dengan menggabungkan tiga kelompok utama. Kelompok teknologi mencakup ("Generative AI" OR "Artificial Intelligence" OR "Large Language Models" OR "ChatGPT" OR "LLM"). Kelompok subjek mencakup ("Arabic language" OR

"Bahasa Arab" OR "Arabic NLP"), sedangkan kelompok konteks pendidikan meliputi ("Learning" OR "Education" OR "Teaching" OR "Assessment"). Rentang publikasi dibatasi pada tahun 2019 hingga 2026 untuk menggambarkan perkembangan teknologi dari periode awal model transformator generatif hingga munculnya model bahasa yang secara khusus dikembangkan untuk bahasa Arab.

Untuk mempertahankan kualitas hasil sintesis, penelitian ini menetapkan kriteria inklusi dan eksklusi secara terstruktur. Artikel yang dimasukkan harus ditulis dalam bahasa Inggris, Arab, atau Indonesia serta diterbitkan pada jurnal *peer-reviewed* atau prosiding konferensi internasional yang relevan. Penelitian juga harus memiliki data empiris yang dapat dipertanggungjawabkan melalui pendekatan kuantitatif, kualitatif, atau metode campuran, atau memberikan kontribusi berupa kerangka teoretis maupun arsitektural yang jelas, termasuk pengembangan *benchmark* evaluasi seperti AraLingBench dan AlGhafa serta dataset untuk Sistem Penilaian Esai Otomatis seperti LAILA dan QAES. Sebaliknya, penelitian yang hanya membahas AI secara umum tanpa keterkaitan dengan ekosistem bahasa Arab, penelitian komputer murni yang tidak membahas implikasi pedagogis, serta editorial atau tulisan opini tanpa dukungan data empiris dikeluarkan dari kajian (Grant & Booth, 2009; Ferrari, 2015).

Setelah proses penghapusan artikel duplikat dilakukan, kelayakan penelitian diperiksa melalui penyaringan terhadap judul dan abstrak, kemudian dilanjutkan dengan pemeriksaan teks lengkap. Informasi yang diekstraksi dari artikel yang memenuhi kriteria meliputi karakteristik demografis, metode penelitian, jenis arsitektur LLM yang digunakan, keterampilan berbahasa (*maharah*) yang menjadi fokus, hasil akurasi, serta berbagai kendala teknis yang ditemukan. Selanjutnya, data dianalisis secara tematik untuk menemukan pola empiris dan mengelompokkan temuan ke dalam tema utama. Tema tersebut mencakup perkembangan perangkat GenAI, reliabilitas penggunaannya dalam pembelajaran, serta inovasi *prompt engineering* dalam pendidikan bahasa Arab.



3. HASIL DAN DISKUSI

Analisis tematik terhadap penelitian primer yang berhasil dikumpulkan menunjukkan bahwa pemanfaatan Kecerdasan Buatan Generatif dalam pembelajaran bahasa Arab telah berkembang melampaui sekadar digitalisasi materi pembelajaran konvensional. GenAI kini turut memengaruhi cara struktur linguistik direpresentasikan, diproses, dihasilkan, dan dievaluasi. Berdasarkan hasil tersebut, pembahasan berikut difokuskan pada beberapa dimensi utama yang menggambarkan perkembangan terkini dalam ekosistem penelitian ini.

3.1 Evolusi dan Lanskap Model GenAI untuk Bahasa Arab

Perkembangan teknologi komputasi global telah membawa perubahan terhadap lanskap teknologi pembelajaran bahasa Arab. Sebelum tahun 2022, berbagai sistem pembelajaran adaptif pada umumnya masih menggunakan pendekatan berbasis aturan (*rule-based systems*) atau teknik *machine learning* parametrik konvensional (Ouahrani & Bennouar, 2020). Sistem generasi awal tersebut memiliki keterbatasan karena sebagian besar hanya mampu melakukan analisis leksikal sederhana, mendeteksi kesalahan ejaan berdasarkan pencocokan kamus (*dictionary-matching*), dan belum mampu memahami konteks semantik secara mendalam (Suharmawan, 2023). Kemunculan *Large Language Models* kemudian membawa perubahan signifikan terhadap kemampuan teknologi tersebut dalam mendukung pembelajaran bahasa.

Dominasi ChatGPT dan Keterbatasan Model Multibahasa Generik

Berdasarkan analisis penggunaan AI yang dilakukan oleh Umbar et al. (2025), ChatGPT yang menggunakan arsitektur GPT dari OpenAI menjadi salah satu alat yang paling banyak ditemukan dalam penelitian empiris mengenai pembelajaran bahasa. Kemampuannya dalam menghasilkan teks, memperbaiki tata bahasa, serta menjalankan fungsi sebagai mitra percakapan interaktif membuat ChatGPT banyak digunakan sebagai asisten virtual dalam pembelajaran bahasa (Rachbini et al., 2023). Pada sejumlah perguruan tinggi di Indonesia dan Malaysia, mahasiswa memanfaatkan teknologi tersebut untuk memperoleh terjemahan secara cepat dan membantu menemukan referensi, yang dalam penelitian tertentu dikaitkan dengan peningkatan efisiensi aktivitas akademik (Sahrir et al., 2025).

Walaupun model multibahasa berukuran besar seperti GPT-4 dan Gemini mempunyai kemampuan yang luas, literatur juga menunjukkan adanya sejumlah keterbatasan arsitektural. Model generalis tersebut dilatih menggunakan jumlah data yang sangat besar dengan proporsi bahasa Inggris dan struktur bahasa Indo-Eropa yang relatif dominan. Kondisi tersebut dapat memengaruhi performanya ketika digunakan untuk memproses bahasa Arab. Dalam situasi tertentu, model dapat mengalami fenomena yang disebut "keterlupaan katastropik" (*catastrophic forgetting*) atau mentransfer pola sintaksis bahasa Inggris ke dalam struktur bahasa Arab (Samaan et al., 2023). Akibatnya, keluaran model terkadang memiliki karakter seperti hasil terjemahan literal, kurang tepat dalam menempatkan struktur *mubtada muakkhor* (subjek yang ditunda), serta menunjukkan penurunan kualitas gramatikal hingga 13% dibandingkan dengan generasi teks

dalam bahasa Inggris (Khan, 2020; Ni et al., 2022).

Kemunculan LLM Spesifik Bahasa Arab (*Arabic-Centric LLMs*)

Keterbatasan model generalis tersebut mendorong penelitian AI menuju pengembangan LLM yang sejak awal dilatih (*from scratch*) atau disesuaikan secara khusus untuk merepresentasikan karakteristik bahasa Arab. Dalam perkembangan ini, JAIS dan ALLaM menjadi dua model yang mendapat perhatian penting dalam lanskap LLM berorientasi bahasa Arab.

JAIS dikembangkan melalui inisiatif G42 bersama Cerebras Systems sebagai LLM *open-source* bilingual Arab-Inggris yang menggunakan arsitektur transformator *decoder-only* berbasis GPT-3 (Alkaabi & Almaamari, 2025). Varian terbesar JAIS, yaitu JAIS 70B, dilatih menggunakan korpus pra-pelatihan sekitar 1,6 triliun token yang mencakup *Modern Standard Arabic* (MSA), transkripsi berbagai dialek regional, serta data untuk pemetaan terjemahan lintas bahasa. Dalam konteks pendidikan bahasa, JAIS 70B digunakan untuk mendukung aktivitas seperti merangkum literatur akademik berbahasa Arab, menghasilkan bahan bacaan yang disesuaikan dengan kerangka kemahiran CEFR (*Common European Framework of Reference*), serta menjalankan percakapan tutor yang mempertimbangkan konteks budaya lokal. Aspek-aspek tersebut menjadi salah satu area yang sering menjadi tantangan bagi LLM generalis.

Sementara itu, ALLaM yang dikembangkan oleh *Saudi Data and AI Authority* (SDAIA) menggunakan pendekatan pelatihan yang memiliki kemiripan dengan konsep Akuisisi Bahasa Kedua (*Second Language Acquisition*). Model ALLaM-1-13b-instruct dikembangkan dengan memanfaatkan infrastruktur NVIDIA/MegatronLM dan menjalani proses pra-pelatihan dalam dua tahap. Tahap pertama menggunakan sekitar 4 triliun token berbahasa Inggris untuk membangun kemampuan inferensi dan struktur logika umum. Pada tahap berikutnya, model dilatih menggunakan sekitar 1,2 triliun token yang terdiri atas data bahasa Arab dan Inggris (Bari et al., 2024). Strategi transfer pengetahuan secara bertahap tersebut diarahkan untuk membangun kemampuan morfologi bahasa Arab sekaligus mempertahankan kemampuan penalaran analitis yang telah diperoleh sebelumnya.

Perkembangan JAIS dan ALLaM menunjukkan adanya perubahan arah dalam pengembangan teknologi pembelajaran bahasa Arab. Pengembangan model tidak lagi hanya berorientasi pada pelokalan teknologi yang telah tersedia, tetapi juga mengarah pada pembangunan model yang memiliki pemahaman lebih spesifik terhadap karakteristik linguistik dan epistemologi bahasa Arab.

3.2 Peningkatan Pedagogis dan Keterlibatan Pembelajar

Sejumlah penelitian empiris menunjukkan bahwa penggunaan GenAI berpotensi mendukung efektivitas pembelajaran pada berbagai kompetensi berbahasa (*maharah*). Salah satu karakteristik yang paling menonjol adalah kemampuan LLM dalam membangun lingkungan belajar yang dapat disesuaikan dengan kebutuhan individu, bersifat interaktif, serta mampu beradaptasi dengan tingkat kemampuan pembelajar (Madjid, 2022).

Dalam penelitian komparatif berskala besar mengenai persepsi dan penerimaan mahasiswa pada institusi pendidikan tinggi di Malaysia dan Indonesia, Sahrir et al. (2025) mengkaji pemanfaatan GenAI berdasarkan sejumlah parameter. Survei kuantitatif tersebut mencakup aspek motivasi, penguasaan literasi, dan pemantauan perkembangan pembelajaran. Hasil penelitian tersebut kemudian dirangkum dalam Tabel 1.

Tabel 1. Pemanfaatan GenAI

Dimensi Evaluasi Pedagogis	Rata-rata Skor (Skala 1-6)	Tingkat Efektivitas
Keterlibatan Belajar (Learning Engagement)		
Menyederhanakan kemudahan akses ke materi	5.42	Sangat Tinggi
Membuat proses belajar lebih menarik	5.39	Sangat Tinggi
Meningkatkan motivasi intrinsik belajar bahasa Arab	5.32	Sangat Tinggi
Peningkatan Keterampilan (Skill Enhancement)		
Akselerasi pembelajaran kosakata baru (*Mufrodat*)	5.34	Sangat Tinggi
Mendorong penguasaan bacaan (*Qira'ah*)	5.33	Sangat Tinggi
Meningkatkan pemahaman tata bahasa (*Nahwu*)	5.26	Sangat Tinggi
Meningkatkan praktik pengucapan (*Kalam*)	5.22	Sangat Tinggi
Pemahaman Konteks (Contextual Understanding)		
Menjelaskan dan mengurai kata polisemi bahasa Arab	4.88	Menengah
Membedakan antara dialek Arab formal dan informal	4.82	Menengah

Sumber: Evaluasi Kuantitatif Persepsi Mahasiswa terhadap Efektivitas GenAI dalam Pembelajaran Bahasa Arab (Data diadaptasi dari temuan Sahrir et al., 2025).

Data tersebut memperlihatkan adanya hubungan yang kuat antara penggunaan antarmuka AI yang responsif dengan peningkatan motivasi intrinsik pembelajar. Salah satu fenomena psikologis yang menarik adalah perubahan posisi GenAI dari sekadar perangkat teknologi menjadi semacam "mitra latih tanding" (*sparring partner*) virtual yang memungkinkan mahasiswa berlatih dalam lingkungan yang relatif bebas dari rasa takut terhadap penilaian (Qiao & Zhao, 2023). Berkurangnya kecemasan dalam menggunakan bahasa Arab dapat mendorong mahasiswa untuk lebih berani mencoba struktur kalimat yang lebih kompleks, memperkaya kosakata, serta mengeksplorasi berbagai bentuk wacana secara mandiri tanpa kekhawatiran terhadap penilaian sosial.

Dalam keterampilan menulis (*Kitabah*) dan penerjemahan teks, manfaat penggunaan GenAI juga terlihat secara lebih konkret. Mahasiswa dapat memanfaatkan *chatbot* berbasis AI untuk menerjemahkan teks secara cepat, melakukan proses koreksi mandiri (*self-correction*), serta mengembangkan struktur sintaksis yang lebih koheren dibandingkan dengan hasil tulisan awal yang mereka kerjakan secara manual (Rachbini et al., 2023; Suharmawan, 2023). Selain itu, model generatif mampu memberikan penjelasan mengenai kesalahan bahasa, menawarkan alternatif

dalam penyusunan kembali frasa, dan menghasilkan variasi gaya penulisan (*stylistic variations*). Kemampuan tersebut dapat membantu mahasiswa mengembangkan sensitivitas terhadap aspek stilistika dalam penggunaan bahasa.

3.3 Hambatan pada Pemahaman Tata Bahasa (*Nahwu*) dan Morfologi (*Shorof*)

Walaupun GenAI menunjukkan berbagai manfaat dalam pembelajaran, penerapannya juga memperlihatkan sejumlah keterbatasan teknis ketika berhadapan dengan karakteristik inti bahasa Arab, terutama pada aspek tata bahasa (*nahwu*) dan morfologi (*shorof*). Bahasa Arab memiliki sistem derivasi non-konkatenatif yang memungkinkan satu akar kata (*juzr*) menghasilkan berbagai bentuk berdasarkan pola *wazan*. Perubahan tersebut dapat membentuk makna yang berbeda, mulai dari verba aktif dan pasif hingga nomina pelaku dan nomina tempat (Attruk & Mustapha, 2026; Hadi & Qohar, 2024). Kompleksitas tersebut semakin meningkat karena teks bahasa Arab dalam penggunaan sehari-hari sering ditulis tanpa tanda diakritik (*harakat*). Hilangnya informasi vokalisasi dapat menghasilkan ambiguitas semantik yang cukup kompleks bagi sistem komputasional (Toyin et al., 2024; Nelken & Shieber, 2025).

Ketika LLM digunakan sebagai tutor morfologi, sistem tidak hanya diharapkan mampu mengidentifikasi dan memperbaiki kesalahan konjugasi, tetapi juga menghasilkan contoh perubahan bentuk kata (*tasrif*) secara langsung (Qader & Yalcin Arslan, 2019). Namun, penelitian menunjukkan bahwa model LLM masih dapat mengalami kesulitan ketika harus melakukan restorasi diakritik (*diacritization*). Permasalahan tersebut dapat memengaruhi proses derivasi kata benda partisipatif, termasuk kesulitan dalam membedakan *Ism Fa'il* yang menunjukkan pelaku dengan *Ism Maf'ul* yang menunjukkan objek ketika informasi vokalisasi tidak tersedia. Sebagai salah satu upaya mengurangi kelemahan tersebut, pengembangan sistem pembelajaran bahasa mulai mengombinasikan kemampuan LLM dengan perangkat diagnostik linguistik yang lebih spesifik. Contohnya adalah penggunaan *sarf.one* untuk membantu memeriksa ketepatan konjugasi verba dan *ar-radif* untuk mendukung ketepatan pemilihan kosakata serta hubungan sinonim.

3.4 Validasi Keterbatasan melalui *Benchmark* Linguistik dan Distorsi Kultural

Keterbatasan LLM dalam memahami tata bahasa Arab tidak hanya terlihat melalui pengamatan kualitatif, tetapi juga telah diuji melalui berbagai *benchmark* linguistik yang dikembangkan untuk mengukur kemampuan model secara lebih sistematis. Salah satunya adalah *AraLingBench*, yaitu dataset evaluasi yang dirancang untuk menguji kemampuan intrinsik model bahasa dan melibatkan anotasi dari para ahli linguistik Arab yang berasal dari institusi akademik (Zbib et al., 2026).

Pengujian terhadap lebih dari 30 varian LLM berukuran besar melalui skenario *zero-shot* pada lima kategori kemampuan linguistik menunjukkan adanya kesenjangan performa yang cukup jelas. Temuan tersebut memberikan gambaran bahwa kefasihan keluaran yang dihasilkan AI tidak selalu mencerminkan kemampuan linguistik yang benar-benar setara dengan kompetensi penutur atau ahli bahasa Arab.

Tabel 2. Performa Model LLM dalam Menguasai Kategori Linguistik Bahasa Arab (Data diadaptasi dari hasil evaluasi AraLingBench oleh Zbib et al., 2026).

Kategori Kompetensi Linguistik	**Rata-rata Kinerja LLM (Median Skor)**	**Tingkat Kesulitan untuk LLM**
Pemahaman Ejaan (*Spelling*)	~60%	Terendah (Termudah)
Pemahaman Bacaan (*Reading*)	~58%	Rendah
Penguasaan Tata Bahasa (*Grammar*)	~53%	Tinggi
Kemampuan Morfologi (*Morphology*)	~50%	Sangat Tinggi
Analisis Sintaksis (*Syntax*)	~48%	Tertinggi (Tersulit)

Sumber : evaluasi AraLingBench oleh Zbib et al., 2026

Temuan yang diperoleh melalui metrik *AraLingBench* memiliki implikasi penting terhadap perkembangan NLP bahasa Arab. Perbedaan antara kemampuan LLM yang relatif baik dalam aspek ejaan dan kelemahannya dalam memahami sintaksis menunjukkan bahwa model AI saat ini masih cenderung mengenali pola ortografis pada tingkat permukaan (*surface-level mapping*), tetapi belum sepenuhnya mampu merepresentasikan struktur gramatikal hierarkis yang menjadi dasar pembentukan kalimat bahasa Arab (Zbib et al., 2026). Mekanisme prediksi token berikutnya (*next-token prediction*) yang menjadi dasar kerja berbagai model GPT juga menghadapi keterbatasan ketika harus memahami hubungan sintaksis yang memiliki jarak jauh (*long-distance dependencies*) dalam konstruksi kalimat bahasa Arab yang kompleks.

Selain persoalan linguistik, aspek kultural dan teologis juga menjadi perhatian penting dalam penerapan GenAI untuk pembelajaran bahasa Arab, terutama pada lingkungan pendidikan Islam (Alkaabi & Almaamari, 2025). Para pendidik mengemukakan kekhawatiran mengenai keterbatasan GenAI dalam memahami keragaman dialek regional serta dalam mempertahankan konteks dan orisinalitas budaya. Hasil evaluasi melalui *AlGhafa*, salah satu *benchmark* untuk menguji kemampuan LLM, juga menunjukkan adanya persoalan dalam menghasilkan maupun mengutip teks-teks sakral. Model dapat menghasilkan keluaran yang tidak akurat, misalnya kesalahan dalam penempatan diakritik pada reproduksi ayat Al-Qur'an yang berpotensi mengubah makna, maupun kekeliruan ketika menjelaskan persoalan dalam yurisprudensi fikih.

Salah satu contoh yang ditemukan berkaitan dengan ketidakmampuan LLM memahami konsensus hukum mengenai persoalan "puasa pada hari yang diragukan" (*Fasting on the day of doubt*). Kondisi tersebut menunjukkan adanya risiko epistemologis ketika keluaran AI diterima secara mutlak oleh mahasiswa tanpa melalui proses verifikasi, pendampingan intelektual dari dosen, maupun pemeriksaan terhadap sumber-sumber rujukan yang otoritatif.

3.5 Revolusi Sistem Penilaian Esai Otomatis (*Automated Essay Scoring*)

Di tengah berbagai keterbatasan tersebut, salah satu bidang yang mengalami perkembangan signifikan melalui pemanfaatan GenAI adalah Sistem Penilaian Esai Otomatis (*Automated Essay Scoring*/AES). Dalam perspektif historis, perkembangan teknologi AES untuk bahasa Arab memperlihatkan perubahan yang sejalan dengan kemajuan algoritma komputasional. Perkembangan tersebut dapat diklasifikasikan ke dalam tiga generasi utama, sebagaimana disajikan pada Tabel 3.

Tabel 3: Trajektori Evolusi Sistem Penilaian Esai Otomatis (AES) untuk Bahasa Arab (Sintesis dari kajian literatur).

Generasi Sistem AES	Teknologi yang Mendasari	Metodologi Analisis	Karakteristik Kinerja
Generasi Pertama	*Traditional Machine Learning*	SVM (Support Vector Machine), LSA (Latent Semantic Analysis), AraComLex	Mengandalkan ekstraksi fitur manual, pemrosesan leksikal permukaan, korelasi moderat, tidak memahami konteks semantik.
Generasi Kedua	*Deep Learning Architecture*	Jaringan Saraf Tiruan, RNN, LSTM, Bidirectional GRU, AraBERT	Menangkap hubungan kontekstual yang dalam. AraBERT mencapai korelasi Kuadratik (*QWK*) 0.88, meniadakan fitur rekayasa manual.
Generasi Ketiga	*Large Language Models (LLMs)* & GenAI	AraBART, GPT-4, JAIS, ALLaM, dengan *Prompt Engineering*	*Trait-centric evaluation* (penilaian berdasarkan dimensi linguistik spesifik), memberikan umpan balik formatif kualitatif, *few-shot learning*.

Generasi ketiga AES yang memanfaatkan LLM membawa perubahan mendasar dalam paradigma penilaian. Sistem tidak lagi terbatas pada pemberian skor numerik secara umum, tetapi mulai diarahkan pada evaluasi diagnostik berdasarkan karakteristik atau aspek tertentu (*trait-centric evaluation*) (Ghazawi & Simpson, 2024; El-Haj et al., 2024). Perkembangan tersebut didukung oleh tersedianya *dataset* beranotasi dalam jumlah besar yang digunakan untuk melakukan kalibrasi terhadap model. Salah satu contohnya adalah LAILA (*Linguistic Assessment with traIts for Learning Arabic*), yang menyediakan 7.859 esai tulisan tangan dan memberikan penilaian berdasarkan tujuh dimensi, yaitu relevansi, organisasi, kosakata, gaya, pengembangan gagasan, mekanika ejaan, dan ketepatan tata bahasa (Bashendy et al., 2026). Selain LAILA, hadir pula korpus seperti QAES dan AR-AES yang memberikan sumber data tambahan bagi peneliti untuk melatih LLM dalam mengenali berbagai tingkat kemampuan menulis mahasiswa (Almujaiwel et al., 2025).

Dalam perkembangan AES terbaru, pendekatan *Multi-Trait Specialization* (MTS) digunakan dengan memanfaatkan konsep sistem "multi-agen" pada LLM. Alih-alih meminta model menghasilkan satu skor keseluruhan, sistem dirancang agar LLM menjalankan beberapa peran evaluator dengan fokus penilaian yang berbeda. Sebagai contoh, satu *prompt* diarahkan secara khusus untuk mengevaluasi aspek tata bahasa, sedangkan *prompt* lainnya digunakan untuk menilai organisasi dan kohesi tulisan. Skor dari setiap aspek tersebut kemudian digabungkan menjadi

hasil penilaian akhir (Lee et al., 2024).

Selain digunakan untuk penilaian secara langsung, LLM juga mulai dimanfaatkan dalam proses *Data Augmentation* melalui teknik Injeksi Kesalahan (*Error Injection*). Dengan menggunakan LLM dan arsitektur transformator, peneliti dapat menghasilkan secara sintetis ribuan esai bahasa Arab yang mewakili berbagai tingkat kemampuan berdasarkan kerangka CEFR, mulai dari A1 hingga C2. Esai sintetis yang telah diberi anotasi tersebut kemudian dimanfaatkan sebagai data tambahan untuk proses pra-pelatihan sehingga dapat membantu mengatasi keterbatasan jumlah esai autentik berbahasa Arab yang tersedia (Qwaider et al., 2025).

3.6 Rekayasa Perintah (*Prompt Engineering*) dan Kolaborasi *Human-in-the-Loop* (HITL)

Berbagai perkembangan dalam evaluasi dan pembelajaran berbasis GenAI tidak terlepas dari kemajuan dalam bidang *prompt engineering* atau rekayasa perintah. Kemampuan LLM dalam memahami dan menghasilkan wacana berbahasa Arab sangat dipengaruhi oleh kejelasan, struktur, serta spesifikasi instruksi yang diberikan. Perintah yang tidak dirancang dengan baik dapat menyebabkan distorsi dalam pembentukan verba dan menurunkan ketepatan makna yang dihasilkan oleh model (Khan, 2020).

Penelitian eksperimental yang dilakukan oleh Awasthi menunjukkan bahwa *well-designed prompts* tidak hanya dapat meningkatkan kualitas keluaran teks, tetapi juga berpotensi mendorong kemampuan pemecahan masalah dan diskusi tingkat tinggi di kalangan pembelajar. Pada saat yang sama, perintah yang dirancang secara terstruktur memungkinkan batasan ruang lingkup analisis AI ditentukan secara lebih jelas.

Pada penerapan AES menggunakan dataset QAES, metode *few-shot prompting* yang dilengkapi dengan rubrik penilaian serta contoh esai dengan nilai tinggi dan rendah menunjukkan kemampuan dalam mengurangi *systematic penalty* yang sebelumnya dapat diberikan secara tidak proporsional oleh LLM terhadap aspek mekanika dan tata bahasa ketika menggunakan pendekatan *zero-shot* (Bashendy et al., 2024).

Rekayasa perintah juga memiliki peran dalam pengolahan semantik yang lebih kompleks, termasuk penerjemahan bahasa Arab ke dalam bahasa basis data melalui tugas *Text-to-SQL* pada dataset Ar-SParC. Penelitian tersebut menggunakan pendekatan rekayasa perintah hibrida dengan melibatkan *GAT Corrector* dan *GAT Verifier* berbasis LLM yang bekerja secara paralel untuk melakukan perbaikan dan validasi terhadap kueri yang memiliki konteks lintas domain (Ghazawi & Simpson, 2024).

Implikasi perkembangan tersebut terhadap pembelajaran bahasa Arab adalah semakin pentingnya penguasaan literasi *prompt*. Kemampuan menyusun instruksi yang efektif untuk LLM secara bertahap tidak lagi hanya menjadi keterampilan teknis yang bersifat khusus (*niche*), tetapi mulai berkembang menjadi bagian dari kompetensi literasi digital yang relevan bagi mahasiswa dan akademisi bidang bahasa Arab.

3.7 Integrasi Kerangka Etis dan Epistemologi Pendidikan Tinggi Islam

Di balik perkembangan teknologi dan meningkatnya penggunaan GenAI, terdapat sejumlah persoalan etis yang perlu mendapatkan perhatian. Kemudahan dalam mengakses sistem penerjemahan dan generator esai berbasis AI dapat mendorong terjadinya pelepasan beban kognitif (*cognitive offloading*). Ketika mahasiswa terlalu sering menyerahkan tugas analitis, seperti menguraikan struktur sintaksis puisi klasik atau mengidentifikasi bentuk konjugasi pasif, kepada AI, terdapat risiko bahwa proses pembentukan memori jangka panjang dan kemandirian intelektual tidak berkembang secara optimal (Attruk & Mustapha, 2026).

Sebagai respons terhadap persoalan tersebut, terdapat dorongan teoretis agar institusi pendidikan mengembangkan Kerangka Kompetensi AI (*AI Competency Framework*) yang disesuaikan dengan karakteristik pendidikan tinggi Islam. Kerangka tersebut menempatkan penggunaan teknologi dalam pengawasan filosofi pendidikan yang mengedepankan *adab* sebagai bagian dari etika digital sekaligus menempatkan pencarian kebenaran yang bersumber dari referensi otoritatif sebagai prinsip penting (Sahrir et al., 2025).

Dalam konteks tersebut, pendekatan **Human-in-the-Loop (HITL)** menjadi salah satu pendekatan yang relevan. LLM tidak ditempatkan sebagai sumber pengetahuan yang berdiri sendiri, melainkan sebagai perangkat pendukung komputasional. Mahasiswa dan dosen tetap memegang peran utama dalam memeriksa ketepatan linguistik, menilai kesesuaian konteks, serta memastikan bahwa penggunaan AI tidak mengabaikan integritas budaya dan nilai-nilai keagamaan.

4. KESIMPULAN

Tinjauan literatur sistematis (*Systematic Literature Review*) ini menunjukkan bahwa Kecerdasan Buatan Generatif memiliki potensi besar dalam mengubah ekosistem pembelajaran bahasa Arab. Berbagai perangkat GenAI dapat berfungsi sebagai pendukung pedagogis dengan menyediakan lingkungan belajar yang adaptif, membantu mengurangi kecemasan dalam menggunakan bahasa, memperluas penguasaan kosakata, serta memperluas akses terhadap evaluasi formatif melalui sistem AES berbasis *trait*. Perkembangan model *Large Language Models* yang dirancang secara khusus untuk ekosistem bahasa Arab, seperti JAIS dan ALLaM, juga menunjukkan arah pengembangan teknologi yang semakin memperhatikan karakteristik linguistik dan kebutuhan bahasa Arab.

Namun demikian, berbagai kemajuan tersebut tetap disertai keterbatasan teknis yang perlu diperhatikan. Di balik kemampuan menghasilkan teks yang tampak fasih pada tingkat permukaan, LLM masih menghadapi persoalan ketika harus memahami struktur hierarkis bahasa Arab secara mendalam. Keterbatasan tersebut terlihat pada kemampuan memahami sintaksis (*nahwu*), proses derivasi morfologis (*shorof*), serta interpretasi teks yang tidak menggunakan tanda diakritik. Hasil evaluasi melalui *benchmark* seperti AraLingBench turut menunjukkan bahwa kefasihan keluaran model tidak selalu menunjukkan penguasaan struktur linguistik secara akurat. Oleh sebab itu, penggunaan LLM secara tidak kritis dapat membawa risiko munculnya kesalahan

gramatikal, ketidakakuratan dialektal, maupun kekeliruan dalam memahami teks yang memiliki dimensi hukum dan keagamaan.

Dengan demikian, integrasi GenAI dalam pembelajaran bahasa Arab perlu menempatkan teknologi sebagai instrumen pendukung, bukan sebagai pengganti otoritas akademik dan pedagogis manusia. Pendekatan *Human-in-the-Loop* (HITL) dapat digunakan untuk memastikan bahwa keluaran AI tetap melalui proses pemeriksaan dan validasi oleh pendidik maupun pembelajar. Pendidik memiliki peran penting dalam memeriksa ketepatan struktur bahasa, memberikan arahan berdasarkan konteks sosial dan budaya, serta membangun kemandirian analitis peserta didik.

Untuk penelitian selanjutnya, pengembangan teknologi dapat diarahkan pada pendekatan hibrida yang menggabungkan *knowledge graph* atau graf pengetahuan dengan metode *deep learning* sehingga kemampuan model dalam memahami struktur linguistik dapat ditingkatkan. Pada saat yang sama, institusi pendidikan dapat mempertimbangkan penguatan literasi *prompt engineering* dalam kurikulum sebagai bagian dari kompetensi digital mahasiswa. Pengembangan tersebut perlu tetap ditempatkan dalam kerangka *adab*, etika penggunaan teknologi, dan integritas intelektual yang menjadi bagian penting dari pendidikan tinggi Islam.

DAFTAR PUSTAKA

Alkaabi, M. H., & Almaamari, A. S. (2025). Generative AI implementation and assessment in Arabic language teaching. *International Journal of Online Pedagogy and Course Design (IJOPCD)*, 15(1), 1–18. https://doi.org/10.4018/IJOPCD.368037

Abdurahman, I., Salim, A., Wangsadanureja, M., Alkhotiri, I., & Alpiansyah, R. (2026). Modernization of Human Resource Management in Islamic Boarding Schools: An Analysis of the Outlines of Personnel and Work Systems (GBKSK) Minhajul Haq. *Labina : Journal of Islamic Education Management*, *1*(1), 29-39. https://doi.org/10.66891/8wxyjc16

Alpiansyah, R., Affandy, A., & Dzulfikri. (2026). Branding Strategies For Islamic Educational Institutions: The Use Of Digital Marketing And Influencer Marketing To Attract Gen Z Prospective Students To Islamic-Based Schools. *Labina : Journal of Islamic Education Management*, *1*(2), 01-14. https://doi.org/10.66891/5brb6v48

Alpiansyah, R., Abdurahman, I., Alkhotiri, I., Wangsadanureja, M., & Dzulfikri. (2026). MANAJEMEN KURIKULUM DAN EKSTRAKURIKULER BERBASIS KARAKTER: INTEGRASI PENDIDIKAN JASMANI, PEMBIASAAN SPIRITUAL, DAN NEUROSAINS DALAM PEMBENTUKAN ADAB ISLAMI. *Labina : Journal of Islamic Education Management*, *1*(1), 62-90. https://doi.org/10.66891/v4dpda25

Attruk, M. F. H., & Mustapha, N. H. (2026). Generative AI in Arabic grammar learning: A critical review of pedagogical benefits and linguistic limitations. *Takuana*, 5(1), 577–594. https://doi.org/10.56113/takuana.v5i1.493

Bari, M. S., et al. (2024). ALLaM: A large language model for Arabic and English. *Proceedings of ArabicNLP*.

Bashendy, M., Massoud, W., Eltanbouly, S., Albatarni, S., Sayed, M., Abir, A., Bouamor, H., & Elsayed, T. (2026). LAILA: A large trait-based dataset for Arabic automated essay scoring. *Proceedings of the 19th Conference of the European Chapter of the Association for Computational Linguistics*, 1, 3074–3091.

Eden, A., et al. (2024). Navigating AI in education: Collaborative models for digital literacy. *Journal of Educational Technology*.

El-Haj, M., et al. (2024). Advancing automated essay scoring in Arabic. *Computers and Education*.

Ferrari, R. (2015). Writing narrative style literature reviews. *Medical Writing*, 24(4), 230–235.

Ghazawi, R., & Simpson, E. (2024). Automated essay scoring in Arabic: A dataset and analysis of a BERT-based system. *arXiv preprint*.

Grant, M. J., & Booth, A. (2009). A typology of reviews: An analysis of 14 review types and

associated methodologies. *Health Information & Libraries Journal*, 26(2), 91–108.

Hadi, S., & Qohar, A. (2024). Integrasi AI dalam pembelajaran sintaksis bahasa Arab. *Jurnal Inovasi Pendidikan*.

Khan, M. (2020). Challenges in semantic translation and prompt generation for low-resource languages. *Journal of Applied Linguistics*.

Lee, S., Cai, Y., Meng, D., Wang, Z., & Wu, Y. (2024). Unleashing large language models' proficiency in zero-shot essay scoring. *arXiv preprint arXiv:2404.04941*.

Madjid, A. (2022). Kecerdasan buatan dalam lanskap pendidikan modern. *Jurnal Pendidikan Sosial dan Humaniora*.

Nelken, R., & Shieber, S. M. (2025). Arabic diacritization using weighted finite-state transducers. *Harvard University Press*.

Ni, J., et al. (2022). Cross-lingual challenges in natural language generation. *IEEE Transactions on Computational Linguistics*.

Ouahrani, T., & Bennouar, D. (2020). AR-ASAG: An unsupervised Arabic automatic short answer grading system. *Journal of Natural Language Processing*.

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., & Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, 372(71).

Qader, R., & Yalcin Arslan, M. (2019). The impact of collaborative writing and automated feedback on L2 writing. *Linguistics and Education*.

Qiao, H., & Zhao, A. (2023). Artificial intelligence-based language learning: Illuminating the impact on speaking skills and self-regulation in Chinese EFL context. *IEEE*.

Qwaider, C., Alhafni, B., Chirkunov, K., Habash, N., & Briscoe, T. (2025). Enhancing Arabic automated essay scoring with synthetic data and error injection. *Proceedings of the 20th Workshop on Innovative Use of NLP for Building Educational Applications*, 549–563.

Rachbini, W., Evi, S., & Suyanto. (2023). Integrasi ChatGPT dalam pendidikan: Peluang dan tantangan. *Jurnal Ilmiah Multidisiplin*.

Sahrir, M. S., Albantani, A. M., Arifin, F., & Fathudin, F. (2025). The use of generative artificial intelligence (AI) in Arabic language education: Insights and implications between Malaysia and Indonesia. *International Journal of Research and Innovation in Social Science*, 9(3), 3638–3646. https://doi.org/10.47772/IJRISS.2025.90300288

Samaan, A., et al. (2023). Evaluating grammatical accuracy of machine translation in Arabic. *Language Resources and Evaluation*.

Suharmawan, W. (2023). Pemanfaatan ChatGPT sebagai alat penerjemahan bahasa asing. *Jurnal*

Sarjana Ilmu Budaya.

Toyin, A., et al. (2024). MultiDiac: Multilingual diacritic restoration benchmark. *Computational Linguistics Journal*.

Umbar, K., Azkiyah, S. N., Maswani, M., Wahab, M. A., & Zulkifli, Z. (2025). A systematic review of AI-powered language teaching trends, innovations, and challenges. *Ascarya: Journal of Islamic Science, Culture, and Social Studies*, 5(2), 248–261. https://doi.org/10.53754/rth2mc61

Zapata, G. C. (2025). Generative AI technologies, multiliteracies, and language education (1st ed.). *Routledge*.

Zbib, M., Hammoud, H. A. A. K., & Ghanem, B. (2026). AraLingBench: A human-annotated benchmark for evaluating Arabic linguistic capabilities of large language models. *Proceedings of the 2nd Workshop on NLP for Languages Using Arabic Script (AbjadNLP 2026)*, 385–393.