



## Assessment and evaluation of an artificial intelligence-based deep learning curriculum

Aam Hamdani<sup>1</sup>, Amay Suherman<sup>2</sup>, Bambang Darmawan<sup>3</sup>, Enda Permana<sup>4</sup>

<sup>1,2,3,4</sup>Universitas Pendidikan Indonesia, Kota Bandung, Indonesia

[aam\\_hamdani@upi.edu](mailto:aam_hamdani@upi.edu)<sup>1</sup>, [a\\_suherman@upi.edu](mailto:a_suherman@upi.edu)<sup>2</sup>, [bambang\\_darmawan@upi.edu](mailto:bambang_darmawan@upi.edu)<sup>3</sup>, [enda\\_per@upi.edu](mailto:enda_per@upi.edu)<sup>4</sup>

### ABSTRACT

The rapid development of Artificial Intelligence (AI) and the implementation of deep learning curricula in education have necessitated reforms to learning assessment and evaluation. This article examines the conceptual framework and assessment practices appropriate for AI-based deep learning curricula: how assessments can measure critical thinking, collaboration, knowledge transfer, and data literacy competencies; how AI-based platforms enable personalization, automated assessment, and learning analytics; and the ethical and validity challenges of using AI for evaluation. The method used is a systematic literature review and a comparative analysis of current practices in school and higher education contexts. The results indicate that AI-based assessments can improve learning effectiveness by providing real-time feedback and progress tracking; however, their success depends heavily on authentic task design, teacher readiness, infrastructure, and ethical/data security arrangements. Recommendations are provided for educational institutions to adopt a combination of authentic assessments (portfolios, projects, observations) and AI metrics, and to conduct regular algorithm audits to ensure the validity, reliability, and fairness of assessments.

### ARTICLE INFO

#### Article History:

Received: 12 Aug 2025

Revised: 16 Nov 2025

Accepted: 23 Nov 2025

Publish online: 27 Dec 2025

#### Keywords:

artificial intelligence;  
assessment; authentic  
assessment; deep learning  
curriculum; learning  
evaluation; personalized  
learning

#### Open access

Curricula: Journal of Curriculum  
Development is a peer-reviewed  
open-access journal.

### ABSTRAK

Perkembangan pesat di bidang Artificial Intelligence (AI) dan penerapan kurikulum pembelajaran mendalam (deep learning) dalam ranah pendidikan memunculkan kebutuhan untuk mereformasi asesmen dan evaluasi pembelajaran. Artikel ini mengkaji kerangka konseptual dan praktik asesmen yang sesuai untuk kurikulum deep-learning berbasis AI: bagaimana asesmen dapat mengukur kompetensi berpikir kritis, kolaborasi, transfer pengetahuan, dan literasi data; bagaimana platform berbasis AI memungkinkan personalisasi, penilaian otomatis, serta analitik pembelajaran; serta tantangan-etika dan validitas dalam penggunaan AI untuk evaluasi. Metode yang digunakan adalah studi literatur sistematis dan analisis komparatif terhadap praktik terkini dalam konteks sekolah dan perguruan tinggi. Hasil menunjukkan bahwa asesmen berbasis AI dapat meningkatkan efektivitas pembelajaran dengan menyediakan umpan balik real-time dan pelacakan kemajuan, namun keberhasilannya sangat bergantung pada desain tugas otentik, kesiapan guru, infrastruktur, dan pengaturan etika/keamanan data. Rekomendasi diberikan agar institusi pendidikan mengadopsi kombinasi penilaian otentik (portofolio, proyek, observasi) dengan metrik AI, serta melakukan audit algoritma secara berkala untuk memelihara validitas, reliabilitas, dan keadilan penilaian.

**Kata Kunci:** asesmen; asesmen otentik; evaluasi pembelajaran; kurikulum deep learning; kecerdasan buatan; personalisasi pembelajaran

### How to cite (APA 7)

Hamdani, A., Suherman, A., Darmawan, B., & Permana, E. (2025). Assessment and evaluation of an artificial intelligence-based deep learning curriculum. *Curricula: Journal of Curriculum Development*, 4(2), 1797-1810.

### Peer review

This article has been peer-reviewed through the journal's standard double-blind peer review, where both the reviewers and authors are anonymised during review.

### Copyright

2025, Aam Hamdani, Amay Suherman, Bambang Darmawan, Enda Permana. This an open-access is article distributed under the terms of the Creative Commons Attribution-ShareAlike 4.0 International (CC BY-SA 4.0) <https://creativecommons.org/licenses/by-sa/4.0/>, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author, and source are credited. \*Corresponding author: [aam\\_hamdani@upi.edu](mailto:aam_hamdani@upi.edu)

## INTRODUCTION

Dalam era kompetensi abad ke-21, pendidikan tidak hanya menekankan penguasaan fakta atau prosedur semata, tetapi juga kemampuan murid untuk berpikir kritis, memecahkan masalah kompleks, bekerja dalam tim, dan mentransfer pengetahuan ke situasi baru. Istilah “*deep learning*” muncul sebagai sebuah konsep yang kemudian diimplementasikan dalam bentuk kurikulum dan dioperasionalkan sebagai pendekatan pedagogis. Pada penelitian ini menempatkan *deep learning* sebagai pendekatan pedagogis yang bertujuan menciptakan pembelajaran lebih bermakna dan kritis serta menghasilkan solusi baru dengan wawasan kreatif terhadap permasalahan di lingkungan. Konsep ini tidak hanya berfokus pada penguasaan materi, tetapi juga pemahaman konseptual mendalam, hubungan antar ide, penerapan dalam berbagai konteks, dan refleksi diri. Konsep ini menjadi semakin relevan untuk murid mampu berpikir lebih kritis dan mendalam, berkolaborasi secara aktif, dan memecahkan masalah secara bermakna (Halim, 2025; Kasi et al., 2025; Nafi’ah & Faruq, 2025).

Tantangan ini semakin menonjol karena proses pembelajaran di sekolah cenderung bersifat dangkal (*surface learning*), asesmen didominasi tes tertulis, dan guru belum memiliki alat evaluasi yang memadai untuk memetakan kompetensi tingkat tinggi. Kondisi ini berdampak langsung terhadap kualitas pembelajaran murid pada kemampuan berpikir kritis dan keterampilan pemecahan masalah tidak tergalikan secara optimal, transfer pengetahuan ke konteks baru menjadi terbatas, motivasi belajar menurun, serta kesiapan lulusan untuk memenuhi tuntutan dunia kerja dan industri menjadi kurang. Oleh karena itu, pembelajaran dan asesmen ideal seharusnya bersifat personal, adaptif, dan otentik sehingga mampu menangkap kemampuan berpikir tingkat tinggi secara komprehensif (Putri et al., 2025).

Integrasi teknologi *Artificial Intelligence* (AI) dalam proses pembelajaran turut membuka peluang penting untuk menjembatani kesenjangan tersebut. Teknologi AI dapat mendukung personalisasi pembelajaran, adaptasi konten secara dinamis, dan otomatisasi umpan balik bagi murid dan guru (Yuliansyah & Saidah, 2025). Misalnya, sistem manajemen pembelajaran berbasis AI dapat memantau jejak belajar murid, memberikan rekomendasi materi lanjutan, atau mengidentifikasi area kesulitan yang dialami murid. Integrasi sistem manajemen berbasis AI dalam penerapan kurikulum *deep learning* di sekolah dasar menghasilkan peningkatan rata-rata skor N-Gain sebesar 58,13%, dibandingkan kelompok kontrol 37,48% (Ali et al., 2025). Angka N-Gain tersebut menunjukkan efektivitas intervensi berbasis AI dalam meningkatkan pemahaman konseptual murid, termasuk pada kemampuan analitis dan *problem solving*.

Namun demikian, penerapan asesmen dalam kerangka kurikulum yang mengadopsi prinsip-prinsip *deep learning* berbasis AI membawa tantangan yang cukup besar. Asesmen tradisional yang bersifat ujian tertulis tidak cukup untuk mengukur kompetensi tingkat tinggi seperti transfer pengetahuan, kolaborasi, dan kreativitas. Namun, penggunaan AI dalam evaluasi juga dapat menambahkan lapisan kompleksitas, termasuk masalah validitas dan reliabilitas, bias algoritma, keadilan (*fairness*), serta privasi dan etika data (Bulut et al., 2024). Isu-isu ini telah disorot dalam berbagai kajian, termasuk integrasi antara teori validitas asesmen dan kerangka *responsible AI*, serta faktor keandalan dalam pengumpulan data yang memengaruhi performa dan keadilan sistem AI (Burstein & LaFlair, 2024; Inel et

*al.*, 2023). Oleh karena itu, penting untuk merancang kerangka asesmen yang holistik—yang menggabungkan tugas otentik, analitik pembelajaran berbasis AI, dan pengawasan ahli—agar evaluasi tidak hanya menjadi alat pengukuran akhir, tetapi juga menjadi bagian integral dari pembelajaran mendalam.

Berdasarkan permasalahan tersebut, artikel ini menawarkan solusi berupa analisis komprehensif mengenai rancangan dan implementasi asesmen dan evaluasi pembelajaran dalam kurikulum *deep-learning* berbasis AI. Dalam konteks studi literatur ini, variabel utama yang dibahas mencakup konsep asesmen, evaluasi pembelajaran, kurikulum *deep learning*, AI, serta asesmen otentik dan personalisasi pembelajaran. Analisis literatur menunjukkan bahwa banyak kajian dan praktik pendidikan saat ini masih menghadapi kendala, seperti asesmen yang terlalu berfokus pada tes tertulis, keterbatasan instrumen yang mampu mengukur kemampuan berpikir tingkat tinggi, serta belum optimalnya pemanfaatan AI dalam mendukung proses evaluasi pembelajaran (Bulut *et al.*, 2024; Chinta *et al.*, 2025; ; Yan *et al.*, 2023). Padahal, secara ideal, kurikulum *deep learning* menuntut asesmen yang autentik, adaptif, dan terintegrasi dengan analitik pembelajaran berbasis AI agar mampu menggambarkan capaian belajar secara komprehensif.

Studi ini mengembangkan kerangka konseptual yang mengintegrasikan prinsip pembelajaran mendalam, asesmen otentik, serta pemanfaatan AI dalam asesmen formatif dan sumatif, sekaligus menyoroti aspek validitas, reliabilitas, *fairness*, dan peran supervisi guru dalam menjaga kualitas evaluasi. Selain itu, penelitian ini mengkaji kesenjangan antara teori kurikulum *deep learning* dan praktik asesmen berbasis AI di berbagai studi terdahulu. Secara khusus, artikel ini akan: 1) Mengidentifikasi karakteristik asesmen yang sesuai untuk pembelajaran mendalam; 2) Mengeksplorasi peran AI dalam mendukung asesmen formatif dan sumatif; 3) Membahas tantangan validitas, reliabilitas, dan etika dalam asesmen berbasis AI; serta 4) Memberikan rekomendasi bagi praktisi pendidikan untuk merancang sistem evaluasi yang efektif, etis, dan relevan dengan kebutuhan pembelajaran abad ke-21.

## LITERATURE REVIEW

### ***Deep Learning Pedagogy***

*Deep learning pedagogy* menekankan pemahaman konseptual, berpikir tingkat tinggi, dan transfer pengetahuan melalui tugas otentik serta refleksi metakognitif dengan efektivitas bergantung pada desain tugas dan kesiapan pengajar (Sadykova *et al.*, 2024; Suherman *et al.*, 2025). Pembelajaran mendalam dalam penelitian ini dimaksudkan sebagai pendekatan pedagogis yang berakar pada sebuah konsep dan diimplementasikan melalui kurikulum yang mengadopsi prinsip-prinsip *deep learning*, yaitu penekanan pada pemahaman konseptual yang kuat, kemampuan berpikir tingkat tinggi, serta transfer pengetahuan ke berbagai konteks baru. Tidak seperti pembelajaran permukaan (*surface learning*) yang berfokus pada hafalan atau penyelesaian tugas secara prosedural, pembelajaran mendalam mengharuskan murid mengembangkan keterampilan analisis, sintesis, evaluasi, serta refleksi terhadap proses belajarnya.

Literatur internasional menjelaskan bahwa *deep learning* mencakup integrasi ide, elaborasi makna, dan pembangunan hubungan antar konsep sehingga pengetahuan menjadi lebih fleksibel dan dapat diterapkan secara adaptif (Chen & Singh, 2024). Kajian nasional juga

menguatkan bahwa pembelajaran mendalam relevan dengan tuntutan kurikulum modern yang menekankan kompetensi, penalaran, dan kemampuan memecahkan masalah dalam konteks nyata (Nafi'ah & Faruq, 2025).

Dari perspektif pedagogis, terdapat sejumlah karakteristik utama yang menandai pendekatan ini, antara lain penggunaan tugas otentik, pemecahan masalah berbasis konteks, pembelajaran berbasis proyek atau desain, serta stimulasi metakognisi untuk meningkatkan kesadaran belajar. Pendekatan *design-based learning* dapat mendorong terjadinya pembelajaran mendalam melalui eksplorasi, kolaborasi, dan refleksi atas proses desain yang dilakukan murid (Weng *et al.*, 2023).

## **Asesmen**

Asesmen sebagai pedagogi memosisikan tugas otentik, portofolio, dan penilaian formatif sebagai bagian integral pengajaran untuk memfasilitasi *deep learning*, dengan fokus pada umpan balik berkelanjutan dan penilaian proses yang memperbaiki pembelajaran murid (Sappaile, 2025). Selain itu, asesmen memegang peran sentral sebagai bagian dari pedagogi, bukan sekadar alat pengukuran hasil akhir. Konsep *assessment as pedagogy* menekankan asesmen yang berfungsi memberikan umpan balik berkelanjutan, menuntut performa otentik, serta memberikan ruang bagi murid untuk menunjukkan proses berpikirnya (Timperley & Schick, 2025). Selaras dengan itu, literatur menegaskan bahwa *deep learning* hanya dapat dicapai apabila asesmen, aktivitas belajar, dan tujuan pembelajaran dirancang secara koheren.

Asesmen yang mendukung *deep learning* menekankan otentisitas tugas, penilaian proses dan produk, serta penilaian formatif yang berkelanjutan. Tugas otentik (misalnya proyek, studi kasus, portofolio, performa) yang meniru konteks dunia nyata memaksa murid untuk mengintegrasikan, mengelaborasi, dan mentransfer konsep sehingga mendorong pemikiran tingkat tinggi dan refleksi metakognitif. Desain rubrik analitik, umpan balik berulang, dan instrumen yang mengukur proses (bukan hanya hasil akhir) adalah ciri penting karena memungkinkan pendidik menilai aspek kognitif tingkat tinggi seperti analisis, sintesis, dan pemecahan masalah (Baines *et al.* 2024; Banihashem *et al.*, 2025; Cahyani *et al.*, 2024). Selain itu, asesmen yang mendukung *deep learning* memerlukan keselarasan kurikulum, instruksi, asesmen, kapasitas pendidik untuk merancang tugas otentik, dan dukungan teknologi untuk personalisasi serta umpan balik *real-time*.

## **Artificial Intelligence (AI)**

*Artificial Intelligence* (AI) dalam asesmen (*automated scoring*, *LLM feedback*, *learning analytics*) menawarkan efisiensi dan personalisasi, tetapi menimbulkan isu validitas inferensi, bias/*fairness*, dan privasi sehingga memerlukan audit manusia dan tata kelola *responsible AI* (Aloisi, 2023). Salah satu potensi paling nyata dari AI dalam asesmen adalah mempercepat proses koreksi dan memberikan umpan balik otomatis pada skala besar, khususnya untuk tugas terbuka seperti esai dan soal pemrograman. Sistem *automated scoring* dan *AI-powered grading* mampu menurunkan beban kerja penilai sekaligus menghasilkan skor yang konsisten dan cepat sehingga waktu untuk intervensi pedagogis menjadi lebih singkat. Bukti kajian terkini menunjukkan bahwa *AI-powered grading*

meningkatkan efisiensi koreksi tanpa mengorbankan konsistensi, terutama bila dikombinasikan dengan mekanisme audit manusia untuk memastikan validitas penilaian (Deepshikha, 2025; Tan et al., 2025).

Selain itu, AI membuka peluang untuk *feedback real-time*, personalisasi, dan pemrosesan data besar (*learning analytics*) yang mendukung evaluasi formatif berkelanjutan. Dengan analitik pembelajaran, platform berbasis AI dapat memberi umpan balik segera dan *actionable*, memetakan pola kesulitan pelajar, serta menyesuaikan jalur pembelajaran secara adaptif sehingga intervensi menjadi lebih tepat sasaran. Studi-studi sistematis dan eksperimen menunjukkan bahwa kombinasi *adaptive testing*, *Intelligent Tutoring Systems* (ITS), dan *LLM-based support* dapat meningkatkan personalisasi dan efektivitas pembelajaran dengan catatan perlu adanya tata kelola data dan pengawasan kualitas algoritma. Contoh implementasi lokal juga menunjukkan peningkatan *learning gain* setelah adopsi sistem AI yang menyediakan umpan balik dan penyesuaian konten secara otomatis (Campos, 2025; Weidlich et al., 2025).

Penerapan AI dalam asesmen menimbulkan tantangan serius terkait validitas, reliabilitas, dan integritas inferensi penilaian: model AI kadang menilai aspek yang berbeda dari tujuan pembelajaran yang dimaksudkan, sehingga argumen validitas harus dibangun ulang untuk tiap sistem; selain itu, hasil scoring otomatis perlu diuji konsistensi dan ketahanan terhadap manipulasi (*robustness*). Hal ini telah didiskusikan luas dalam literatur yang menghubungkan prinsip validasi asesmen klasik dengan prinsip *responsible AI* yaitu menyarankan standar, audit, dan dokumentasi yang ketat agar skor AI dapat dipercaya (Burstein & LaFlair, 2024).

## METHODS

Penelitian ini menggunakan pendekatan kualitatif melalui metode studi literatur sistematis guna menganalisis secara mendalam konsep asesmen dan evaluasi pembelajaran dalam kurikulum *deep learning* berbasis kecerdasan buatan. Subjek penelitian terdiri atas berbagai sumber literatur berupa artikel jurnal, sedangkan objek penelitian adalah konsep, model, dan praktik asesmen, evaluasi pembelajaran, AI dalam personalisasi pembelajaran, serta implementasi kurikulum *deep learning* yang dibahas dalam sumber-sumber tersebut. Pemilihan literatur dilakukan berdasarkan kriteria relevansi, keterbaruan isu, kesesuaian konteks pendidikan, serta kontribusinya terhadap kajian asesmen dan kurikulum berbasis AI, sehingga literatur yang dianalisis benar-benar merepresentasikan perkembangan teoritis maupun praktik terkini. Indikator keberhasilan penelitian ini ditentukan dengan kemampuan penelitian untuk menggabungkan dan menyusun pemahaman menyeluruh dari berbagai literatur secara komprehensif sehingga menghasilkan gambaran yang jelas mengenai model asesmen dan evaluasi yang efektif dalam kurikulum *deep learning* berbasis AI. Sementara itu, alat ukur penelitian berupa teknik analisis isi (*content analysis*) dan analisis sitematik, yang digunakan untuk mengklasifikasikan temuan literatur dan mengidentifikasi pola-pola konseptual. Berikut adalah tahapan metodenya secara rinci.

### Pemilihan Literatur

Proses pencarian literatur dilakukan melalui database terbuka dan jurnal pendidikan/teknologi, dengan fokus publikasi tahun 2023-2025 untuk memastikan kekinian

penelitian. Kata kunci pencarian meliputi: “*deep learning curriculum*”, “*artificial intelligence assessment education*”, “*AI in educational measurement*”, “*deep learning pedagogy assessment*”, “*AI-based assessment tools*”. Dari hasil awal, sejumlah artikel ditinjau seperti 1) *Measurement and ethics* dalam AI (Bulut *et al.*, 2024); 2) Pendekatan *deep learning* di pendidikan dasar (Nafi’ah & Faruq, 2025); serta 3) Efektivitas *AI-based management systems*, dan lainnya sebagai dasar referensi terkini (Ali *et al.*, 2025) .

### **Kriteria Inklusi dan Eksklusi**

Terdapat beberapa kriteria inklusi yakni: 1) Penelitian yang secara eksplisit membahas asesmen atau evaluasi pembelajaran dalam konteks kurikulum *deep learning* atau AI; 2) Penelitian empiris atau tinjauan sistematis; 3) Publikasi dalam rentang 2023-2025; 4) Konteks pembelajaran formal (sekolah dasar, menengah, atau perguruan tinggi). Sedangkan untuk kriteria eksklusi yaitu: 1) Penelitian yang hanya fokus pada algoritma teknis AI tanpa kaitan dengan asesmen pembelajaran; 2) Publikasi sebelum 2023; 3) Konteks non-formal (misalnya pelatihan korporasi) yang tidak relevan langsung dengan kurikulum sekolah/kampus.

### **Proses Seleksi dan Ekstraksi Data**

Setelah pencarian, proses seleksi dilakukan dengan *screening* judul dan abstrak, kemudian *full-text* apabila relevan. Setiap artikel yang lolos dianalisis untuk ekstraksi data: jenis asesmen (formatif, sumatif, portofolio, proyek, kolaboratif), peran AI (umpan balik otomatis, adaptasi jalur belajar, analitik murid), *outcome* pembelajaran (misalnya peningkatan skor, keterlibatan murid, transfer pengetahuan), serta isu validitas, reliabilitas, etika, kesiapan guru dan infrastruktur. Data diekstraksi ke *Spreadsheet* untuk dianalisis komparatif.

### **Analisis Komparatif dan Sintesis**

Setelah ekstraksi, penelitian dilakukan dengan menganalisis pola dan temuan lintas studi, yaitu apabila asesmen berbasis AI berhasil, apa faktor pendukungnya, apa tantangan yang muncul, bagaimana desain tugas otentik dikaitkan dengan AI, dan bagaimana aspek validitas/reliabilitas serta keadilan (*fairness*) ditangani. Hasil analisis kemudian disintesis ke dalam kerangka konseptual asesmen dalam kurikulum *deep-learning* berbasis AI, yang mencakup elemen-elemen kunci seperti tugas otentik, analitik data, personalisasi, dan audit algoritma.

### **Keterbatasan Metode**

Terdapat beberapa keterbatasan dalam penelitian ini di antaranya: 1) Karena berbasis literatur; 2) Tidak ada pengumpulan data primer lapangan sendiri; 3) Terkadang detail metodologi dalam artikel yang dikaji terbatas; dan 4) Karena teknologi AI dan asesmen terus berkembang, maka hasil yang disajikan perlu dipandang sebagai kondisi saat ini dan dapat berubah. Meski demikian, penelitian ini dapat menjadi fondasi bagi penelitian empiris lanjutan.

## RESULTS AND DISCUSSION

Karakteristik asesmen yang cocok untuk kurikulum *deep learning* memiliki beberapa karakteristik kunci yakni: 1) Tugas bersifat otentik (*real-world*, proyek, kolaboratif); 2) Memberi peluang transfer pengetahuan ke konteks baru; 3) Memungkinkan refleksi murid; dan 4) Menyediakan umpan balik yang mendalam. Tugas otentik menjadi sangat penting karena hanya dengan situasi nyata atau simulasi kompleks murid dapat menginternalisasi konsep dan menghubungkan antar ide. Dalam konteks AI, tugas dapat dirancang sedemikian rupa agar murid menggunakan alat AI sebagai bagian dari proses (misalnya eksplorasi data dengan AI dan kolaborasi antar-murid menggunakan platform AI). Namun, asesmen akhir tetap mengukur kontribusi kognitif dan reflektif murid. AI memainkan peran penting terutama pada tahap formatif yang memungkinkan pemantauan kemajuan murid secara *real-time*, memberikan umpan balik langsung, dan merekomendasikan jalur belajar yang disesuaikan. Namun untuk asesmen sumatif, penggunaan AI masih memerlukan kehati-hatian, karena penilaian otomatis atau analitik harus divalidasi terhadap rubrik penilaian tradisional dan didukung oleh penilaian manusia.

Penerapan asesmen berbasis AI menghadirkan tantangan signifikan dalam menjaga validitas dan reliabilitas penilaian. Validitas asesmen masih menjadi isu utama, khususnya terkait sejauh mana instrumen berbasis AI mampu mengukur kompetensi yang dituju secara akurat, sementara reliabilitas berkaitan dengan konsistensi hasil penilaian antar murid dan waktu. Selain itu, penggunaan algoritma penilaian otomatis berpotensi menimbulkan bias, terutama ketika *dataset* tidak representatif atau ketika algoritma cenderung memprioritaskan tipe jawaban tertentu. Untuk meminimalkan risiko tersebut, peneliti menemukan bahwa pentingnya desain asesmen yang mengombinasikan rubrik penilaian yang teruji, keterlibatan penilai manusia, serta algoritma yang bersifat transparan. Uji coba awal asesmen diperlukan untuk mengevaluasi reliabilitas sistem, sementara analisis perbedaan capaian antar kelompok murid, seperti berdasarkan latar belakang sosial-ekonomi, menjadi langkah krusial dalam memastikan aspek *fairness* dalam penilaian berbasis AI.

Implementasi asesmen berbasis AI dipengaruhi oleh memiliki beberapa tantangan sekaligus faktor pendukung yang saling berkaitan yaitu sebagai berikut:

1. Kesiapan guru menjadi aspek krusial, karena guru dituntut harus memiliki literasi data, pemahaman tentang analitik AI, dan mampu menafsirkan umpan balik yang dihasilkan. Berbagai studi mengindikasikan bahwa keterbatasan pelatihan dan pendampingan bagi guru masih menjadi hambatan signifikan dalam pemanfaatan AI secara optimal;
2. Faktor infrastruktur teknologi, terutama terkait dengan ketersediaan perangkat, konektivitas internet, dan sistem manajemen pembelajaran berbasis AI;
3. Desain tugas juga perlu dirancang agar memungkinkan penggunaan AI sebagai alat bantu, bukan pengganti berpikir murid, sementara rubrik penilaian harus disusun secara eksplisit dan selaras dengan kompetensi yang akan dinilai;
4. Aspek etika, privasi, dan keamanan data juga perlu menjadi prioritas utama, karena penggunaan AI melibatkan pengumpulan dan analisis data murid, sehingga perlu kebijakan atas perlindungan privasi, persetujuan murid/orang tua, serta audit algoritma untuk menjamin penggunaan AI yang bertanggung jawab.

## Discussion

### Asesmen yang Efektif Untuk Kurikulum *Deep Learning*

Asesmen yang efektif untuk kurikulum *deep learning* harus menempatkan tugas otentik (proyek dunia nyata, studi kasus, simulasi, dan aktivitas kolaboratif) sebagai pusat desainnya karena tugas semacam itu memfasilitasi proses internalisasi konsep, integrasi antar konsep, dan transfer pengetahuan ke konteks baru. Literatur terkini menunjukkan bahwa autentisitas tugas ditandai oleh relevansi kontekstual, kompleksitas situasi, dan kebutuhan aplikasi pengetahuan memungkinkan penilaian yang menangkap proses berpikir mendalam murid, bukan sekadar keluaran produk yang terukur secara kuantitatif. Selain itu, pengukuran yang berfokus pada produk dan proses (misalnya portofolio, rubrik analitik, penilaian *peer/collaborative*) memberi peluang bagi murid untuk merefleksikan strategi kognitifnya dan menunjukkan bukti transfer keterampilan ke situasi lain yang menjadi inti dari pembelajaran mendalam. Penelitian sistematis dan kajian empiris terbaru mendukung pergeseran ke arah asesmen otentik sebagai strategi kunci untuk mewujudkan hasil pembelajaran mendalam (Baines *et al.*, 2025; Vlachopoulos & Makri, 2024).

Dalam konteks integrasi AI, desain tugas harus memungkinkan penggunaan alat AI sebagai bagian dari proses belajar misalnya eksplorasi data dengan *tools* AI, dukungan ITS untuk *scaffolding*, atau kolaborasi berbantuan platform AI tetapi asesmen akhir harus dirancang agar secara eksplisit menilai kontribusi kognitif dan reflektif murid. Literatur 2023–2025 menekankan model-model di mana AI bertindak sebagai *scaffolding* atau *augmenting tool*, bukan sebagai pengganti berpikir murid. Dalam konsep tersebut, instrumen penilaian perlu memasukkan indikator yang menilai argumentasi, justifikasi keputusan, proses desain, dan refleksi metakognitif murid. Selain itu, studi kasus penggunaan AI dalam *Problem-Based Learning* (PBL) menunjukkan bahwa penetapan aturan penggunaan AI, rubrik yang membedakan kontribusi manusia dan keluaran alat, serta mekanisme verifikasi (misalnya artefak proses, wawancara lisan, atau observasi langsung) penting untuk menjaga validitas dan integritas asesmen di era GenAI. Kajian tentang *assessment for learning* berbasis AI juga menegaskan perlunya mekanisme audit, transparansi algoritma, dan kebijakan etis agar umpan balik AI mendukung bukan merusak tujuan pembelajaran mendalam (Gonsalves, 2025; Memarian & Doleck, 2024).

### Peran AI dalam Mendukung Asesmen Formatif dan Sumatif

AI menunjukkan peran yang paling transformasional pada tahap formatif, ketika pemantauan progres, intervensi, dan umpan balik berkala sangat dibutuhkan. Melalui *learning analytics* dan sistem tutor cerdas, AI dapat memantau jejak belajar murid secara *real-time*, memberi umpan balik langsung yang bersifat diagnostik, dan merekomendasikan jalur belajar yang dipersonalisasi berdasarkan pola kesulitan individu, fitur ini dinilai memperkuat proses *assessment for learning* dan *self-regulated learning*. Studi empiris menemukan bahwa *analytic-driven feedback* meningkatkan keterlibatan reflektif dan membantu guru menargetkan intervensi formatif. Selain itu, kajian tentang GenAI juga menunjukkan potensi umpan balik *real-time* yang bersifat dialogis untuk mendukung pembelajaran berkelanjutan (Arslan *et al.*, 2024; Banihashem *et al.*, 2025; Manganello & Fante, 2025).

Namun, untuk asesmen sumatif penggunaan AI mensyaratkan kehati-hatian dan validasi yang ketat. Penilaian otomatis atau analitik harus dipadankan dengan rubrik penilaian tradisional dan diuji terhadap standar validitas, karena tanpa bukti bahwa keluaran algoritma merefleksikan konstruk yang dimaksudkan, inferensi skor bisa tidak akurat. Literatur terbaru menekankan praktik *human-in-the-loop*: verifikasi/peninjauan oleh pendidik, audit algoritmik, dan dokumentasi transparan diperlukan untuk menjamin reliabilitas, *fairness*, dan akuntabilitas dalam penggunaan AI untuk penilaian sumatif. Studi implementasi juga menunjukkan bahwa meskipun guru menghargai kecepatan umpan balik AI pada konteks formatif, mereka cenderung kurang percaya terhadap skor otomatis untuk tujuan sumatif tanpa supervisi manusia dan mekanisme validasi yang jelas (Berg & Papadopoulos, 2025; Kaldaras *et al.*, 2024).

### **Tantangan Validitas, Reliabilitas, dan Etika dalam Asesmen Berbasis AI**

Salah satu tantangan utama dalam implementasi asesmen berbasis AI adalah menjamin validitas apakah skor/indikator yang dihasilkan benar-benar merefleksikan konstruk kompetensi yang dimaksud dan reliabilitas, yakni konsistensi hasil antar waktu, antar item, dan antar kelompok responden. Literatur mutakhir menegaskan bahwa model otomatis sering menangkap pola permukaan bahasa atau fitur teknis yang korelatif dengan skor, tetapi tidak selalu mengukur proses berpikir mendalam yang menjadi target kurikulum *deep learning*; oleh karena itu bukti validitas harus dikumpulkan ulang untuk setiap sistem AI, meliputi bukti isi, respons, hubungan skor dengan indikator eksternal, dan konsekuensi penggunaan skor. Studi-kajian dan pedoman baru merekomendasikan kombinasi bukti empiris (misalnya korelasi dengan rubrik analitik dan penilaian ahli), dokumentasi pengembangan model, serta mekanisme *human-in-the-loop* untuk menangani isu interpretasi skor (Kaldaras *et al.*, 2024; Kinnear *et al.*, 2024).

Risiko bias dan ketidakadilan (*fairness*) juga menjadi perhatian utama: *training* data yang tidak representatif, perbedaan linguistik, dan fenomena *sampling* dapat menyebabkan algoritma memberikan perlakuan yang tidak setara terhadap kelompok beragam (misalnya pembelajar bahasa kedua atau kelompok sosio-ekonomi tertentu). Oleh karena itu desain asesmen berbasis AI harus memasukkan langkah-langkah mitigasi untuk menguji reliabilitas dan kesetaraan skor, analisis diferensial item antar subkelompok, penggunaan metrik *fairness* (misalnya *disparate impact*, ABROCA), audit algoritma berkala, serta penggabungan rubrik yang teruji dan penilaian ahli sebagai verifikasi akhir. Studi-studi internasional juga menekankan pentingnya transparansi model, pelaporan metrik bias, dan keterlibatan evaluator manusia dalam *loop* keputusan agar inferensi penilaian tetap sah dan adil; kajian lokal menunjukkan kebutuhan tambahan berupa peningkatan literasi AI bagi pendidik dan penyesuaian konteks data lokal sebelum implementasi skala luas (Ali *et al.*, 2025; Andersen *et al.*, 2025; Chinta *et al.*, 2025).

Hasil analisis dari berbagai literatur menunjukkan beberapa implikasi praktis yang dapat diterapkan dalam konteks ini yakni: 1) Menggunakan rubrik analitik yang jelas sebagai standar konstruk; 2) Uji reliabilitas serta analisis perbedaan antar kelompok sebelum penerapan sumatif; 3) Terapkan *human-in-the-loop* (verifikasi pendidik) untuk kasus-kasus bernilai tinggi; dan 4) Lakukan audit algoritmik dan pelaporan transparan metrik validitas dan *fairness* secara berkala. Pendekatan terpadu ini menggabungkan bukti empiris, verifikasi

manusia, dan tata kelola teknis lebih memungkinkan asesmen berbasis AI menghasilkan inferensi yang valid, andal, dan adil dalam konteks kurikulum *deep learning* (Johnson & McCaffrey, 2023; Kinnear et al., 2024; Zheng et al., 2024).

### **Tantangan dan Faktor Pendukung dalam Asesmen Berbasis AI**

Implementasi asesmen berbasis AI sangat dipengaruhi oleh kesiapan guru. Literasi data, pemahaman tentang analitik AI, dan kemampuan menafsirkan umpan balik analitik merupakan prasyarat agar AI dapat dipakai secara pedagogis bukan sekadar teknis. Banyak studi menyoroti bahwa kurangnya pelatihan profesional dan program pengembangan kapasitas menjadi hambatan utama, karena apabila guru belum terlatih, *output* AI mudah disalahartikan atau justru mengurangi praktik penilaian reflektif yang seharusnya. Oleh karena itu program pelatihan untuk *AI-literacy* dan *Professional Development* (PD) yang berfokus pada interpretasi data, desain tugas, serta praktik *human-in-the-loop* sangat diperlukan (Tan et al., 2025).

Infrastruktur teknologi merupakan faktor pendukung kritis. Ketersediaan perangkat memadai, konektivitas internet yang stabil, dan integrasi sistem manajemen pembelajaran (LMS) dengan modul AI menjadi prasyarat operasional. Tanpa infrastruktur yang andal, sistem analitik tidak bisa memproses data real-time atau memberikan umpan balik adaptif sehingga potensi AI untuk *assessment-for-learning* tidak maksimal. Penelitian implementasi dan tinjauan literatur menunjukkan pula bahwa kesenjangan infrastruktur (termasuk ketergantungan pada vendor komersial LMS) seringkali menimbulkan masalah interoperabilitas, biaya, dan skala adopsi di sekolah-sekolah dengan sumber daya terbatas (Pardosi et al., 2024; Vergara et al., 2024).

Dalam desain tugas dan rubrik, tantangannya adalah merancang tugas yang memposisikan AI sebagai *scaffolding* atau alat bantu proses berpikir, bukan sebagai pengganti berpikir murid, sekaligus menyusun rubrik eksplisit yang memisahkan dan menilai kontribusi kognitif manusia dari keluaran otomasi AI. Literatur terbaru menekankan dua praktik penting: 1) rubrik analitik yang teruji untuk menangkap aspek proses (argumentasi, justifikasi, refleksi); dan 2) mekanisme verifikasi (artefak proses, wawancara lisan, atau portofolio) untuk memastikan validitas inferensi penilaian. Beberapa studi eksperimen dengan LLM dan *chatbot* juga merekomendasikan prosedur penyesuaian rubrik saat AI diintegrasikan, agar reliabilitas dan kesesuaian konstruk tetap terjaga (Fernandez-Sanchez et al., 2024; Hmoud et al., 2024).

Isu etika, privasi, dan keamanan data adalah tantangan non-teknis yang krusial. Penggunaan AI memerlukan pengumpulan dan pemrosesan data murid yang sensitif, sehingga kebijakan perlindungan data, persetujuan orang tua/murid, enkripsi, dan audit algoritmik harus dijalankan. Kajian etika menekankan pula kebutuhan transparansi model, dokumentasi bias, serta mekanisme akuntabilitas ketika keputusan berdampak tinggi (misalnya penentuan kelulusan atau peringkat). Dalam praktik, institusi perlu menerapkan kebijakan *responsible AI* termasuk audit berkala dan pelaporan metrik *fairness* serta membangun kepercayaan melalui komunikasi terbuka dengan pemangku kepentingan (Akgun & Greenhow, 2022; Lim et al., 2025).

Hasil kajian literatur mengidentifikasi beberapa implikasi praktis yang relevan untuk diperhatikan dalam penerapan asesmen berbasis AI yakni: 1) Menjalankan program PD untuk *AI-literacy* guru yang praktis dan berbasis kasus (*case-based*); 2) Memprioritaskan investasi terhadap interoperabilitas dan konektivitas infrastruktur LMS; 3) Merancang rubrik analitik dan prosedur verifikasi proses saat menggunakan AI; 4) Menerapkan kebijakan privasi, persetujuan eksplisit, dan audit algoritma berkala. Pendekatan terpadu menggabungkan kemampuan manusia, kesiapan teknologi, desain asesmen yang matang, dan tata kelola etis menjadi fondasi agar asesmen berbasis AI dapat berfungsi efektif dan adil dalam kurikulum *deep learning* (Hmoud *et al.*, 2024; Lim *et al.*, 2025; Vergara *et al.*, 2024).

## CONCLUSION

Berdasarkan analisis literatur yang telah dilakukan, penelitian ini mengajukan hipotesis konseptual bahwa asesmen dalam kurikulum *deep learning* berbasis AI akan lebih efektif apabila dirancang sebagai sistem terpadu yang menggabungkan asesmen otentik, analitik pembelajaran berbantuan AI, dan penilaian ahli sebagai komponen sentral. Integrasi ini diperkirakan mampu meningkatkan akurasi penilaian kompetensi tingkat tinggi sekaligus mendukung personalisasi pembelajaran.

Tujuan penelitian telah tercapai melalui pemetaan karakteristik asesmen yang selaras dengan *deep learning*, penjelasan peran AI dalam mendukung asesmen formatif maupun sumatif, identifikasi tantangan validitas, serta perumusan rekomendasi desain asesmen yang sesuai dengan prinsip pembelajaran abad ke-21. Kajian ini menunjukkan bahwa meskipun AI memiliki potensi besar dalam membantu proses evaluasi, kualitas asesmen tetap sangat dipengaruhi oleh keterlibatan aktif pendidik, kesiapan ekosistem teknologi, dan keselarasan desain asesmen dengan tujuan kurikulum.

Sebagai tindak lanjut, diperlukan upaya keberlanjutan untuk memperkuat kapasitas pendidikan dalam literasi AI dan desain asesmen, membangun infrastruktur data yang aman dan transparan, serta mengembangkan mekanisme audit algoritmik untuk memastikan *fairness* dan akuntabilitas sistem. Penelitian lanjutan juga disarankan untuk menguji model asesmen terpadu ini dalam konteks implementasi nyata, sehingga diperoleh bukti empiris mengenai efektivitas, tantangan praktis, serta strategi adaptasi yang paling sesuai bagi institusi pendidikan.

## AUTHOR'S NOTE

Penulis menyatakan bahwa tidak ada konflik kepentingan terkait publikasi artikel ini. Penulis menegaskan bahwa data dan isi artikel bebas dari plagiarisme. Terima kasih kepada Direktorat Penelitian dan Pengabdian Kepada Masyarakat Universitas Pendidikan Indonesia Tahun Anggaran 2025.

## REFERENCES

Akgun, S., & Greenhow, C. (2022). Artificial intelligence in education: Addressing ethical challenges in K-12 settings. *AI and Ethics*, 2, 431-440.

- Ali, E. Y., Nugraha, R. G., & Bagja, N. (2025). The effectiveness of AI-Based education management systems in implementing deep learning curriculum in elementary schools. *Journal of Integrated Elementary Education, 5*(2), 339-352.
- Aloisi, C. (2023). The future of standardised assessment: Validity and trust in algorithms for assessment and scoring. *European Journal of Education, 58*(1), 98-110.
- Andersen, N., Mang, J., Goldhammer, F., & Zehner, F. (2025). Algorithmic fairness in automatic short answer scoring. *International Journal of Artificial Intelligence in Education, 38*, 1-38.
- Arslan, B., Lehman, B., Tenison, C., Sparks, J. R., López, A. A., Gu, L., & Zapata-Rivera, D. (2024). Opportunities and challenges of using generative AI to personalize educational assessment. *Frontiers in Artificial Intelligence, 7*(1), 1-8.
- Baines, S., Otermans, P., Tree, D., & Worsfold, N. (2025). Measuring and mapping authentic assessment with a novel quantitative typology. *Teacher in Higher Education, 30*(3), 663-682.
- Banihashem, S. K., Gasevic, D., Noroozi, O., Jarodzka, H., Brinke, D. J., & Draschler, H. (2025). Optimizing formative assessment with learning analytics. *Review of Educational Research, 1*(1), 1-49.
- Berg, S. V. D., & Papadopoulos, P. M. (2025). Summative assessment with artificial intelligence: Qualitative analysis and comparison of technology acceptance in student and teacher populations. *Innovations in Education and Teaching International, 62*(5), 1529-1544.
- Bulut, O., Beiting-Parrish, M., Casabianca, J. M., Slater, S. C., Jiao, H., Song, D., & Walsh, C. (2024). The rise of artificial intelligence in educational measurement: Opportunities and ethical challenges. *Chinese/English Journal of Educational Measurement and Evaluation, 5*(3), 1-34.
- Burstein, J., & LaFlair, G. T. (2024). Where assessment validation and responsible AI meet. *Europe Knowledge Development Institute, 1*(1), 1-32.
- Cahyani, V. P., Ahmad, F., Soekarman, & Nawati, M. Z. M. (2024). Authentic assessment as a solution to enhance deep learning and maintain academic integrity in higher education. *Journal of Educational Analytics (JEDA), 3*(3), 531-544.
- Campos, C. M. (2025). The impact of artificial intelligence on personalized learning in higher education: A systematic review. *Trends in Higher Education, 4*(2), 1-15.
- Chen, J., & Singh, C. K. S. (2024). A systematic review on deep learning in education: Concepts, factors, models and measurements. *Journal of Education and Educational Research, 7*(1), 125-129.
- Chinta, S. V., Wang, Z., Yin, Z., Hoang, N., Gonzalez, M., Quy, T. L., & Zhang, W. (2024). FairAIED: Navigating fairness, bias, and ethics in educational AI applications. *CORR, 1*(1), 1-59.
- Deepshikha, D. (2025). A comprehensive review of AI-Powered grading and tailored feedback in universities. *Discover Artificial Intelligence, 5*(1), 1-18.

- Fernandez-Sanchez, A., Lorenzo-Castineiras, J., & Sanchez-Bello, A. (2024). Navigating the future of pedagogy: The integration of AI tools in developing educational assessment rubrics. *European Journal of Education, 60*(1), 1-13.
- Gonsalves, C. (2025). Contextual assessment design in the age of generative AI. *Journal of Learning Development in Higher Education, 34*(1), 1-14.
- Halim, M. F. (2025). Transforming educational leadership: The role of school principals in implementing deep learning-based curricula. *Journal of Educational Management Research, 4*(1), 308-321.
- Hmoud, M., Swaity, H., Anjass, E., & Aguaded-Ramirez, E. (2024). Rubric development and validation for assessing tasks' solving via AI chatbots. *The Electronic Journal of E-Learning, 22*(6), 1-17.
- Inel, O., Draws, T., & Aroyo, L. (2023). Collect, measure, repeat: Reliability factors for responsible AI data collection. *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing, 11*(1), 51-64.
- Johnson, M. S., & McCaffrey, D. F. (2023). Evaluating fairness of automated scoring in educational measurement. *Advancing Natural Language Processing in Educational Assessment, 142*(1), 1-10.
- Kaldaras, L., Akaeze, H. O., & Reckase, M. D. (2024). Developing valid assessments in the era of generative artificial intelligence. *Sec. Assessment, Testing and Applied Measurement, 9*(1), 1-10.
- Kasi, Y. F., Bai, D. V., Novia, N., Deporos, S. R. C., & Mabbaya, A. D. (2025). Implementation of deep learning in school curriculum: Perspective of teachers in Nagekeo Regency. *Jurnal Penelitian Pendidikan Paedagogia, 28*(2), 320-328.
- Kinncar, B., St-Onge, C., Schumacher, D. J., Marceau, M., & Naidu, T. (2024). Validity in the next era of assessment: Consequences, social impact, and equity. *Perspectives on Medical Education, 13*(1), 452-459.
- Lim, T., Gottipati, S., & Cheong, M. (2025). What students really think: Unpacking AI ethics in educational assessments through a triadic framework. *International Journal of Educational Technology in Higher Education, 22*(56), 1-32.
- Manganello, F., & Fante, C. (2025). Scoping review on the role of learning analytics in assessing and fostering creativity in educational contexts. *Journal of Learning Analytics, 12*(2), 5-18.
- Memarian, B., & Doleck, T. (2024). A Review of assessment for learning with artificial intelligence. *Computers in Human Behavior: Artificial Humans, 2*(1), 1-11.
- Nafi'ah, J., & Faruq, D. J. (2025). Conceptualizing deep learning approach in primary education: Integrating mindful, meaningful, and joyful. *Journal of Educational Research and Practice, 6*(2), 78-89.
- Pardosi, V. B. A., Xu, S., Umurohmi, U. Nurdiana, & Sabur, F. (2024). Implementation of an artificial intelligence based learning management system for adaptive learning. *Al-Fikrah: Jurnal Manajemen Pendidikan, 12*(1), 149-161.

- Putri, D. N., Mayasril, N., & Gusman, R. (2025). Asesmen abad 21: Menakar kompetensi, bukan sekedar nilai. *Didaktik: Jurnal Ilmiah PGSD STKIP Subang*, 11(2), 221-228.
- Sadykova, A., Iskakova, M., Ismailova, G., Ishmukhametova, A., Sovetova, A., & Mukasheva, K. (2024). The impact of a metacognition-based course on school students' metacognitive skills and biology comprehension. *Frontiers in Education*, 9(1), 1-9.
- Sappaile, B. I. (2025). A comparative analysis of the merdeka belajar framework as a means to achieve deep learning in elementary school students: A literature review. *Cosmos: Jurnal Ilmu Pendidikan, Ekonomi dan Teknologi*, 2(4), 944-963.
- Suherman, A., Supriyadi, T., Safari, I., Saptani, E., Fauzi, R. A., Sudirjo, E., Komiljon O'g'li, T. S., & Abdisamiyevich. (2025). Bridging teacher readiness and deep learning-based teaching practice: Assessing the effectiveness of the active model for enhancing teacher pedagogy. *International Research Journal of Multidisciplinary Scope*, 6(4), 284-298.
- Tan, L. Y., Hu, S., Yeo, D. J., & Cheong, K. H. (2025). A comprehensive review on automated grading systems in STEM using AI techniques. *Mathematics*, 13(1), 1-10.
- Tan, X., Cheng, G., & Ling, M. H. (2025). Artificial intelligence in teaching and teacher professional development: A systematic review. *Computers and Education: Artificial Intelligence*, 8(1), 1-19.
- Timperley, C., & Schick K. (2024). Assessment as pedagogy: Inviting authenticity through relationality, vulnerability and wonder. *Teaching in Higher Education*, 30(3), 592-607.
- Vergara, D., Lampropoulos, G., Antón-Sancho, Á., & Fernández-Arias, P. (2024). Impact of artificial intelligence on learning management systems: A bibliometric review. *Multimodal Technologies and Interaction*, 8(9), 1-19.
- Vlachopoulos, D., & Makri, A. (2024). A systematic literature review on authentic assessment in higher education: Best practices for the development of 21st century skills, and policy considerations. *Studies in Educational Evaluation*, 83(1), 1-13.
- Weidlich, J., Fink, A., Frey, A., Jivet, I., Gombert, S., Menzel, L., Giorgashvili, T., Yau, J., & Drachsler, H. (2025). Highly informative feedback using learning analytics: How feedback literacy moderates students perceptions of feedback. *International Journal of Education Technology in Higher Education*, 22, 43(1), 1-25.
- Weng, C., Chen, C., & Ai, X. (2022). A pedagogical study on promoting students' deep learning through design-based learning. *International Journal of Technology and Design Education*, 33(1), 1653-1647.
- Yan, L., Sha, L., Zhao, L., Li, Y., Maldano, R. M., Chen, G., Li, X., Jin, Y., & Gasevic, D. (2023). Practical and ethical challenges of large language models in education: A systematic scoping review. *British Journal of Educational Technology*, 55(1), 90-112.
- Yuliansyah, H., & Saidah, N. (2025). Transformasi kurikulum berbasis kecerdasan buatan sebagai strategi pendidikan abad 21. *Jurnal Pendidikan Islam Educen*, 9(1), 53-66.
- Zheng, E. L., Jin, W., Hamarneh, G., & Lee, S. S. J. (2024). From human-in-the-loop to human-in-power. *The American Journal of Bioethics*, 24(9), 84-86.