Amorti: Jurnal Studi Islam Interdisipliner
Vol. 5, No. 1, Januari 2026,pp. 1688-1701
http://journal.amorfati.id/index.php/amorti | | ISSN 2962-9209

Interdisciplinary Scope of Contemporary Islamic Studies • Interdisciplinary Islamic Education • Interdisciplinary Islamic Law, Ethics, and Fatwa Studies • Interdisciplinary Islamic Economics, Finance, and Philanthropy • Islam, Society, and Culture • Islam, Politics, Governance, and Public Policy • Islam, Gender, Family, and Human Rights • Islam, Media, Communication, and Digital Da'wah • Islam, Health, Environment, and Sustainable Development • Islam, Science, Technology, Artificial Intelligence, and Digital Transformation

Transformasi Paradigma Evaluasi Pembelajaran di Era AI: Dari *Assessment of Learning* Menuju *Intelligent, Humanistic, and Personalized Assessment* (IHPA)

Herman Tadjuddin Mapatarani[a,1,*]

[a] Institut Agama Islam Negeri (IAIN) Bone, Indonesia (NIM 860322025009)
[1] hermantadjuddin.m@gmail.com
*Penulis koresponden

ARTICLE INFO

Article history
Received:
19-12-2025
Revised:
19-01-2026
Accepted:
23-01-2026

Keywords
adaptive assessment;
artificial intelligence;
evaluasi pembelajaran;
human-centred AI

ABSTRACT

The diffusion of artificial intelligence into educational practice has outpaced the conceptual frameworks available for evaluating learning. Recent work has advanced automated scoring, adaptive testing, and generative feedback, yet the frameworks proposed so far tend to treat the technology either as an instructional and administrative aid or as a risk to be governed, and few specify the point at which algorithmic analysis must give way to pedagogical judgement. This article addresses that gap by reconstructing the paradigm of learning assessment and proposing a conceptual model termed Intelligent, Humanistic, and Personalized Assessment (IHPA). The study employs library research with an analytical-conceptual approach; forty-two sources retrieved from international and national databases under explicit inclusion criteria were examined through content analysis, thematic coding, cross-theme comparison, and conceptual synthesis. The discussion traces the movement from assessment of learning towards assessment for and as learning, and further towards intelligent assessment driven by learning analytics and adaptive testing. IHPA rests on humanism, constructivism, and human-centred artificial intelligence, and operates through seven stages: learner profiling, learning data collection, artificial intelligence analysis, adaptive assessment, humanistic decision making, personalized feedback, and continuous improvement. Its specific contribution lies in the fifth stage, which makes teacher interpretation a mandatory filter between algorithmic output and assessment decisions, and in extending the framework to the moral and spiritual dimensions central to Islamic education. The model remains conceptual, and empirical testing of its validity and practicality is an agenda for further research.

This is an open-access article under the CC–BY-SA license.

Pendahuluan

Kehadiran kecerdasan buatan mengubah cara lembaga pendidikan mengelola informasi tentang belajar. Sistem yang mampu mengolah rekam jejak aktivitas peserta didik dalam jumlah besar, mengenali pola pengerjaan tugas, serta menyusun rekomendasi tindak lanjut kini tersedia pada banyak platform pembelajaran. Fungsi teknologi tersebut bergeser dari

sekadar media penyampai materi menjadi perangkat yang turut menopang perencanaan, pelaksanaan, hingga penilaian pembelajaran (Nasihuddin, 2024; Tan et al., 2025). Pergeseran itu menyentuh salah satu wilayah paling sensitif dalam pendidikan, yaitu keputusan mengenai capaian belajar seseorang.

Selama beberapa dekade praktik penilaian di ruang kelas didominasi pendekatan *assessment of learning*, yakni pengukuran hasil belajar melalui tes sumatif yang dilaksanakan setelah pembelajaran berakhir. Pendekatan ini memberi informasi ringkas mengenai tingkat pencapaian kompetensi dan mudah dipertanggungjawabkan secara administratif. Akan tetapi, rekam nilai akhir jarang mampu menjelaskan bagaimana peserta didik sampai pada capaian tersebut, pada titik mana ia mengalami hambatan, serta strategi belajar apa yang sesungguhnya ia gunakan. Kritik atas keterbatasan itu telah lama dikemukakan, terutama karena penilaian yang berorientasi pada skor cenderung mendorong pembelajaran yang diarahkan kepada tes daripada penguasaan kompetensi yang bermakna (Wiggins, 1998).

Koreksi terhadap keadaan tersebut melahirkan paradigma *assessment for learning*, yang menempatkan penilaian sebagai bagian dari proses pembelajaran dan bukan penutupnya. Informasi hasil penilaian dimanfaatkan sebagai umpan balik agar peserta didik mengetahui jarak antara capaian dan sasaran belajarnya, sementara guru memperoleh dasar untuk menyesuaikan strategi mengajar (Sadler, 1989; Black & Wiliam, 1998). Perkembangan berikutnya, *assessment as learning*, menggeser posisi peserta didik dari objek menjadi subjek penilaian melalui refleksi diri dan penilaian antarteman, sehingga kemampuan metakognitif tumbuh bersamaan dengan penguasaan materi (Earl, 2003; Zimmerman, 1990). Ketiga paradigma itu berkembang dalam konteks ketika data belajar masih dikumpulkan dan ditafsirkan secara manual oleh guru.

Kondisi teknis itulah yang berubah secara mendasar. Analitik pembelajaran memungkinkan pengumpulan dan pelaporan data belajar berlangsung terus-menerus selama pembelajaran, bukan hanya pada akhir satuan materi (Siemens, 2013). Pengujian adaptif terkomputerisasi menyesuaikan tingkat kesulitan butir dengan kemampuan yang terduga dari respons sebelumnya, sehingga estimasi kemampuan menjadi lebih presisi dengan jumlah butir yang lebih sedikit (Yang et al., 2022). Sistem rekomendasi berbasis pembelajaran mesin mampu menyusun lintasan belajar yang berbeda bagi peserta didik yang berbeda (Ruan & Lu, 2025), dan umpan balik yang dahulu bersifat kelas kini dapat dirancang pada tingkat individu dengan variasi isi, bentuk, serta waktu penyampaian (Maier & Klotz, 2022; Villegas-Ch et al., 2025). Kemampuan teknis ini menjanjikan evaluasi yang cerdas dalam pengolahan data sekaligus lebih adil dalam memperlakukan perbedaan individual.

Bukti empiris terkini memperlihatkan bahwa kemampuan teknis tersebut telah diuji pada praktik penilaian yang sesungguhnya, dengan hasil yang terbelah. Perbandingan antara penilaian tulisan oleh pengajar dan oleh kecerdasan buatan generatif menemukan bahwa keselarasan umpan balik dan keabsahan skor yang dihasilkan mesin belum setara dengan penilaian manusia, sekaligus memunculkan persoalan tergerusnya kewenangan profesional pengajar (Anbreen et al., 2026). Penerapan penilaian komparatif berbantuan pembelajaran mesin menghemat waktu penilaian secara berarti, tetapi keandalannya bergantung pada rancangan algoritma dan jumlah perbandingan yang tersedia (Gray et al., 2024). Mahasiswa menilai umpan balik yang dihasilkan kecerdasan buatan generatif cepat dan rinci, namun kurang memadai pada aspek yang menuntut pemahaman konteks (Grion et al., 2026), sedangkan pengajar yang merancang ulang praktik umpan baliknya melaporkan pergeseran peran dari penulis komentar menjadi penilai kelayakan komentar mesin (Cooper et al., 2026). Pola temuan tersebut menunjukkan bahwa persoalan pokoknya tidak lagi terletak pada kemampuan mesin menilai, melainkan pada sejauh mana keluarannya boleh menentukan keputusan.

Sejumlah kerangka kerja memang telah diusulkan untuk menata penggunaan kecerdasan buatan dalam penilaian, tetapi masing-masing menggarap satu sisi persoalan. Kerangka umpan balik tiga pihak mengatur pembagian peran antara mahasiswa, pengajar, dan mesin dalam penyusunan komentar atas tugas yang kompleks, namun berhenti pada tahap pemberian

umpan balik (Venetsanos, 2026). Kerangka perancangan ulang penilaian menekankan keadilan dan kelayakan praktik di tengah maraknya kecerdasan buatan generatif, dengan fokus pada bentuk tugas dan bukan pada alur keputusan evaluatif (Laidlaw, 2026). Kerangka ketergantungan epistemik mengarahkan perhatian pada tanggung jawab mahasiswa dalam menilai keluaran mesin (Du & Yuan, 2026), sementara kerangka tata kelola menerjemahkan prinsip etika menjadi aturan yang dapat dieksekusi pada tataran kelembagaan (Evangelista & Salman Bukhari, 2026). Perhatian lain terpusat pada pengendalian risiko penyalahgunaan, yakni keabsahan penilaian dan integritas akademik (Ali & Maroulis, 2026; Mangubat & Mangubat, 2026). Di sisi lain, kajian di Indonesia sebagian besar masih menempatkan kecerdasan buatan sebagai media atau alat bantu administrasi pembelajaran, sehingga pembahasannya berhenti pada pemilihan aplikasi, peningkatan minat belajar, atau efisiensi tugas kepengurusan (Lustani et al., 2025; Samsudin, 2025; Afif & Nawawi, 2024). Telaah sistematis atas arah penelitian evaluasi berbasis kecerdasan buatan menegaskan bahwa kerangka yang menghubungkan kapasitas analitik mesin dengan pertimbangan pedagogis guru masih jarang dirumuskan (Yasfi, 2026). Celah yang tersisa, dengan demikian, adalah ketiadaan model yang merangkai kapabilitas analitik, pertimbangan kemanusiaan, dan personalisasi ke dalam satu siklus evaluasi yang utuh sekaligus menetapkan secara tegas pada tahap mana penafsiran guru wajib hadir sebelum keluaran algoritma berubah menjadi keputusan penilaian.

Celah itu bukan persoalan akademik semata. Penilaian berbasis algoritma terbukti dapat menghasilkan keputusan yang tidak setara bagi kelompok peserta didik tertentu apabila data pelatihannya tidak mewakili keragaman populasi (Baker & Hawn, 2022; Idowu et al., 2024). Perbandingan antara penilaian mesin dan penilaian manusia pada tugas menulis akademik juga menunjukkan pola ketidaksesuaian yang belum sepenuhnya dapat dijelaskan (Cui, 2026). Persoalan tersebut bertaut dengan keamanan data pribadi, keterbukaan cara kerja algoritma, dan akuntabilitas keputusan, yang menjadi alasan UNESCO (2023) menuntut agar pemanfaatan kecerdasan buatan dalam pendidikan berpijak pada penghormatan martabat manusia. Ditambah kesenjangan akses teknologi antarwilayah yang masih lebar (Martin et al., 2024) serta kesiapan kompetensi digital pendidik yang belum merata (Baizhanov et al., 2026), penerapan kecerdasan buatan dalam evaluasi tidak dapat diserahkan kepada logika efisiensi.

Berdasarkan keadaan tersebut, artikel ini bermaksud merekonstruksi paradigma evaluasi pembelajaran pada era kecerdasan buatan dan menawarkan sebuah model konseptual yang disebut *Intelligent, Humanistic, and Personalized Assessment* (IHPA). Pembahasan diarahkan pada empat hal: menelusuri pergeseran paradigma evaluasi dari pendekatan tradisional menuju evaluasi berbasis kecerdasan buatan; merumuskan landasan dan karakteristik model IHPA; menjelaskan tahapan operasionalnya dalam pembelajaran; serta menganalisis implikasi, tantangan, dan prospek penerapannya. Kebaruan kajian ini terletak pada tiga hal. Pertama, IHPA merangkai analitik mesin, pertimbangan humanistik, dan personalisasi ke dalam satu siklus evaluasi tujuh tahap, sementara kerangka yang telah ada umumnya menggarap satu sisi persoalan saja. Kedua, model ini menempatkan tahap pengambilan keputusan humanistik sebagai penyaring wajib antara keluaran algoritma dan keputusan penilaian, sehingga otoritas profesional guru memperoleh kedudukan struktural di dalam alur evaluasi dan bukan sekadar imbauan etis. Ketiga, kerangka tersebut diperluas ke konteks pendidikan Islam melalui pembagian kerja yang tegas antara hal yang dapat direkam mesin dan penilaian atas akhlak serta penghayatan nilai yang menuntut kehadiran pendidik. Dengan demikian, model yang ditawarkan diharapkan menjadi rujukan konseptual bagi perancangan sistem evaluasi yang akurat secara akademik sekaligus tetap manusiawi dalam pelaksanaannya.

Metode

Kajian ini merupakan penelitian kepustakaan (*library research*) dengan pendekatan analitis-konseptual, yaitu penelusuran literatur yang diarahkan untuk membangun konstruksi

teoretis, bukan tinjauan sistematis yang mengukur besaran efek suatu intervensi. Penelusuran dilakukan pada pangkalan data Scopus, Crossref, ScienceDirect, SpringerLink, Taylor & Francis Online, dan DOAJ untuk literatur internasional, serta Garuda dan portal jurnal terindeks Sinta untuk literatur nasional, dengan kata kunci *artificial intelligence in assessment*, *learning analytics*, *adaptive assessment*, *personalized feedback*, *human-centred AI in education*, evaluasi pembelajaran, dan pendidikan Islam. Sebuah sumber dinyatakan layak apabila memenuhi empat syarat: membahas evaluasi atau asesmen pembelajaran yang bersinggungan dengan kecerdasan buatan, analitik pembelajaran, personalisasi, atau etika penggunaannya; berupa artikel jurnal bertelaah sejawat yang memiliki DOI, buku rujukan utama dalam teori evaluasi dan teori belajar, atau dokumen kebijakan lembaga internasional; tersedia dalam teks penuh; serta berbahasa Indonesia atau Inggris. Rentang terbit dibatasi pada 2021–2026 untuk sumber empiris agar gambaran perkembangan teknologinya mutakhir, sedangkan sumber teoretis klasik yang menjadi fondasi paradigma evaluasi tidak dibatasi waktu. Tulisan opini tanpa dasar data atau teori, prosiding tanpa telaah sejawat, dan sumber yang hanya mengulang isi sumber lain dikeluarkan dari kajian. Melalui penyaringan tersebut diperoleh 42 sumber yang dianalisis, terdiri atas 36 artikel jurnal ber-DOI serta 6 buku dan dokumen kebijakan.

Analisis ditempuh dalam empat langkah. Pertama, analisis isi atas setiap sumber untuk mengidentifikasi gagasan pokok, klaim utama, bukti pendukung, dan keterbatasan yang diakui penulisnya. Kedua, pengodean dan pengelompokan gagasan ke dalam empat tema, yaitu pergeseran paradigma evaluasi, kapabilitas analitik kecerdasan buatan, persoalan etis dan tata kelola data, serta kesiapan kompetensi pendidik. Ketiga, analisis komparatif antartema untuk menemukan titik temu dan pertentangan antarsumber, khususnya antara temuan yang menonjolkan keunggulan teknis dan temuan yang menyoroti risiko penggunaannya. Keempat, sintesis konseptual yang merangkai hasil perbandingan tersebut menjadi konstruksi model. Keabsahan penafsiran dijaga melalui triangulasi jenis sumber, yakni menyandingkan temuan penelitian akademik dengan dokumen kebijakan lembaga internasional, serta pemeriksaan ulang atas kesesuaian antara setiap klaim dalam pembahasan dan sumber yang mendukungnya. Kesimpulan ditarik secara induktif dari pola hubungan antarkonsep yang terbentuk. Model yang dihasilkan bersifat konseptual sehingga pengujian validitas dan keefektifannya secara empiris berada di luar cakupan kajian ini.

Hasil dan Pembahasan

1. Evolusi Paradigma Evaluasi Pembelajaran

Evaluasi merupakan komponen yang menentukan arah seluruh kegiatan pembelajaran. Dalam rumusan klasik, evaluasi dipahami sebagai proses sistematis untuk menetapkan sejauh mana tujuan pendidikan telah tercapai, sehingga kedudukannya melekat pada perumusan tujuan, pemilihan pengalaman belajar, dan pengorganisasian pengalaman tersebut (Tyler, 1949). Kerangka ini menempatkan evaluasi tidak semata sebagai kegiatan pengukuran, tetapi juga sebagai instrumen pengendalian mutu penyelenggaraan pendidikan. Konsekuensinya, setiap perubahan pada cara data belajar dikumpulkan dan ditafsirkan akan berpengaruh langsung terhadap mutu keputusan pendidikan yang dihasilkan.

Paradigma pertama yang mendominasi praktik adalah *assessment of learning*. Penilaian dilaksanakan setelah proses pembelajaran selesai dengan tujuan menetapkan tingkat pencapaian kompetensi; guru berperan sebagai penilai tunggal, sedangkan peserta didik berada pada posisi yang dinilai. Keberhasilan belajar direpresentasikan melalui skor, nilai, atau peringkat. Kekuatan pendekatan ini terletak pada kejelasan dan daya bandingnya, namun informasi yang dihasilkannya bersifat terminal. Penilaian yang terpusat pada hasil akhir cenderung mengabaikan proses berpikir, kreativitas, keterampilan kolaborasi, dan perkembangan karakter, sekaligus mendorong budaya belajar yang diarahkan pada tes (Wiggins, 1998).

Paradigma *assessment for learning* lahir sebagai koreksi. Sadler (1989) menunjukkan

bahwa umpan balik hanya bermakna apabila peserta didik mengetahui standar yang dituju, mampu membandingkan capaiannya dengan standar tersebut, dan memiliki cara untuk memperkecil selisihnya. Tinjauan Black dan Wiliam (1998) atas ratusan studi memperkuat temuan itu dengan bukti bahwa penilaian formatif yang dijalankan secara konsisten meningkatkan hasil belajar, terutama pada peserta didik berkemampuan rendah. Dalam kerangka ini penilaian berfungsi diagnostik: hasilnya dipakai untuk mengubah pembelajaran yang sedang berlangsung, bukan untuk melaporkan pembelajaran yang telah berakhir.

Perkembangan lanjutan, *assessment as learning*, memindahkan sebagian tanggung jawab penilaian kepada peserta didik. Melalui refleksi diri dan penilaian antarteman, peserta didik dilatih mengenali kekuatan dan kelemahan belajarnya serta memilih strategi yang sesuai (Earl, 2003). Kemampuan semacam itu merupakan inti pembelajaran yang teratur secara mandiri, yang mencakup penetapan sasaran, pemantauan proses, dan penilaian atas hasil belajar sendiri (Zimmerman, 1990). Bukti empiris terbaru menunjukkan bahwa kemampuan tersebut berkaitan erat dengan literasi umpan balik, yakni kesiapan peserta didik memahami, menilai, dan memanfaatkan masukan yang diterimanya (Chen et al., 2025).

Memasuki dasawarsa ketika data belajar tersimpan secara digital, evaluasi memperoleh basis teknis yang berbeda. Analitik pembelajaran tumbuh sebagai bidang tersendiri yang mengumpulkan, mengukur, menganalisis, dan melaporkan data tentang peserta didik untuk memahami serta mengoptimalkan proses belajar (Siemens, 2013). Pada titik ini muncul paradigma keempat, yaitu evaluasi cerdas, yang menandai perpindahan orientasi dari pengukuran hasil menuju pemantauan proses dan peramalan capaian. Telaah sistematis mengenai arah pengembangannya memperlihatkan bahwa potensi terbesarnya justru terletak pada fungsi diagnostik dan preventif, bukan pada otomatisasi penilaian akhir (Yasfi, 2026). Pergeseran keempat paradigma tersebut bukan penggantian yang saling meniadakan, melainkan pelapisan: evaluasi cerdas hanya bermakna apabila prinsip umpan balik dan keterlibatan peserta didik yang diwariskan dua paradigma sebelumnya tetap dipertahankan.

2. Kecerdasan Buatan sebagai Penggerak Transformasi Evaluasi

a. Kapabilitas Analitik dan Pembentukan Profil Belajar

Kemampuan kecerdasan buatan yang paling langsung berpengaruh pada evaluasi adalah pengolahan jejak belajar. Setiap interaksi peserta didik dengan sistem pembelajaran meninggalkan rekaman: waktu pengerjaan, urutan langkah penyelesaian, pola kesalahan yang berulang, frekuensi membuka sumber belajar, hingga keterlibatan dalam diskusi. Kumpulan rekaman itu diolah menjadi profil belajar yang jauh lebih kaya daripada nilai ujian (Siemens, 2013). Profil tersebut kemudian disajikan kepada guru melalui papan pemantau; studi pada jenjang pendidikan dasar dan menengah memperlihatkan bahwa penyajian data multimoda semacam itu membantu guru mengatur kelas dan mengambil keputusan pada saat pembelajaran masih berjalan (Possaghi et al., 2025).

Nilai praktis kapabilitas ini terletak pada pendeteksian dini. Sistem dapat menandai penurunan aktivitas, keterlambatan pengumpulan tugas, atau melemahnya interaksi jauh sebelum penurunan tersebut tampak pada nilai. Guru memperoleh peluang melakukan intervensi ketika masalah masih dapat diperbaiki. Namun ketersediaan papan pemantau tidak otomatis mengubah praktik. Penelitian intervensi menunjukkan bahwa manfaat papan pemantau baru muncul apabila disertai pelatihan yang menolong guru menafsirkan data serta menghubungkannya dengan tindakan pedagogis yang relevan (Khulbe & Tammets, 2023). Data, dengan sendirinya, bukan keputusan.

b. Asesmen Adaptif dan Umpan Balik Terpersonalisasi

Transformasi kedua menyentuh instrumen penilaian. Pengujian adaptif terkomputerisasi memilih butir berikutnya berdasarkan respons terhadap butir sebelumnya, sehingga peserta didik menerima soal yang sepadan dengan kemampuannya. Ketika mekanisme ini

digabungkan dengan model siklus ingatan belajar, sistem tidak hanya menaksir kemampuan dengan lebih presisi, tetapi juga menentukan waktu pengulangan materi yang paling menguntungkan bagi retensi (Yang et al., 2022). Pengembangan mutakhir memakai pembelajaran penguatan untuk menyusun lintasan belajar yang menyesuaikan diri secara berkelanjutan terhadap perubahan capaian peserta didik (Ruan & Lu, 2025). Instrumen yang seragam untuk seluruh kelas, dengan demikian, bukan lagi satu-satunya pilihan yang tersedia secara teknis.

Transformasi ketiga berkenaan dengan umpan balik. Kerangka klasifikasi atas umpan balik dalam lingkungan belajar digital memperlihatkan bahwa keefektifannya bergantung pada kesesuaian antara isi masukan, tingkat kerincian, dan waktu penyampaiannya dengan kebutuhan penerima (Maier & Klotz, 2022). Sistem tutor cerdas yang adaptif pada bidang sains dan matematika menunjukkan peningkatan capaian yang berarti ketika umpan baliknya disusun berdasarkan pola kesalahan individual (Villegas-Ch et al., 2025), dan meta-analisis mutakhir menemukan bahwa sistem semacam itu juga memperkuat kemampuan peserta didik mengatur proses belajarnya sendiri (Zhang et al., 2026). Pada pendidikan tinggi, kajian mengenai umpan balik yang dihasilkan kecerdasan buatan generatif melaporkan penilaian mahasiswa yang umumnya positif terhadap kecepatan dan kerinciannya, dengan catatan bahwa masukan tersebut dipandang kurang memadai pada aspek yang menuntut pemahaman konteks (Grion et al., 2026). Pengalaman pengajar yang mencoba merancang ulang praktik umpan baliknya dengan bantuan teknologi ini memperkuat catatan tersebut: peran pengajar bergeser dari penulis komentar menjadi penilai kelayakan komentar yang dihasilkan mesin (Cooper et al., 2026).

c. Persoalan Etis: Bias Algoritmik, Privasi, dan Akuntabilitas

Kapabilitas teknis di atas membawa risiko yang setara besarnya. Bias algoritmik dalam pendidikan muncul ketika model dilatih pada data yang tidak mewakili seluruh kelompok peserta didik, sehingga galat prediksi terdistribusi tidak merata dan kelompok tertentu dirugikan secara sistematis (Baker & Hawn, 2022). Pengujian pada sistem pemantauan kemajuan belajar menemukan perbedaan akurasi antarkelompok yang cukup berarti, sekalipun model yang digunakan tampak berkinerja baik secara agregat (Idowu et al., 2024). Pada penilaian tulisan akademik, perbandingan antara penilaian mesin dan penilaian manusia juga memperlihatkan inkonsistensi yang berpola (Cui, 2026). Temuan-temuan ini menunjukkan bahwa persoalannya tidak terletak pada kecanggihan model, melainkan pada asumsi bahwa keluaran model dapat diperlakukan sebagai keputusan.

Dimensi kedua adalah perlindungan data. Sistem yang analitis menuntut data dalam jumlah besar, sementara data belajar mencakup informasi yang sensitif mengenai kemampuan, perilaku, dan kadang kondisi psikologis peserta didik. UNESCO (2023) menegaskan perlunya keterbukaan cara kerja sistem, kejelasan pihak yang bertanggung jawab atas keputusan, serta jaminan atas hak peserta didik sebagai prasyarat pemanfaatan kecerdasan buatan dalam pendidikan. Tinjauan kebijakan pendidikan digital lintas negara menempatkan tuntutan tersebut dalam kerangka tata kelola: keandalan teknis tidak memadai tanpa kerangka kelembagaan yang mengatur kepemilikan data, kewenangan penggunaan, dan mekanisme koreksi kesalahan (OECD, 2023). Dari kedua dimensi itulah berkembang gagasan kecerdasan buatan yang berpusat pada manusia, yakni penempatan mesin sebagai pendukung keputusan sementara pertimbangan akhir tetap menjadi kewenangan manusia. Gagasan inilah yang menjadi pijakan konstruksi model pada bagian berikut.

3. Konstruksi Model Intelligent, Humanistic, and Personalized Assessment

Model IHPA dirumuskan sebagai jawaban atas ketegangan antara kapasitas analitik mesin dan tuntutan pertimbangan kemanusiaan dalam evaluasi. Asumsi dasarnya sederhana: keberhasilan evaluasi tidak cukup diukur dari ketepatan angka yang dihasilkan, tetapi dari sumbangannya terhadap perkembangan peserta didik secara intelektual, emosional, sosial, moral, dan spiritual. Karena itu kecerdasan buatan ditempatkan sebagai sistem pendukung

keputusan, sedangkan guru tetap menjadi penafsir akhir atas seluruh bukti yang terkumpul. Berbeda dari model evaluasi yang berpusat pada hasil belajar, IHPA memandang evaluasi sebagai proses yang berlangsung berkelanjutan sehingga setiap aktivitas belajar berpotensi menjadi sumber informasi bagi pendampingan.

a. Landasan Filosofis

Landasan pertama adalah humanisme. Tujuan pendidikan dalam pandangan ini adalah pengembangan seluruh potensi manusia, bukan pembentukan kecakapan kognitif semata, sehingga penilaian harus memperhitungkan keunikan latar belakang, minat, dan perkembangan psikologis setiap peserta didik. Keberhasilan evaluasi diukur pula dari pengaruhnya terhadap motivasi, kepercayaan diri, dan pertumbuhan karakter. Landasan kedua adalah konstruktivisme, yang menegaskan bahwa pengetahuan dibangun melalui interaksi peserta didik dengan lingkungan dan dengan orang lain yang lebih mampu, khususnya pada wilayah perkembangan yang masih memerlukan bantuan (Vygotsky, 1978). Implikasinya bagi evaluasi jelas: yang perlu diketahui bukan hanya apa yang telah dikuasai, tetapi bagaimana penguasaan itu terbentuk dan bantuan seperti apa yang masih dibutuhkan. Landasan ketiga adalah kecerdasan buatan yang berpusat pada manusia, yang menuntut agar teknologi memperkuat martabat manusia, menjunjung keadilan, menjaga privasi, dan menjamin keterbukaan dalam pengambilan keputusan (UNESCO, 2023). Atas dasar ketiga landasan itu, IHPA tidak memberi kewenangan kepada mesin untuk menetapkan hasil evaluasi.

b. Landasan Teoretis

Secara teoretis, IHPA merupakan sintesis empat kerangka. Teori penilaian formatif menyumbang prinsip bahwa evaluasi wajib menghasilkan informasi yang dapat ditindaklanjuti oleh penerimanya (Sadler, 1989; Black & Wiliam, 1998). Teori pembelajaran yang dipersonalisasi menyumbang prinsip kesesuaian antara perlakuan dan karakteristik individu, yang pada tataran instrumen diwujudkan melalui penyesuaian tingkat kesulitan serta ragam bentuk penilaian (Yang et al., 2022; Maier & Klotz, 2022). Analitik pembelajaran menyumbang metode pengumpulan dan pembacaan data belajar secara berkelanjutan sebagai dasar keputusan (Siemens, 2013). Teori pembelajaran yang teratur secara mandiri menyumbang orientasi akhir model ini, yaitu peserta didik yang mampu memantau dan mengarahkan belajarnya sendiri; karena itu hasil evaluasi tidak berhenti pada guru, melainkan dikembalikan kepada peserta didik dalam bentuk yang dapat mereka pahami dan tindaklanjuti (Zimmerman, 1990; Chen et al., 2025).

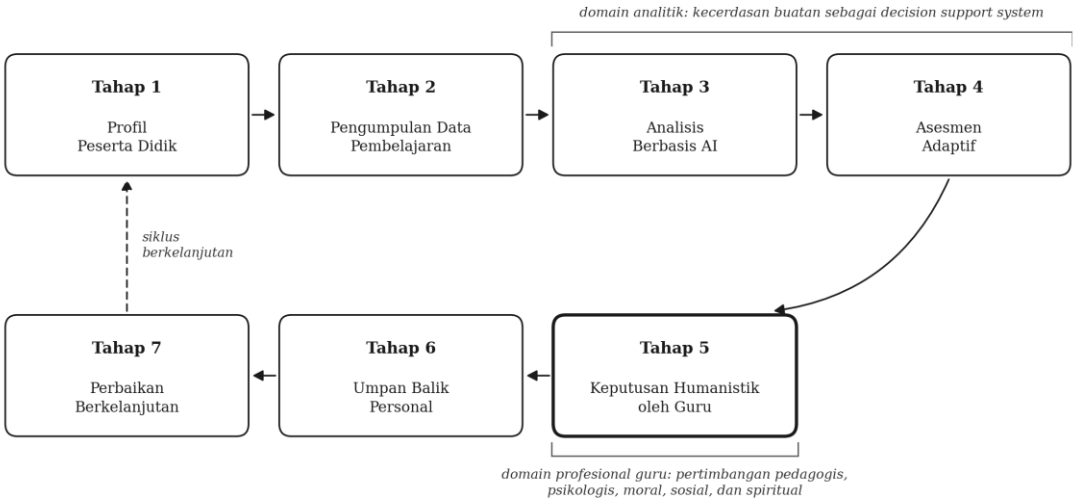
c. Tiga Pilar Karakteristik

Pilar *intelligent* menunjuk pada pemanfaatan pembelajaran mesin, analitik pembelajaran, dan data berskala besar untuk menghasilkan analisis perkembangan belajar secara cepat, berkelanjutan, dan berbasis bukti. Keluaran pilar ini berupa pola belajar, perkiraan risiko kegagalan akademik, dan rekomendasi tindak lanjut, bukan keputusan penilaian. Pilar *humanistic* menempatkan manusia sebagai pusat proses: penilaian mempertimbangkan perkembangan sikap, nilai, empati, dan kemampuan sosial, sementara keputusan akhir berada pada guru yang mengenal konteks kehidupan peserta didiknya. Pilar *personalized* mengarahkan seluruh perangkat penilaian agar sesuai dengan kebutuhan masing-masing peserta didik, baik pada bentuk instrumen, tingkat kesulitan, maupun isi umpan balik. Ketiga pilar tersebut bersifat saling menyeimbangkan: pilar pertama tanpa pilar kedua melahirkan penilaian yang efisien tetapi kehilangan konteks, sedangkan pilar kedua tanpa pilar pertama mengembalikan evaluasi pada keterbatasan pengamatan manual.

d. Sintesis Model Konseptual

Sintesis dari landasan dan pilar di atas menghasilkan model yang bekerja melalui tujuh tahap, sebagaimana disajikan pada Gambar 1. Rangkaian tahap tersebut membentuk siklus, bukan urutan linear yang berhenti pada pelaporan nilai. Dua tahap pertama bersifat

pengumpulan informasi, dua tahap berikutnya merupakan wilayah kerja analitik mesin, tahap kelima merupakan wilayah kewenangan profesional guru, sedangkan dua tahap terakhir mengembalikan hasil evaluasi kepada peserta didik dan kepada perbaikan pembelajaran. Penempatan tahap kelima di antara analisis mesin dan penyampaian umpan balik merupakan ciri yang membedakan IHPA dari model evaluasi berbasis kecerdasan buatan yang menyalurkan keluaran algoritma secara langsung kepada peserta didik.



Gambar 1. Model konseptual *Intelligent, Humanistic, and Personalized Assessment* (IHPA)

4. Implementasi Model IHPA dalam Pembelajaran

a. Tahapan Operasional

Tahap pertama adalah penyusunan profil peserta didik. Guru menghimpun informasi mengenai kemampuan awal, gaya belajar, minat, motivasi, kondisi sosial, serta kebutuhan belajar khusus melalui asesmen diagnostik, angket, wawancara, observasi kelas, dan rekam jejak akademik sebelumnya. Informasi ini berfungsi sebagai data dasar yang menentukan mutu seluruh analisis berikutnya; profil yang dangkal akan menghasilkan rekomendasi yang dangkal pula.

Tahap kedua adalah pengumpulan data pembelajaran yang berlangsung selama proses, bukan pada akhirnya. Sumbernya beragam, meliputi hasil kuis digital, tugas perorangan dan kelompok, portofolio, aktivitas pada sistem pengelolaan pembelajaran, kehadiran, keaktifan diskusi, observasi guru, penilaian antarteman, dan refleksi diri. Keragaman sumber inilah yang menjaga agar gambaran yang terbentuk tidak bergantung pada satu jenis bukti (Possaghi et al., 2025).

Tahap ketiga adalah analisis berbasis kecerdasan buatan. Pada tahap ini sistem mengenali kecenderungan peningkatan atau penurunan capaian, materi yang paling sulit dikuasai, pola kesalahan yang berulang, tingkat konsistensi belajar, serta perkiraan risiko kegagalan akademik, dan menyajikannya dalam bentuk papan pemantau perkembangan belajar (Siemens, 2013). Tahap keempat adalah penyusunan asesmen adaptif. Berpijak pada hasil analisis, guru menetapkan bentuk dan tingkat kesulitan penilaian yang tidak harus seragam: tes objektif, studi kasus, proyek, presentasi, portofolio, unjuk kerja, atau wawancara dipilih menurut karakteristik peserta didik sehingga setiap orang memperoleh kesempatan menunjukkan kompetensinya melalui cara yang paling sesuai (Yang et al., 2022; Ruan & Lu, 2025).

Tahap kelima merupakan inti model ini, yaitu pengambilan keputusan humanistik. Hasil analisis mesin tidak berlaku sebagai keputusan, melainkan sebagai bahan pertimbangan yang ditafsirkan guru dengan memperhitungkan kondisi psikologis, sosial, budaya, moral, dan spiritual peserta didik. Ilustrasinya mudah ditemukan dalam praktik: sistem menandai penurunan capaian seorang peserta didik dan merekomendasikan penurunan tingkat

kesulitan materi, sementara guru mengetahui melalui interaksi langsung bahwa penurunan tersebut berkaitan dengan persoalan keluarga atau kesehatan yang bersifat sementara. Keputusan yang diambil dengan memperhitungkan konteks tersebut lebih adil daripada keputusan yang mengikuti rekomendasi algoritma secara harfiah. Justifikasi teoretisnya kuat, sebab galat prediksi model terbukti tidak terdistribusi merata antarkelompok peserta didik (Baker & Hawn, 2022; Idowu et al., 2024), sehingga penafsiran manusia berfungsi sebagai mekanisme koreksi.

Tahap keenam adalah penyampaian umpan balik yang dipersonalisasi. Masukan tidak berhenti pada angka, tetapi memuat kompetensi yang telah dikuasai, kompetensi yang perlu ditingkatkan, saran strategi belajar, sumber belajar yang sesuai, dan sasaran pembelajaran berikutnya. Kerincian semacam itu hanya bermanfaat bila disampaikan dalam bentuk yang dapat dicerna peserta didik, sehingga penyesuaian bahasa dan waktu penyampaian menjadi bagian dari perancangan umpan balik (Maier & Klotz, 2022; Grion et al., 2026). Tahap ketujuh adalah perbaikan berkelanjutan, yakni pemanfaatan seluruh hasil evaluasi untuk memperbaiki rencana pembelajaran, metode, media, instrumen penilaian, dan layanan akademik. Melalui tahap ini data evaluasi berhenti menjadi arsip administratif dan berubah menjadi bahan pengambilan keputusan mutu.

b. Perbedaan dengan Praktik Evaluasi Konvensional

Perbandingan antara praktik evaluasi konvensional dan model IHPA disajikan pada Tabel 1. Perbandingan tersebut memperlihatkan bahwa perbedaan keduanya tidak terletak pada penggunaan perangkat digital, melainkan pada kedudukan evaluasi dalam keseluruhan pembelajaran.

Tabel 1. Perbandingan praktik evaluasi konvensional dan model IHPA

Aspek	Evaluasi konvensional	Model IHPA
Orientasi	Hasil belajar	Proses sekaligus hasil belajar
Waktu pelaksanaan	Akhir pembelajaran	Berkelanjutan sepanjang pembelajaran
Sumber data	Tes dan tugas tertulis	Beragam sumber data belajar
Instrumen	Seragam untuk seluruh kelas	Adaptif menurut karakteristik individu
Umpan balik	Terbatas dan bersifat umum	Personal dan berkesinambungan
Pengambilan keputusan	Guru berdasarkan hasil tes	Guru dibantu analisis data belajar
Fokus penilaian	Perolehan nilai	Pengembangan kompetensi secara utuh
Pendekatan	Berpusat pada guru	Berpusat pada peserta didik

Kolom kanan tabel memperlihatkan bahwa fungsi evaluasi meluas dari pengukuran hasil belajar menjadi sarana pengembangan peserta didik. Perluasan itu menuntut perubahan kebiasaan kerja: guru tidak lagi menunggu akhir satuan materi untuk memperoleh informasi, dan peserta didik tidak lagi menerima penilaian sebagai putusan yang selesai.

c. Prasyarat Keberhasilan

Prasyarat pertama adalah kompetensi pendidik. Tinjauan sistematis atas pemanfaatan kecerdasan buatan dalam pengajaran dan pengembangan keprofesian menunjukkan bahwa hasil positif hampir selalu bersyarat pada adanya pendampingan yang terstruktur, bukan pada ketersediaan perangkat (Tan et al., 2025). Tinjauan atas kompetensi guru dalam kaitannya dengan penilaian dan analitik memperkuat simpulan itu dengan menyoroti keterampilan membaca data sebagai kemampuan yang paling jarang dimiliki (Baizhanov et al., 2026). Pada konteks pendidikan Islam, pemetaan kompetensi guru dalam mengelola pembelajaran berbasis kecerdasan buatan menemukan kesenjangan serupa antara penguasaan perangkat dan kemampuan memaknai keluarannya secara pedagogis (Saputra et al., 2026).

Prasyarat kedua adalah dukungan kelembagaan berupa infrastruktur digital, sistem pengelolaan pembelajaran yang terpadu, kebijakan perlindungan data peserta didik, dan aturan penggunaan kecerdasan buatan yang jelas. Tanpa kerangka tata kelola semacam itu, pemanfaatan analitik berisiko menimbulkan bias, pelanggaran privasi, dan ketimpangan akses (OECD, 2023; UNESCO, 2023). Implementasi IHPA karena itu perlu dipahami sebagai

perubahan ekosistem evaluasi yang melibatkan guru, peserta didik, satuan pendidikan, dan pembuat kebijakan secara bersamaan, bukan sebagai pengadaan perangkat lunak.

5. Implikasi, Tantangan, dan Prospek Pengembangan

a. Implikasi

Bagi peserta didik, penerapan IHPA menjanjikan pengalaman penilaian yang lebih sesuai dengan kemampuan dan kebutuhannya, sehingga evaluasi berpeluang bergeser dari sumber kecemasan menjadi sarana mengenali potensi diri. Umpan balik yang rinci dan berkesinambungan juga memperkuat kemampuan mengatur belajar sendiri, sebagaimana ditunjukkan meta-analisis atas pengaruh sistem tutor cerdas terhadap kemampuan tersebut (Zhang et al., 2026). Bagi guru, peran bergeser dari pemeriksa hasil menjadi penafsir data sekaligus perancang pembelajaran. Waktu yang sebelumnya terserap pada pekerjaan administratif penilaian dapat dialihkan pada analisis dan pendampingan, dengan syarat kompetensi membaca data terbangun lebih dahulu (Tan et al., 2025).

Bagi kurikulum, IHPA menuntut perancangan yang lentur agar asesmen adaptif, portofolio digital, penilaian unjuk kerja, dan penilaian berbasis proyek dapat ditempatkan dalam satu rangkaian yang utuh. Pada pendidikan tinggi, tuntutan itu sejalan dengan pendekatan pendidikan berbasis capaian, karena seluruh kegiatan penilaian diarahkan untuk mengukur ketercapaian capaian pembelajaran mata kuliah dan capaian pembelajaran program secara berkelanjutan. Bagi kebijakan pendidikan, diperlukan pengaturan mengenai pengelolaan data pendidikan, standar etika penggunaan kecerdasan buatan, dan pengembangan kompetensi digital pendidik, dengan prinsip keterbukaan, keadilan, akuntabilitas, serta perlindungan hak peserta didik sebagai acuan (UNESCO, 2023; OECD, 2023).

b. Tantangan

Tantangan pertama adalah kesenjangan infrastruktur. Tinjauan atas dua dasawarsa penelitian mengenai kesenjangan digital dalam pendidikan menunjukkan bahwa persoalannya tidak berhenti pada ketersediaan perangkat dan jaringan, tetapi mencakup perbedaan kemampuan memanfaatkannya secara produktif (Martin et al., 2024). Pada wilayah dengan keterbatasan akses, penerapan evaluasi berbasis analitik berpotensi memperlebar jarak mutu antarsatuan pendidikan. Tantangan kedua adalah kompetensi digital pendidik yang belum merata, yang menuntut pelatihan berkelanjutan alih-alih sosialisasi sekali jalan (Baizhanov et al., 2026; Saputra et al., 2026).

Tantangan ketiga menyangkut etika dan privasi data. Kebutuhan sistem akan data dalam jumlah besar berhadapan dengan kewajiban melindungi informasi pribadi peserta didik, sehingga setiap satuan pendidikan perlu menetapkan batas pengumpulan, masa simpan, dan kewenangan akses data (UNESCO, 2023). Tantangan keempat adalah bias algoritmik yang menuntut verifikasi manusia atas setiap keluaran yang berimplikasi pada keputusan penilaian (Baker & Hawn, 2022; Idowu et al., 2024; Cui, 2026). Tantangan kelima bersifat kultural, yaitu penolakan terhadap perubahan dari pendidik dan lembaga yang telah lama nyaman dengan pola penilaian konvensional; persoalan integritas akademik yang mengiringi penggunaan kecerdasan buatan generatif memperkuat kehati-hatian tersebut dan perlu dijawab melalui perancangan ulang penilaian, bukan melalui pelarangan (Ali & Maroulis, 2026; Mangubat & Mangubat, 2026).

c. Prospek Pengembangan dalam Konteks Pendidikan Islam

Prospek pengembangan IHPA tampak paling menarik justru pada wilayah yang selama ini dianggap paling sulit diukur, yaitu dimensi akhlak dan spiritualitas. Pendidikan Islam menempatkan keberhasilan belajar tidak hanya pada penguasaan pengetahuan keagamaan, tetapi pada pengamalan nilai dalam kehidupan sehari-hari. Kajian mengenai penguatan karakter islami melalui pembelajaran berbantuan kecerdasan buatan menunjukkan bahwa teknologi dapat berperan pada penyediaan pengalaman belajar yang kontekstual, sementara penilaian atas perkembangan akhlak tetap menuntut pengamatan dan pertimbangan guru (Rahman & Nasution, 2025). Pada pembelajaran Al-Qur'an dan Hadis, pemanfaatan teknologi

ini terbukti menolong pada aspek yang bersifat teknis dan berulang, namun tetap memerlukan penyeliaan atas ketepatan substansi keagamaan (Afif & Nawawi, 2024).

Dengan demikian, pengembangan IHPA pada pendidikan Islam sebaiknya diarahkan pada pembagian kerja yang tegas: mesin menangani pemantauan keteraturan ibadah, konsistensi partisipasi, dan capaian kognitif keagamaan yang dapat direkam secara objektif, sedangkan guru menilai kualitas akhlak, keikhlasan, dan kedalaman penghayatan yang tidak dapat direduksi menjadi data. Pemanfaatan teknologi untuk memperkuat kolaborasi antarpeserta didik dalam pembelajaran keagamaan berbasis digital membuka kemungkinan lain, yaitu penilaian atas sikap sosial yang terbentuk dalam kerja bersama (Samsudin, 2025). Pengembangan instrumen evaluasi digital yang memadukan ranah kognitif, afektif, psikomotorik, dan karakter dengan demikian menjadi agenda yang layak diprioritaskan (Choiriyah, 2025; Lustani et al., 2025). Pada arah itu, IHPA tidak berhenti sebagai model evaluasi berbasis teknologi, tetapi menjadi kerangka penilaian yang selaras dengan tujuan pendidikan Islam maupun tujuan pendidikan nasional.

Kesimpulan

Paradigma evaluasi pembelajaran telah bergerak melalui empat lapisan yang saling melengkapi: dari pengukuran hasil belajar, menuju penilaian yang menopang proses belajar, berlanjut pada penilaian sebagai bagian dari proses belajar itu sendiri, hingga memasuki evaluasi cerdas yang bersandar pada analitik pembelajaran dan pengujian adaptif. Kecerdasan buatan mempercepat pergerakan tersebut dengan menyediakan kemampuan membaca jejak belajar secara berkelanjutan, memperkirakan risiko kegagalan akademik, dan menyusun rekomendasi tindak lanjut. Kemampuan itu tidak menggantikan guru, sebab keputusan evaluatif selalu memuat pertimbangan pedagogis, moral, sosial, dan spiritual yang berada di luar jangkauan algoritma.

Atas dasar sintesis tersebut, artikel ini merumuskan model *Intelligent, Humanistic, and Personalized Assessment* sebagai kerangka konseptual yang memadukan kapasitas analitik mesin dengan otoritas profesional guru. Model ini bekerja melalui tujuh tahap, yaitu penyusunan profil peserta didik, pengumpulan data pembelajaran, analisis berbasis kecerdasan buatan, asesmen adaptif, pengambilan keputusan humanistik, umpan balik yang dipersonalisasi, dan perbaikan berkelanjutan. Penempatan tahap kelima sebagai penyaring atas keluaran algoritma merupakan ciri pembeda utamanya: kecerdasan buatan menghasilkan bahan pertimbangan, sedangkan guru menghasilkan keputusan. Pada konteks pendidikan Islam, kerangka ini dapat diperluas dengan pembagian kerja antara pemantauan yang dapat direkam secara objektif dan penilaian atas akhlak serta penghayatan nilai yang menuntut kehadiran pendidik.

Penerapan model ini menjanjikan evaluasi yang lebih adaptif, inklusif, dan berbasis bukti, namun keberhasilannya bersyarat pada kesiapan infrastruktur, kompetensi membaca data di kalangan pendidik, perlindungan data peserta didik, dan kejelasan tata kelola penggunaan teknologi. Keterbatasan kajian ini terletak pada sifatnya yang konseptual, sehingga validitas, keefektifan, dan kepraktisan model belum diuji secara empiris. Pengujian melalui penelitian pengembangan dan penelitian eksperimen pada berbagai jenjang pendidikan, disertai penyusunan instrumen evaluasi digital yang memadukan ranah kognitif, afektif, psikomotorik, dan karakter, merupakan agenda yang perlu ditempuh untuk menguji daya guna model ini dalam praktik.

Daftar Pustaka

Afif, N., & Nawawi, A. (2024). Optimalisasi pengajaran Al-Qur'an dan Hadis melalui teknologi kecerdasan buatan: Tantangan dan strategi integrasi. *EduInovasi: Journal of Basic Educational Studies, 4*(3), 1829–1848. doi: 10.47467/edu.v4i3.5156

Ali, R., & Maroulis, J. (2026). Assessment validity in the age of generative artificial intelligence: A critical review. *Assessment & Evaluation in Higher Education*, 1–16. doi: 10.1080/02602938.2026.2683046

Anbreen, T., Özturan, T., Shrestha, P., & Maqsood, A. (2026). Human vs. generative artificial intelligence in writing assessment: Investigating feedback alignment, score validity, and teacher agency. *Educational Assessment, Evaluation and Accountability*. doi: 10.1007/s11092-026-09486-z

Baizhanov, N., Sharimbayev, B., Churbanova, Z., Khaidarova, A., & Shinetova, L. (2026). Artificial intelligence and teacher competence: A scoping review of assessment, analytics, and professional development. *Frontiers in Artificial Intelligence, 9*. doi: 10.3389/frai.2026.1901449

Baker, R. S., & Hawn, A. (2022). Algorithmic bias in education. *International Journal of Artificial Intelligence in Education, 32*(4), 1052–1092. doi: 10.1007/s40593-021-00285-9

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice, 5*(1), 7–74. doi: 10.1080/0969595980050102

Chen, S., Zhang, L., & Li, A. W. (2025). Leveraging self-regulated learning to support university students' feedback literacy in blended learning environments. *Assessment & Evaluation in Higher Education, 51*(1), 174–193. doi: 10.1080/02602938.2025.2546349

Choiriyah, M. (2025). Evaluasi pembelajaran berbasis digital: Sistematis terhadap strategi media yang efektif. *Al-Qiyadah, 2*(1), 64–72. doi: 10.64481/v1tnrv30

Cooper, G., Ilich, K., & Sturrock, K. (2026). Using generative artificial intelligence to reimagine feedback in higher education: A collaborative autoethnography. *Assessment & Evaluation in Higher Education*, 1–15. doi: 10.1080/02602938.2026.2653889

Cui, Y. (2026). Artificial intelligence and human assessment consistency and bias in higher education academic writing evaluation. *Discover Artificial Intelligence, 6*(1). doi: 10.1007/s44163-026-01286-0

Du, Y., & Yuan, Y. (2026). Epistemic dependence in AI-mediated learning within higher education: A framework for student judgement and responsibility. *Assessment & Evaluation in Higher Education*, 1–15. doi: 10.1080/02602938.2026.2732250

Earl, L. M. (2003). *Assessment as learning: Using classroom assessment to maximize student learning*. Corwin Press.

Evangelista, E., & Salman Bukhari, S. M. (2026). From ethical principles to executable governance: A policy-as-code framework for trustworthy AI in higher education. *Computers and Education: Artificial Intelligence, 10*, 100594. doi: 10.1016/j.caeai.2026.100594

Gray, A., Rahat, A., Crick, T., & Lindsay, S. (2024). A Bayesian active learning approach to comparative judgement within education assessment. *Computers and Education: Artificial Intelligence, 6*, 100245. doi: 10.1016/j.caeai.2024.100245

Grion, V., Doria, B., Agostini, D., & Slaviero, G. (2026). Artificial intelligence and feedback in university education: Effectiveness and student perceptions. *Assessment & Evaluation in Higher Education*, 1–20. doi: 10.1080/02602938.2026.2697962

Idowu, J. A., Koshiyama, A. S., & Treleaven, P. (2024). Investigating algorithmic bias in student progress monitoring. *Computers and Education: Artificial Intelligence, 7*, 100267. doi: 10.1016/j.caeai.2024.100267

Khulbe, M., & Tammets, K. (2023). Mediating teacher professional learning with a learning analytics dashboard and training intervention. *Technology, Knowledge and Learning, 28*(3), 981–998. doi: 10.1007/s10758-023-09642-0

Laidlaw, P. (2026). Supporting assessment redesign in the age of AI: A critical practice-based framework for equitable change. *Teaching in Higher Education*, 1–24. doi: 10.1080/13562517.2026.2670362

Lustani, A. D., Fadil, A., Soraya, S., & Pratama, A. F. (2025). Pemanfaatan media pembelajaran berbasis AI: Solusi cerdas untuk pendidikan masa kini. *Jurnal Pemikiran Pendidikan dan Keguruan, 1*(1), 70–76. doi: 10.59966/0cw1e858

Maier, U., & Klotz, C. (2022). Personalized feedback in digital learning environments: Classification framework and literature review. *Computers and Education: Artificial Intelligence, 3*, 100080. doi: 10.1016/j.caeai.2022.100080

Mangubat, F. M., & Mangubat, F. N. (2026). From diffusion to negotiation: Generative AI and academic integrity in higher education assessment. *SN Social Sciences, 6*(9). doi: 10.1007/s43545-026-01667-3

Martin, F., Ceviker, E., & Gezer, T. (2024). From digital divide to digital equity: Systematic review of two decades of research on educational digital divide factors, dimensions, and interventions. *Journal of Research on Technology in Education, 58*(2), 396–421. doi: 10.1080/15391523.2024.2425442

Nasihuddin, M. (2024). Peran kecerdasan buatan terhadap transformasi pendidikan di era digital. *Salimiya: Jurnal Studi Ilmu Keagamaan Islam, 5*(4), 410–418. doi: 10.58401/salimiya.v5i4.1919

OECD. (2023). *OECD digital education outlook 2023: Towards an effective digital education ecosystem*. OECD Publishing.

Possaghi, I., Vesin, B., Zhang, F., Sharma, K., Knudsen, C., Bjørkum, H., & Papavlasopoulou, S. (2025). Integrating multi-modal learning analytics dashboard in K-12 education: Insights for enhancing orchestration and teacher decision-making. *Smart Learning Environments, 12*(1). doi: 10.1186/s40561-025-00410-4

Rahman, A., & Nasution, B. (2025). Strategies to strengthen Islamic character through Islamic education learning based on artificial intelligence. *DAYAH: Journal of Islamic Education, 8*(2), 120–130. doi: 10.22373/jie.v8i2.33627

Ruan, S., & Lu, K. (2025). Adaptive deep reinforcement learning for personalized learning pathways: A multimodal data-driven approach with real-time feedback optimization. *Computers and Education: Artificial Intelligence, 9*, 100463. doi: 10.1016/j.caeai.2025.100463

Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science, 18*(2), 119–144. doi: 10.1007/BF00117714

Samsudin, U. (2025). Exploration of artificial intelligence (AI) in increasing student collaboration in digital-based Islamic education learning. *Al-Hayat: Journal of Islamic Education, 9*(1), 216–230. doi: 10.35723/ajie.v9i1.106

Saputra, A., Neliwati, Pristiawan, E., & Siregar, P. (2026). Islamic education teachers' competencies in managing artificial intelligence-based learning. *IJGIE: International Journal of Graduate of Islamic Education, 7*(2), 729–743. doi: 10.37567/ijgie.v7i2.4880

Siemens, G. (2013). Learning analytics: The emergence of a discipline. *American Behavioral Scientist, 57*(10), 1380–1400. doi: 10.1177/0002764213498851

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*, 100355. doi: 10.1016/j.caeai.2024.100355

Tyler, R. W. (1949). *Basic principles of curriculum and instruction*. University of Chicago Press.

UNESCO. (2023). *Guidance for generative AI in education and research*. UNESCO.

Venetsanos, D. T. (2026). A tripartite feedback framework for AI-assisted assessment of complex reports in higher education. *AI in Education, 2*(3), 26. doi: 10.3390/aieduc2030026

Villegas-Ch, W., Buenaño-Fernández, D., Maldonado Navarro, A., & Mera-Navarrete, A. (2025). Adaptive intelligent tutoring systems for STEM education: Analysis of the learning impact and effectiveness of personalized feedback. *Smart Learning Environments, 12*(1). doi: 10.1186/s40561-025-00389-y

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.

Wiggins, G. P. (1998). *Educative assessment: Designing assessments to inform and improve student performance*. Jossey-Bass.

Yang, A. C. M., Flanagan, B., & Ogata, H. (2022). Adaptive formative assessment system based on computerized adaptive testing and the learning memory cycle for personalized learning. *Computers and Education: Artificial Intelligence, 3*, 100104. doi: 10.1016/j.caeai.2022.100104

Yasfi, A. (2026). Future direction of artificial intelligence in higher education assessment: A systematic review. *Cogent Education, 13*(1). doi: 10.1080/2331186X.2026.2728183

Zhang, K., Zhou, Y., Zhang, C., & Chen, J. (2026). The effects of intelligent tutoring systems on self-regulated learning: A meta-analysis. *Educational Psychology Review, 38*(1). doi: 10.1007/s10648-026-10199-4

Zimmerman, B. J. (1990). Self-regulated learning and academic achievement: An overview. *Educational Psychologist, 25*(1), 3–17. doi: 10.1207/s15326985ep2501_2