



Evaluasi Pembelajaran Bahasa Arab Berbasis Deep Learning: Model, Tantangan, dan Peluang di Era Digital

Dadan Mardani¹, Iis Susiawati²

IAI Al-AZIS, Indramayu, Indonesia^{1,2}

e-mail : iis.susiawati@iai-alzaytun.ac.id

Abstrak

Transformasi pendidikan di era digital menuntut adanya inovasi dalam proses evaluasi pembelajaran, termasuk pada bidang Pendidikan Bahasa Arab. Penilaian tradisional sering dianggap kurang mampu menangkap kompleksitas kompetensi berbahasa, terutama dalam keterampilan produktif seperti kalām dan kitābah. Perkembangan Deep Learning menawarkan peluang baru untuk menciptakan sistem evaluasi yang lebih akurat, adaptif, dan efisien. Artikel ini menggunakan metode studi pustaka (library research) dengan menganalisis riset terkini mengenai penerapan Deep Learning dalam evaluasi bahasa, teknologi NLP (Natural Language Processing) bahasa Arab, pengenalan suara (speech recognition), serta sistem automated scoring pada pembelajaran bahasa kedua. Hasil kajian menunjukkan bahwa Deep Learning mampu meningkatkan ketepatan deteksi kesalahan linguistik, mengotomatisasi penilaian keterampilan membaca dan berbicara, serta menyediakan analisis performa yang lebih komprehensif. Berbagai model seperti CNN, RNN, LSTM, dan Transformer telah berhasil digunakan untuk menganalisis fonetik, sintaksis, dan semantik bahasa Arab. Integrasi Deep Learning dalam evaluasi Bahasa Arab membuka peluang untuk menciptakan smart assessment, namun tetap menghadapi tantangan seperti kebutuhan data besar, bias model, keterbatasan korpus Arab, serta kesiapan kompetensi digital guru. Artikel ini menyimpulkan bahwa implementasi AI-driven assessment dapat memperkuat kualitas evaluasi sekaligus mendorong transformasi pembelajaran bahasa Arab di Indonesia.

Kata Kunci: Deep Learning; Evaluasi Pembelajaran Bahasa Arab; AI Assessment; NLP Arab; Pendidikan Era Digital.

Abstract

The transformation of education in the digital era requires innovation in the learning evaluation process, including in the field of Arabic Language Education. Traditional assessments are often considered incapable of capturing the complexity of language competence, especially in productive skills such as kalām and kitābah. The development of Deep Learning offers new opportunities to create more accurate, adaptive and efficient evaluation systems. This article uses a library research method by analyzing the latest research on the application of Deep Learning in language evaluation, NLP (Natural Language Processing) Arabic technology, speech recognition, and automated scoring systems in second language learning. The results of the study show that Deep Learning is able to improve the accuracy of linguistic error detection, automate the assessment of reading and speaking skills, and provide a more comprehensive performance analysis. Various models such as CNN, RNN, LSTM, and Transformer have been successfully used to analyze the phonetics, syntax, and semantics of Arabic. The integration of Deep Learning in Arabic language evaluation opens up opportunities to create smart assessments, but still faces challenges such as the need for big data, model bias, limitations of the Arabic corpus, and the readiness of teachers' digital competencies. This article concludes that the implementation of AI-driven assessment can strengthen the quality of evaluation while driving the transformation of Arabic language learning in Indonesia.

Keywords: Deep Learning; Arabic Language Learning Evaluation; AI Assessment; NLP Arab; Digital Era Education.

PENDAHULUAN

Perkembangan teknologi digital pada dekade terakhir telah membawa perubahan mendasar dalam dunia pendidikan, termasuk dalam konteks pembelajaran bahasa. Di antara berbagai inovasi tersebut, kecerdasan buatan (Artificial Intelligence/AI), khususnya cabang Deep Learning, telah menjadi pusat perhatian karena kemampuannya dalam melakukan analisis data besar, mengenali pola linguistik, serta memberikan umpan balik secara otomatis dan presisi (Zawacki-Richter & others, 2023). Dalam konteks

Pendidikan Bahasa Arab, perubahan ini semakin signifikan mengingat karakter bahasa Arab yang kompleks—baik dari aspek fonologi, morfologi, sintaksis, maupun semantik—sehingga membutuhkan pendekatan evaluasi yang lebih cermat, adaptif, dan efisien.

Di Indonesia, pembelajaran Bahasa Arab masih banyak menggunakan model evaluasi tradisional berupa tes tertulis, lisan, atau penilaian manual yang bergantung pada subjektivitas pendidik. Meskipun pendekatan tersebut memiliki kelebihan dari aspek kedekatan pedagogis, berbagai penelitian menunjukkan bahwa penilaian manual sering kali menghadapi kendala reliabilitas, konsistensi, dan efisiensi (Al-Ajmi & Al-Mutairi, 2024). Evaluasi keterampilan produktif seperti maharah al-kalam dan maharah al-kitābah bahkan cenderung tidak terstandar karena keterbatasan waktu, tenaga, dan kemampuan evaluator untuk menilai secara komprehensif seluruh aspek kebahasaan. Di sisi lain, beban administratif semakin meningkat seiring tuntutan akuntabilitas pendidikan di era digital.

Dalam situasi ini, Deep Learning menghadirkan peluang strategis untuk melakukan transformasi evaluasi pembelajaran Bahasa Arab. Model-model seperti Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), Long Short-Term Memory (LSTM), dan Transformer telah terbukti mampu memproses data teks dan suara bahasa Arab dengan tingkat akurasi yang tinggi. Teknologi ini memungkinkan dikembangkan sistem evaluasi otomatis (automated assessment) seperti speech recognition untuk penilaian keterampilan berbicara, automated writing evaluation untuk analisis kesalahan tata bahasa dan kosakata, serta machine reading assessment untuk menguji kemampuan memahami teks (Hussein & Al-Khalil, 2025).

Lebih jauh, kemajuan dalam bidang Natural Language Processing (Arabic NLP) dan ketersediaan korpus bahasa Arab modern membuka peluang bagi pengembangan platform evaluasi berbasis AI yang lebih kontekstual, inklusif, dan adaptif terhadap kemampuan peserta didik. Sistem ini tidak hanya meningkatkan akurasi penilaian, tetapi juga memberikan real-time feedback, menilai performa mahasiswa secara longitudinal, dan mengurangi bias evaluator manusia. Dengan demikian, integrasi Deep Learning dalam evaluasi pembelajaran Bahasa Arab berpotensi besar untuk meningkatkan kualitas pembelajaran sekaligus memenuhi kebutuhan zaman yang semakin digital dan berbasis data. Namun demikian, penerapan evaluasi berbasis Deep Learning tidak lepas dari tantangan. Keterbatasan korpus bahasa Arab yang terstandarisasi, kebutuhan data berlabel dalam jumlah besar, bias linguistik antar dialek, hingga kesiapan literasi digital para pendidik menjadi hambatan yang perlu diantisipasi. Oleh karena itu, kajian mendalam mengenai model, tantangan, dan peluang penerapan Deep Learning dalam evaluasi pembelajaran Bahasa Arab sangat penting dilakukan untuk memberikan arah yang jelas bagi pengembangan penelitian dan praktik pembelajaran di Indonesia.

Berdasarkan latar belakang tersebut, artikel ini bertujuan untuk: 1). menganalisis konsep dan kebutuhan evaluasi pembelajaran Bahasa Arab di era digital; 2). mengidentifikasi model-model Deep Learning yang relevan untuk penilaian Bahasa Arab; 3). mengkaji tantangan implementasi di konteks pendidikan Indonesia; serta 4) menawarkan peluang dan arah pengembangan evaluasi Bahasa Arab berbasis teknologi AI di masa depan.

Kajian ini diharapkan dapat menjadi rujukan teoretis bagi para pendidik, peneliti, dan pembuat kebijakan dalam merancang evaluasi pembelajaran Bahasa Arab yang lebih modern, objektif, efisien, dan selaras dengan perkembangan teknologi pendidikan global.

TINJAUAN PUSTAKA

Evaluasi Pembelajaran Bahasa Arab

Evaluasi pembelajaran merupakan bagian integral dari proses pendidikan yang berfungsi untuk mengukur ketercapaian tujuan instruksional, menilai perkembangan kemampuan peserta didik, serta memberikan umpan balik bagi guru dalam meningkatkan kualitas pembelajaran. Dalam konteks

Pendidikan Bahasa Arab, evaluasi memiliki kompleksitas tersendiri karena terkait dengan karakteristik linguistik bahasa Arab yang khas, seperti sistem fonologi yang variatif, struktur morfologi berbasis akar-kata (root-pattern system), serta sintaksis yang fleksibel namun padat makna (Al-Tamimi, 2023; Taha & Al-Nasser, 2022).

Evaluasi pembelajaran bahasa Arab meliputi penilaian empat keterampilan utama, yaitu *istimā'* (listening), *kalām* (speaking), *qirā'ah* (reading), dan *kitābah* (writing). Setiap keterampilan memiliki indikator dan instrumen penilaian yang berbeda. Misalnya, penilaian *istimā'* berfokus pada kemampuan memahami informasi lisan dan membedakan unsur fonetik, sedangkan *kalām* menilai kefasihan, ketepatan fonem, struktur sintaksis, dan kemampuan merespons secara komunikatif. Adapun *qirā'ah* mengukur pemahaman literal hingga inferensial terhadap teks Arab, sementara *kitābah* menilai struktur kalimat, kosa kata, koherensi ide, serta keakuratan gramatikal.

Dalam praktik tradisional, evaluasi keterampilan tersebut umumnya dilakukan secara manual melalui tes tertulis, tes lisan, portofolio, dan observasi. Meskipun metode ini telah digunakan secara luas, berbagai penelitian menunjukkan bahwa evaluasi manual memiliki beberapa keterbatasan, antara lain subjektivitas penilaian, variabilitas antar-pengajar dalam memberi skor, keterbatasan waktu, serta kurangnya kemampuan mengevaluasi aspek mikrolinguistik secara mendalam (Al-Ajmi & Al-Mutairi, 2024). Selain itu, penilaian keterampilan produktif seperti *kalām* dan *kitābah* sering kali tidak mampu mencakup keseluruhan performa peserta didik secara komprehensif karena keterbatasan instrumen dan sumber daya manusia.

Dengan munculnya teknologi digital, kebutuhan terhadap model evaluasi yang lebih objektif, adaptif, dan otomatis semakin mendesak. Hal ini menjadi relevan mengingat pembelajaran bahasa Arab di berbagai lembaga edukasi—termasuk madrasah, pesantren, dan perguruan tinggi—mengarah pada pendekatan komunikatif dan integratif yang menuntut instrumen evaluasi lebih tepat dan responsif terhadap perkembangan teknologi.

Teknologi Deep Learning dalam Pendidikan Bahasa

Deep Learning merupakan cabang dari machine learning yang menggunakan arsitektur jaringan saraf berlapis-lapis (deep neural networks) untuk mempelajari pola kompleks dalam data. Teknologi ini menjadi fondasi berbagai aplikasi modern seperti pengenalan wajah, pengolahan bahasa alami (NLP), dan sistem rekomendasi, termasuk dalam bidang pendidikan bahasa.

Dalam beberapa tahun terakhir, Deep Learning menunjukkan performa luar biasa dalam berbagai tugas linguistik, mulai dari speech recognition, text classification, machine translation, hingga automated scoring (Hussein & Al-Khalil, 2025). Hal ini disebabkan oleh kemampuan jaringan saraf dalam mengolah data berjumlah besar serta mengekstraksi fitur-fitur penting secara otomatis tanpa intervensi manusia secara langsung.

Dalam pendidikan 3ahasa Arab, penerapan Deep Learning mencakup beberapa ranah berikut:

1. Speech Recognition untuk Istimā' dan Kalām

Teknologi Automatic Speech Recognition (ASR) berbasis CNN, LSTM, dan Transformer telah digunakan untuk:

- a. mendeteksi fonem Arab (termasuk konsonan berdiftong dan huruf bersifat emphatic),
- b. menilai akurasi pelafalan,
- c. menganalisis prosodi dan intonasi,
- d. memberikan real-time feedback kepada pembelajar.

Penerapan ini sangat bermanfaat dalam evaluasi 3alam, khususnya untuk non-native learners yang sering melakukan kesalahan artikulasi.

2. Natural Language Processing (NLP) untuk Qirā'ah dan Kitābah

Arabic NLP memanfaatkan algoritma Deep Learning untuk memproses struktur kompleks teks Arab. Aplikasi yang relevan di antaranya:

- a. Automated Reading Assessment
- b. Arabic Morphological Analysis
- c. Syntax Parsing
- d. Semantic Similarity Detection
- e. Grammar Error Detection
- f. Automated Essay Scoring (AES)

Model seperti BERT, AraBERT, dan GPT-based architectures telah menghasilkan performa tinggi dalam tugas klasifikasi teks dan analisis Bahasa Arab.

3. OCR dan Handwriting Recognition

Evaluasi kemampuan kitābah juga terbantu melalui sistem pengenalan tulisan tangan Arab (Arabic Handwriting Recognition), terutama untuk konteks madrasah yang masih menggunakan tulisan manual. Model CNN dan Hybrid CNN-RNN terbukti efektif dalam membaca karakter Arab bertanda harakat maupun tanpa harakat.

4. Adaptive Testing dan Learning Analytics

Deep Learning memungkinkan penerapan:

- a. adaptive testing (tes menyesuaikan tingkat kemampuan)
- b. predictive analytics untuk performa belajar
- c. analisis kecepatan membaca (reading speed analytics)
- d. deteksi kesulitan belajar berbasis pola error linguistik

Hal ini membuka peluang besar bagi 4ahasa4 Pendidikan Bahasa Arab untuk meningkatkan kualitas evaluasi dan personalisasi pembelajaran. Di sisi lain, perkembangan teori Deep Learning juga semakin matang. Bengio, Goodfellow, dan Courville (2021) menjelaskan bahwa kemajuan arsitektur neural modern seperti Transformer dan attention mechanism memberikan peluang baru dalam pemrosesan 4ahasa alami. Sementara itu, Mubarak dan Abdul-Mageed (2021) memetakan perkembangan korpus digital Arab dan menekankan pentingnya dataset besar dalam meningkatkan performa model 4ahasa. Dalam pendidikan, Sabbah dan Omar (2024) menunjukkan bahwa intelligent tutoring systems berbasis AI mampu meningkatkan efektivitas pembelajaran nahwu melalui umpan balik otomatis yang terpersonalisasi

Penelitian Terkait (State of the Art)

Berbagai penelitian internasional menunjukkan tren peningkatan minat terhadap penggunaan Deep Learning dalam evaluasi bahasa. Studi oleh Zhao et al. (2024) menemukan bahwa integrasi model Transformer mampu meningkatkan akurasi automated speaking assessment hingga lebih dari 90%. Sementara itu, riset oleh Al-Harthy & Osman (2023) menunjukkan bahwa Arabic grammar error detection berbasis LSTM memiliki tingkat keberhasilan lebih tinggi dibandingkan analisis tradisional berbasis aturan (rule-based analysis).

Penelitian di bidang Arabic NLP juga berkembang pesat melalui pengembangan korpus modern seperti ArTenTen, Tashkeela, dan OpenITI, yang memungkinkan analisis frekuensi, kolokasi, dan struktur sintaksis dengan lebih sistematis. Namun demikian, penelitian tentang integrasi Deep Learning dalam evaluasi pembelajaran Bahasa Arab di konteks Indonesia masih sangat terbatas. Mayoritas riset hanya berfokus pada penggunaan media digital atau platform E-learning, bukan pada AI-based evaluation. Hal inilah yang menjadi celah (research gap) yang ingin dijembatani oleh artikel ini.

State of the art ini menegaskan bahwa penerapan Deep Learning untuk evaluasi bahasa Arab bukan hanya relevan, tetapi juga menawarkan potensi transformasi besar dalam sistem penilaian pendidikan bahasa di era digital.

METODE PENELITIAN

Penelitian ini menggunakan pendekatan kualitatif dengan metode studi pustaka (library research). Pendekatan ini dipilih karena tujuan utama penelitian adalah melakukan eksplorasi mendalam terhadap konsep, model, dan implementasi Deep Learning dalam evaluasi pembelajaran Bahasa Arab berdasarkan temuan-temuan teoretis dan hasil riset terdahulu. Studi pustaka memungkinkan peneliti menelusuri data ilmiah secara komprehensif, terstruktur, dan kritis (Zed, 2014). Sumber data dalam penelitian ini berasal dari dokumen ilmiah yang relevan, meliputi: 1. Artikel jurnal internasional bereputasi (Scopus, WoS) terkait: Deep Learning, Artificial Intelligence dalam pendidikan, Natural Language Processing (NLP), Arabic Speech Recognition, Automated writing/speaking assessment. 2. Prosiding konferensi internasional bidang AI, computational linguistics, dan language assessment. 3. Buku akademik mengenai teori evaluasi pembelajaran, pengajaran bahasa Arab, dan teknologi pendidikan (EdTech). 4. Korpus online dan studi empiris tentang Arabic NLP seperti ArTenTen, OpenITI, Tashkeela, dan QAWiD. 5. Kebijakan pendidikan nasional terkait digitalisasi pembelajaran. Pemilihan sumber dilakukan dengan mempertimbangkan relevance, recency (2020–2025), dan academic credibility.

Teknik pengumpulan data dilakukan melalui: Penelusuran literatur ilmiah menggunakan database seperti Scopus, Semantic Scholar, Google Scholar, dan IEEE Xplore. Seleksi artikel berdasarkan kata kunci seperti Deep Learning, Arabic NLP, AI-based assessment, automated scoring, dan Arabic speech recognition. Ekstraksi informasi dari artikel berupa teori, model algoritma, aplikasi, temuan, dan tantangan implementasi. Pemetaan tematik sesuai kebutuhan penelitian. Semua data yang terkumpul kemudian dikategorikan berdasarkan tema inti: evaluasi bahasa, teknologi Deep Learning, dan implementasi pada bahasa Arab.

Teknik Analisis Data Penelitian ini menggunakan analisis isi (content analysis) dengan langkah-langkah berikut: 1. Reduksi Data yaitu menyeleksi informasi relevan dari berbagai sumber dan membuang data yang tidak terkait langsung dengan evaluasi pembelajaran Bahasa Arab berbasis Deep Learning. 2. Klasifikasi Tematik yaitu mengelompokkan temuan literatur ke dalam kategori: Model-model Deep Learning (CNN, RNN, LSTM, Transformer), Aplikasi Arabic NLP, Evaluasi istīmā', kalām, qirā'ah, dan kitābah, Tantangan dan keterbatasan implementasi, Peluang pengembangan ke depan. 3. Analisis Kritis yaitu melakukan perbandingan temuan antar penelitian untuk menemukan: kekuatan dan kelemahan model, kesenjangan penelitian (research gap), relevansi model Deep Learning untuk konteks Indonesia. Penyusunan Sintesis yaitu mengintegrasikan semua temuan menjadi argumentasi utuh yang menjawab tujuan penelitian: kebutuhan evaluasi Bahasa Arab di era digital, model Deep Learning yang sesuai, tantangan implementasi, peluang pengembangan. Proses analisis dilakukan secara berulang dan reflektif untuk menghasilkan pemahaman komprehensif terhadap fenomena yang dikaji.

Keabsahan Data dijaga melalui: Triangulasi sumber (membandingkan berbagai jurnal dan temuan). Cross-check antar studi untuk menghindari bias tunggal. Pemilihan referensi terbaru (2020–2025) untuk menjaga relevansi. Analisis kritis terhadap metodologi dan konteks penelitian terdahulu. Pendekatan ini memastikan bahwa simpulan penelitian memiliki dasar empiris yang kuat dan dapat dipertanggungjawabkan secara ilmiah.

HASIL DAN PEMBAHASAN

Kebutuhan Evaluasi Pembelajaran Bahasa Arab di Era Digital

Evaluasi pembelajaran Bahasa Arab mengalami tantangan besar dalam konteks pendidikan modern, terutama terkait reliabilitas penilaian, objektivitas skor, dan efektivitas waktu. Penilaian konvensional yang dilakukan secara manual oleh pendidik menghadapi berbagai keterbatasan. Misalnya, penilaian *istimā'* dan *kalām* sering kali dipengaruhi oleh subjektivitas penilai, tingkat kelelahan, dan perbedaan interpretasi terhadap rubrik. Dalam konteks *qirā'ah* dan *kitābah*, penilai kerap kesulitan menganalisis aspek mikrolinguistik seperti morfologi, distribusi harakat, kesalahan sintaksis, serta kohesi dan koherensi teks yang ditulis peserta didik.

Di era digital, pendidik dituntut menyajikan penilaian yang lebih cepat, konsisten, akurat, dan mampu menangkap performa peserta didik secara mendalam. Hal ini tidak dapat sepenuhnya dicapai melalui metode tradisional. Oleh karena itu, kebutuhan terhadap sistem evaluasi berbasis teknologi — khususnya yang menggunakan Deep Learning — menjadi semakin penting untuk memperkuat kualitas asesmen bahasa Arab, baik di tingkat madrasah, pesantren, maupun perguruan tinggi.

Sejumlah penelitian menunjukkan bahwa teknologi AI mampu memperbaiki kualitas penilaian melalui *automated scoring*, *speech recognition*, analisis semantik, dan deteksi kesalahan gramatikal secara otomatis (Al-Harthy & Osman, 2023). Dengan demikian, teknologi Deep Learning tidak hanya menjadi pelengkap evaluasi, tetapi juga berpotensi menjadi paradigma baru dalam penilaian keterampilan berbahasa.

Model-Model Deep Learning untuk Evaluasi Bahasa Arab

Deep Learning menyediakan berbagai arsitektur jaringan saraf yang efektif untuk menganalisis data bahasa Arab, baik teks maupun suara. Model-model berikut telah banyak digunakan dalam aplikasi evaluasi linguistik modern.

1. Evaluasi *Istimā'* (Listening): Speech Recognition

Evaluasi keterampilan *istimā'* membutuhkan kemampuan sistem untuk mengidentifikasi fonem, melacak perbedaan intonasi, dan memahami ujaran dalam variasi dialek. Model Deep Learning yang relevan antara lain:

- a. CNN (Convolutional Neural Networks) untuk ekstraksi fitur akustik.
- b. RNN/LSTM untuk memproses rangkaian fonetik yang bersifat temporal.
- c. Transformer-based ASR (seperti Whisper dan SpeechT5) untuk menangani variasi suara yang kompleks.

Dalam konteks Bahasa Arab, pengenalan suara menjadi lebih menantang karena keberadaan huruf-huruf *emphatic*, *qalqalah*, variasi vokal, dan dialek regional. Namun penelitian terbaru menunjukkan bahwa akurasi ASR Arab telah meningkat pesat melalui penggunaan dataset besar seperti MGB-3, QASR, dan Arabic Speech Corpus. Sistem ini dapat menjadi instrumen evaluasi otomatis untuk tugas *listening comprehension* maupun *phonetic accuracy*.

2. Evaluasi *Kalām* (Speaking): Automatic Speech Assessment

Deep Learning memungkinkan penilaian keterampilan berbicara secara otomatis melalui:

- a. analisis pronunciation (kesesuaian fonem dan artikulasi),
- b. evaluasi fluency (kecepatan, kelancaran, pause),
- c. deteksi prosody (intonasi, tekanan),
- d. penilaian coherence (keterkaitan ide dalam ujaran).

Model seperti Deep Speech, wav2vec 2.0, dan HuBERT digunakan untuk mengukur kualitas produksi suara pembelajar. Dalam evaluasi Bahasa Arab, teknologi ini dapat membantu dosen menilai *kalām* tanpa harus melakukan tes lisan secara berulang.

3. Evaluasi Qirā'ah (Reading): Teks dan Pemahaman

Untuk keterampilan qirā'ah, Deep Learning digunakan dalam:

- a. analisis reading comprehension,
- b. klasifikasi pemahaman literal, inferensial, hingga evaluatif,
- c. deteksi kesalahan dalam membaca teks Arab berharakat,
- d. pemetaan hubungan antaride melalui semantic similarity.

Model AraBERT, Arabic GPT, dan Transformer encoder terbukti mampu memproses teks Arab dengan tingkat akurasi tinggi dalam tugas pemahaman teks. Sistem ini dapat membantu mengkategorikan kemampuan membaca mahasiswa berdasarkan pola jawaban yang dianalisis oleh model.

Selain itu, studi oleh El-Hajj dan Abu-Ali (2023) menunjukkan bahwa model transformer dapat digunakan untuk mengevaluasi kualitas argumentasi dan koherensi teks secara otomatis dalam tulisan akademik bahasa Arab. Penelitian ini memperkuat potensi automated essay scoring (AES) berbasis Deep Learning sebagai solusi modern untuk penilaian kitābah. Mekheimer (2025) bahkan menambahkan bahwa integrasi computational linguistics ke dalam pedagogi bahasa Arab semakin penting untuk memastikan evaluasi yang akurat dan relevan dengan perkembangan teknologi.

4. Evaluasi Kitābah (Writing): Grammar Checker & Essay Scoring

Evaluasi keterampilan menulis adalah bagian yang paling kompleks dalam pembelajaran Bahasa Arab. Deep Learning mampu mengotomatisasi proses ini melalui:

- a. deteksi kesalahan gramatikal (i'rāb, struktur kalimat, pola morfologis),
- b. analisis kesalahan kosakata dan pemilihan kata,
- c. penilaian kohesi dan koherensi teks,
- d. penilaian kualitas argumentasi,
- e. automated essay scoring berbasis Transformer.

Model seq2seq, Bi-LSTM, dan Transformers telah berhasil digunakan dalam Arabic grammar correction dan analisis kesalahan tulisan tangan. Dengan demikian, Deep Learning berpotensi menjadi alat evaluasi yang mampu menilai tulisan peserta didik secara cepat, objektif, dan terukur.

Keunggulan Deep Learning dalam Evaluasi Bahasa Arab

Integrasi Deep Learning dalam evaluasi pembelajaran Bahasa Arab menawarkan sejumlah keunggulan, antara lain:

1. Objektivitas Penilaian, Deep Learning menilai berdasarkan algoritma yang konsisten, bukan persepsi manusia, sehingga mengurangi subjektivitas dan meningkatkan reliabilitas penilaian.
2. Efisiensi Waktu dan Sumber Daya, Penilaian otomatis menghemat waktu pendidik terutama dalam tugas-tugas berat seperti koreksi tulisan atau penilaian ujian lisan.
3. Ketepatan Analisis Mikrolinguistik, Deep Learning mampu mendeteksi kesalahan kecil dalam fonem, morfologi, dan sintaksis yang sering terlewat oleh penilai manusia.
4. Adaptive and Personalized Assessment, Model dapat menyesuaikan tingkat soal sesuai kemampuan peserta didik, memungkinkan evaluasi lebih adil dan akurat.
5. Real-time Feedback, Mahasiswa dapat segera mengetahui kesalahan mereka dan memperbaikinya tanpa menunggu proses penilaian manual.

Keunggulan-keunggulan tersebut menjadikan Deep Learning sebagai pendekatan potensial dalam menghadirkan sistem evaluasi modern yang sejalan dengan tuntutan pembelajaran abad ke-21.

Keunggulan lainnya adalah kemampuan Deep Learning untuk menganalisis teks secara semantik. Younes dan Habash (2022) menunjukkan bahwa model Arab NLP modern kini mampu membaca struktur semantik dan melakukan penilaian kesepadanan makna, sebuah aspek penting dalam evaluasi

qirā'ah dan kitābah. Hal ini memberikan peluang bagi pengajar untuk memahami performa siswa secara lebih mendalam berdasarkan aspek makna, bukan sekadar kesalahan struktural.

Tantangan Implementasi di Indonesia

Meskipun potensinya besar, implementasi Deep Learning dalam evaluasi pembelajaran Bahasa Arab di Indonesia menghadapi beberapa kendala, antara lain:

1. Keterbatasan Korpus dan Dataset Bahasa Arab, Bahasa Arab memiliki keragaman dialek dan bentuk tulisan yang kompleks. Dataset berkualitas tinggi masih terbatas, terutama untuk konteks pembelajar non-native seperti Indonesia.
2. Kurangnya Literasi Digital Dosen dan Guru, Integrasi teknologi membutuhkan kesiapan kompetensi digital pendidik, termasuk cara mengakses, mengelola, dan menginterpretasi output AI. Hal ini masih menjadi tantangan signifikan di banyak lembaga pendidikan.
3. Infrastruktur Teknologi yang Belum Merata, Deep Learning membutuhkan perangkat keras tertentu seperti GPU dan koneksi internet yang stabil. Tidak semua institusi pendidikan memiliki fasilitas tersebut.
4. Risiko Bias Algoritmik, Model AI yang dilatih dengan dataset terbatas dapat menghasilkan bias dalam menilai dialek, intonasi, atau gaya menulis tertentu.
5. Kesiapan Regulasi dan Etika, Belum adanya pedoman nasional terkait pemanfaatan AI dalam evaluasi pembelajaran, termasuk isu privasi data, keamanan, dan integritas akademik.

Tantangan-tantangan ini harus diantisipasi agar penerapan Deep Learning tidak menimbulkan kesenjangan baru dalam pendidikan. Selain itu, Wahyuni dan Prasetyo (2023) menegaskan bahwa transformasi digital dalam pendidikan Indonesia menghadapi tantangan kesenjangan infrastruktur dan kesiapan SDM, yang relevan pula dalam konteks penerapan evaluasi berbasis AI untuk pembelajaran bahasa Arab.

Peluang dan Arah Pengembangan Masa Depan

Meskipun memiliki kendala, prospek pengembangan evaluasi Bahasa Arab berbasis Deep Learning di Indonesia sangat terbuka. Beberapa peluang strategis yang dapat dimanfaatkan antara lain:

1. Pengembangan Smart Assessment di Perguruan Tinggi Islam, Prodi PBA dapat mengembangkan platform evaluasi otomatis untuk empat maharah, bekerja sama dengan fakultas sains data atau informatika.
2. Integrasi AI dalam Kurikulum PBA, Materi computational linguistics, Arabic NLP, dan assessment berbasis AI dapat diintegrasikan ke dalam mata kuliah teknologi pembelajaran.
3. Kolaborasi Penelitian antara Kampus dan Pesantren, Dataset linguistik dapat dikembangkan melalui kolaborasi antara kampus, pesantren, dan lembaga pendidikan bahasa.
4. Adopsi AI Tools oleh Guru Madrasah, Guru-guru MI, MTs, MA dapat memanfaatkan aplikasi evaluasi berbasis AI untuk memperbaiki kualitas penilaian.
5. Perkembangan Kebijakan Nasional, Kementerian Agama dan Kemendikbud dapat mengeluarkan regulasi khusus terkait AI dalam evaluasi pembelajaran sehingga implementasinya lebih terarah. Dengan memanfaatkan peluang tersebut, Indonesia dapat berada pada garda depan dalam pengembangan evaluasi Bahasa Arab berbasis Deep Learning di Asia Tenggara. Di sisi lain, Zaid dan Al-Hassan (2025) menegaskan bahwa penilaian berbasis Deep Learning dapat menjadi fondasi bagi kebijakan transformasi pendidikan bahasa di kawasan Asia Barat dan Asia Tenggara. Integrasi ini diyakini mampu menghasilkan transformative assessment yang lebih adil, adaptif, dan berbasis bukti (evidence-based evaluation).

SIMPULAN

Perkembangan teknologi kecerdasan buatan, khususnya Deep Learning, membuka peluang besar dalam meningkatkan kualitas evaluasi pembelajaran Bahasa Arab di Indonesia. Di tengah tantangan era digital yang menuntut proses evaluasi yang lebih objektif, cepat, akurat, dan berbasis data, model-model Deep Learning hadir sebagai solusi inovatif yang mampu menjawab berbagai keterbatasan evaluasi tradisional, terutama dalam menilai keterampilan produktif seperti *istimā'*, *kalām*, *qirā'ah*, dan *kitābah*.

Hasil kajian menunjukkan bahwa arsitektur Deep Learning—seperti CNN, RNN, LSTM, dan Transformer—telah terbukti efektif dalam mengotomatisasi evaluasi bahasa melalui speech recognition, analisis fonetik, deteksi kesalahan gramatikal, pemahaman teks otomatis, serta automated essay scoring. Penerapan teknologi ini berpotensi meningkatkan reliabilitas penilaian, memberikan real-time feedback, mendukung pembelajaran adaptif, serta menghemat waktu dan beban kerja pendidik.

Meski demikian, implementasi Deep Learning dalam evaluasi pembelajaran Bahasa Arab di Indonesia masih menghadapi sejumlah tantangan. Keterbatasan dataset bahasa Arab, kesiapan kompetensi digital pendidik, infrastruktur teknologi yang belum merata, serta isu bias algoritmik dan etika penggunaan AI menjadi faktor-faktor yang perlu diperhatikan secara serius. Tantangan tersebut menuntut adanya strategi kolaboratif antara perguruan tinggi, lembaga pendidikan Islam, pemerintah, dan peneliti untuk mengembangkan sistem evaluasi yang lebih kuat, inklusif, dan kontekstual.

Artikel ini menyimpulkan bahwa integrasi Deep Learning dalam evaluasi pembelajaran Bahasa Arab bukan hanya relevan secara teknologis, tetapi juga strategis untuk memperkuat transformasi pendidikan di Indonesia. Ke depan, pengembangan smart assessment, perluasan korpus bahasa Arab modern, peningkatan literasi digital pendidik, serta penerbitan kebijakan nasional terkait AI dalam pendidikan menjadi arah pengembangan yang sangat penting. Dengan langkah ini, evaluasi Bahasa Arab dapat bergerak menuju sistem yang lebih modern, adaptif, dan mampu mendukung pembelajaran bahasa yang berkualitas di era digital.

DAFTAR PUSTAKA

- Al-Ajmi, M., & Al-Mutairi, F. (2024). Challenges of traditional assessment in Arabic language classrooms: A pedagogical review. *Journal of Arabic Linguistics and Education*, 12(1), 44–59.
- Al-Harthy, S., & Osman, M. (2023). Arabic grammar error detection using deep neural networks. *Natural Language Engineering*, 29(4), 611–629.
- Al-Tamimi, R. (2023). Arabic morphological complexity and its implications for second language assessment. *Arabic Linguistics Review*, 7(2), 201–225.
- Bengio, Y., Goodfellow, I., & Courville, A. (2021). *Deep Learning* (2nd ed.). MIT Press.
- El-Hajj, W., & Abu-Ali, N. (2023). Automatic essay scoring for Arabic using transformer-based models. *Computer Speech & Language*, 82, 101522.
- Hussein, A., & Al-Khalil, M. (2025). Deep Learning in Arabic language education: Applications, limitations, and future directions. *Journal of Applied Linguistics and AI Technologies*, 5(1), 15–36.
- Mekheimer, M. (2025). Arabic computational linguistics and machine learning in language pedagogy. *Journal of Arabic and Applied Linguistics*, 9(1), 55–78.
- Mubarak, H., & Abdul-Mageed, M. (2021). Arabic NLP advances: Corpora, models, and resources. *ACL Anthology*, 1–16.
- Osman, T., Khalil, H., Miltan, M., Shaalan, K., & Alfrjani, R. (2023). Exploiting Functional Discourse Grammar to Enhance Complex Arabic Relation Extraction using a Hybrid Semantic Knowledge Base—Machine Learning Approach. *ACM Transactions on Asian and Low-Resource Language Information Processing*, 22(8), 1–30. <https://doi.org/10.1145/3610581>

- Sabbah, S., & Omar, A. (2024). Intelligent tutoring systems for Arabic grammar learning. *Interactive Learning Environments*, 32(4), 587–603.
- Taha, M., & Al-Nasser, F. (2022). Problems of evaluating Arabic writing skills: A review of modern approaches. *Arab World English Journal*, 13(4), 251–269.
- Wahyuni, S., Gema Febriansyah, S. S. K., & Jasrial, D. (2023). *Metodologi Pendidikan Bahasa. Cendikia Mulia Mandiri*.
- Younes, M., & Habash, N. (2022). Arabic text analytics and implications for language teaching. *Language Resources and Evaluation*, 56(4), 1123–1145.
- Zaid, S., & Al-Hassan, R. (2025). Transformative assessment in language education using Deep Learning. *Journal of Modern Educational Technologies*, 8(1), 77–98.
- Zawacki-Richter, O. & others. (2023). Systematic review of AI and Deep Learning in education. *Educational Research Review*, 38, 100512.
- Zed, M. (2014). *Metode Penelitian Kepustakaan*. Yayasan Pustaka Obor Indonesia.
- Zhao, W., Liu, Q., & Zhang, Y. (2024). Transformer-based models for automated speaking assessment. *Artificial Intelligence in Education*, 35(1), 45–62.