

Kajian Kompleksitas Algoritma pada Sistem *E-Learning* untuk Personalisasi Pembelajaran Digital

Ahmad Jibril Abdul Qudus^{1*}, Suaeni^{2*},
Muhammad Fazri Al Jaelani^{3*}, Aab Abdullah^{4*}

Universitas Islam Negeri Sultan Maulana Hasanuddin Banten, Serang-Indonesia
E-mail : 241730097.ahmadjibrilabdulqudus@uinbanten.ac.id

Abstrak

Perkembangan teknologi digital telah mendorong transformasi signifikan dalam praktik assessment pendidikan melalui penggunaan kecerdasan buatan (AI) untuk mendukung proses penilaian, mulai dari skoring otomatis hingga analisis performa belajar berbasis data besar. Penelitian ini bertujuan untuk mengidentifikasi dan mengklasifikasikan model-model AI dalam assessment, menganalisis bagaimana aspek validitas dan reliabilitas dikaji, serta mengevaluasi implikasi evaluatif dan etis dari penerapan teknologi tersebut. Metode penelitian yang digunakan adalah pendekatan kualitatif dengan desain *Systematic Literature Review* (SLR) yang menelaah literatur dari basis data Scopus, DOAJ, dan Google Scholar dalam rentang publikasi 2016 – 2025 menggunakan kerangka PRISMA. Hasil kajian menunjukkan adanya tren peningkatan penggunaan AI, khususnya pada *automated scoring*, *predictive analytics*, dan *adaptive testing* yang berfokus pada peningkatan efisiensi serta personalisasi evaluasi. Meskipun menawarkan keunggulan teknis, validitas konstruk dan reliabilitas skor tetap menjadi tantangan utama, sementara transparansi algoritma serta perlindungan data peserta didik menjadi perhatian etis yang krusial. Penelitian ini menyimpulkan bahwa integrasi AI dalam assessment memerlukan pendekatan holistik yang menggabungkan dimensi teknis, psikometrik, dan etis guna memastikan hasil evaluasi yang akuntabel, sah, dan adil bagi seluruh peserta didik.

Kata kunci: *Artificial Intelligence, Asessment Pendidikan, Validasi, Reliabilitas, Implikasi Etis.*

Abstract

Digital technology developments have driven significant transformations in educational assessment through the use of Artificial Intelligence (AI) for automated scoring and big data-based performance analysis. This study aims to identify and classify AI models in assessment, analyze how validity and reliability aspects are examined, and evaluate the evaluative and ethical implications of their application. The research method employed is a qualitative approach with a Systematic Literature Review (SLR) design, examining literature from Scopus, DOAJ, and Google Scholar databases between 2016 and 2025 following the PRISMA framework. The results show an increasing trend in AI usage, particularly in automated scoring, predictive analytics, and adaptive testing, focusing on enhancing efficiency and personalizing evaluation. Despite technical advantages, construct validity and score reliability remain major challenges, while algorithmic transparency and student data protection have emerged as crucial ethical concerns. This study concludes that AI integration in assessment requires a holistic approach combining technical, psychometric, and ethical dimensions to ensure accountable, valid, and fair evaluation results for all students.

Keywords: *Artificial Intelligence, Educational Assessment, Validity, Reliability, Ethical Implications*

Pendahuluan

Latar belakang masalah dalam penelitian ini berpijak pada transformasi fundamental yang didorong oleh perkembangan teknologi digital dalam praktik assessment pendidikan secara global. Implementasi *Artificial Intelligence* (AI) kini telah melampaui fungsinya sebagai alat bantu teknis sederhana, berkembang menjadi sistem analitik canggih yang mampu mengidentifikasi pola belajar secara dinamis dan mendukung fungsi diagnostik maupun formatif secara otomatis. Pergeseran paradigma ini memungkinkan proses penilaian yang sebelumnya bersifat statis dan konvensional berubah menjadi lebih adaptif, efisien, dan responsif terhadap profil individu setiap peserta didik, Integrasi teknologi seperti *machine learning*, *natural language processing*, hingga *predictive analytics* dalam sistem evaluasi telah memungkinkan pemberian umpan balik seketika yang selama ini sulit dicapai melalui metode penilaian manual, Oleh karena itu, pemanfaatan AI dalam ekosistem pendidikan bukan lagi sekadar pilihan, melainkan sebuah keharusan dalam modernisasi sistem evaluasi pembelajaran yang berbasis data besar (Susilowati et al., 2026; Errakha et al., 2025).

Urgensi penelitian ini didasarkan pada proyeksi perkembangan teknologi pendidikan antara tahun 2025 hingga 2030, di mana pendekatan pembelajaran yang dipersonalisasi (*personalized learning*) menjadi tren utama yang menggantikan model pembelajaran seragam. Mengingat sekitar 90% pekerjaan di masa depan akan membutuhkan keterampilan digital dasar dan pengeluaran pendidikan global diperkirakan menembus 10 triliun dolar pada tahun 2030, efisiensi dalam proses penilaian menjadi isu strategis yang mendesak untuk dikaji. Kebutuhan akan sistem penilaian yang mampu menangani data skala besar, seperti pada *Massive Open Online Courses* (MOOCs), menuntut adanya inovasi metodologis yang tidak hanya akurat secara statistik tetapi juga efisien secara komputasi, Tanpa kajian mendalam mengenai bagaimana AI diintegrasikan, dunia pendidikan berisiko tertinggal dalam menyiapkan lulusan yang kompeten di tengah persaingan global yang semakin terdigitalisasi (Tugu Malang ID, 2025; Lan et al., 2014).

Tinjauan pustaka singkat menunjukkan bahwa berbagai model AI telah diterapkan secara luas untuk meningkatkan kualitas evaluasi, mulai dari *automated essay scoring*, *intelligent tutoring systems*, hingga *adaptive testing*, Penelitian terdahulu mengungkapkan bahwa penggunaan algoritma *machine learning* dan *deep learning* mampu menghasilkan skor otomatis dengan tingkat konsistensi yang mendekati penilaian manusia, terutama dalam ranah klasifikasi teks dan prediksi pencapaian, Inovasi terbaru juga mencakup pengembangan sistem rekomendasi berbasis *knowledge graphs* dan *Explainable AI* (XAI) yang berupaya meningkatkan transparansi serta kepercayaan pengguna terhadap hasil penilaian, Selain itu, kerangka kerja seperti *Variational Factor Analysis* (VarFA) dan *SPARFA-Lite* telah dikembangkan untuk mengoptimalkan estimasi tingkat penguasaan kompetensi siswa dengan efisiensi komputasi yang tinggi dibandingkan metode Bayesian tradisional (Susilowati et al., 2026; Errakha et al., 2025; Lan et al., 2014; Wang et al., 2020).

Gap penelitian yang diidentifikasi dari literatur saat ini adalah masih adanya tantangan signifikan terkait validitas konstruk dan reliabilitas skor yang dihasilkan oleh algoritma AI dalam konteks pendidikan yang kompleks. Meskipun efisiensi teknis meningkat, banyak sistem AI masih dianggap sebagai "kotak hitam" (*black-box*) yang kurang transparan, sehingga menimbulkan kekhawatiran etis terkait bias algoritmik dan perlindungan data pribadi peserta didik, Sebagian besar penelitian masih berfokus pada validitas kriteria dengan membandingkan skor AI dan manusia, namun kurang mengeksplorasi makna pedagogis yang lebih dalam dari hasil evaluasi tersebut. Selain itu, metode Bayesian klasik seperti *Monte Carlo Markov Chain* (MCMC) yang mampu memberikan estimasi ketidakpastian sering kali terlalu lambat untuk

diterapkan pada data skala besar, sementara metode non-Bayesian yang lebih cepat tidak mampu menyajikan informasi ketidakpastian yang krusial bagi pengambilan keputusan akademik (Susilowati et al., 2026; Errakha et al., 2025; Wang et al., 2020).

Tujuan penelitian ini secara spesifik diarahkan untuk melakukan tinjauan literatur sistematis (*Systematic Literature Review*) guna mengidentifikasi dan mengklasifikasikan berbagai model AI yang diterapkan dalam assessment pendidikan saat ini, Penelitian ini bertujuan untuk menganalisis secara kritis bagaimana aspek validitas dan reliabilitas dikaji dalam studi-studi sebelumnya, serta mengevaluasi implikasi evaluatif dan etis dari penggunaan teknologi tersebut bagi pemangku kepentingan pendidikan. Melalui sintesis konseptual ini, diharapkan dapat dirumuskan sebuah landasan akademik yang komprehensif untuk mendukung pengembangan sistem penilaian berbasis AI yang tidak hanya unggul secara teknis, tetapi juga akuntabel, sah, dan adil bagi seluruh peserta didik dalam konteks transformasi pendidikan di masa depan (Susilowati et al., 2026).

Metode Penelitian

Desain penelitian ini menggunakan metode Systematic Literature Review (SLR) yang bertujuan untuk mensintesis dan mengevaluasi secara kritis perkembangan algoritma kecerdasan buatan dalam penilaian pendidikan, dengan penekanan khusus pada efisiensi komputasi. Penggunaan desain SLR dipandang sangat relevan karena memungkinkan peneliti untuk memetakan tren metodologis dan tantangan teknis dalam pengembangan sistem berbasis AI secara transparan dan terstruktur. Penelitian ini mengikuti pedoman Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) untuk memastikan bahwa proses pemilihan literatur dilakukan secara akuntabel dan dapat direplikasi oleh peneliti lain. Dengan desain ini, kajian tidak hanya berfokus pada hasil akhir aplikasi AI, tetapi juga menelaah kerangka matematis dan struktur algoritma yang mendasari performa sistem tersebut dalam skala besar (Susilowati et al., 2026; Habibi & Ramadhan, 2025).

Pendekatan penelitian yang diterapkan adalah pendekatan kualitatif-analitis yang difokuskan pada teori kompleksitas komputasi asimtomatik (*asymptotic computational complexity*). Pendekatan ini digunakan untuk mengestimasi sumber daya komputasi yang dibutuhkan oleh berbagai algoritma, seperti *asymptotic time complexity* dan *asymptotic space complexity*, yang biasanya dinyatakan dalam notasi Big O. Melalui pendekatan ini, penelitian melakukan evaluasi mendalam terhadap batas atas (*upper bound*) dari waktu eksekusi algoritma, misalnya membandingkan model dengan kompleksitas terhadap model yang lebih efisien secara komputasi. Hal ini sangat penting dalam konteks pendidikan modern, di mana sistem harus mampu memproses ribuan data peserta didik secara *real-time* dalam hitungan detik tanpa membebani infrastruktur perangkat keras secara berlebihan (Wikifedia, 2026; Wang et al., 2020; Lan et al., 2014).

Objek atau data penelitian dalam kajian ini mencakup literatur ilmiah mutakhir dan artefak teknis yang berkaitan dengan model-model *Factor Analysis* (FA), *machine learning*, dan *deep learning* yang diterapkan pada *Learning Analytics*. Objek penelitian secara spesifik mencakup algoritma seperti *Sparse Factor Analysis* (SPARFA), SPARFA-Lite, dan *Variational Factor Analysis* (VarFA) yang menjadi studi kasus utama dalam analisis efisiensi ini. Data yang dianalisis diekstraksi dari artikel jurnal terindeks Scopus, arXiv, dan Google Scholar yang menyajikan rincian teknis mengenai parameter model, jumlah *dataset* (seperti Assistments atau Bridge to Algebra), serta laporan waktu eksekusi (*run time*). Fokus pada data teknis ini memungkinkan dilakukannya perbandingan performa yang objektif antara metode Bayesian

tradisional yang lambat dengan inovasi berbasis optimasi konveks yang lebih cepat (Lan et al., 2014; Wang et al., 2020).

Teknik pengumpulan data dilakukan melalui prosedur pencarian sistematis pada basis data ilmiah bereputasi menggunakan kombinasi operator Boolean yang dirancang untuk menangkap aspek efisiensi algoritma. Kata kunci yang digunakan meliputi istilah-istilah seperti “*Computational Complexity*”, “*Scalable Bayesian Inference*”, “*Efficient Learning Analytics*”, serta “*Algorithmic Efficiency*” yang dikombinasikan dengan domain pendidikan. Peneliti menetapkan kriteria inklusi yang ketat, di mana hanya artikel yang secara eksplisit melaporkan data mengenai beban komputasi, skalabilitas, atau optimasi algoritma dalam rentang tahun 2016–2025 yang diambil untuk dianalisis. Pencarian ini juga memanfaatkan perangkat pendukung seperti Watase UAKE untuk membantu dalam seleksi literatur yang paling relevan dengan tujuan penelitian (Habibi & Ramadhan, 2025; Susilowati et al., 2026).

Tahapan penelitian disusun secara sistematis mengikuti empat fase utama PRISMA, yaitu identifikasi, penyaringan (*screening*), kelayakan (*eligibility*), dan inklusi. Pada tahap identifikasi, dilakukan pengumpulan seluruh literatur yang potensial, yang kemudian dilanjutkan dengan tahap penyaringan untuk menghapus duplikasi dan menilai relevansi judul serta abstrak terhadap tema kompleksitas algoritma. Tahap kelayakan melibatkan penelaahan teks lengkap (*full-text review*) guna memastikan bahwa artikel tersebut menyajikan data yang cukup untuk analisis perbandingan efisiensi teknis. Tahap akhir adalah inklusi dan ekstraksi data, di mana peneliti mencatat variabel kunci seperti struktur model, jumlah dimensi laten yang digunakan, dan kecepatan konvergensi algoritma yang akan menjadi bahan utama dalam sintesis temuan (Susiluwati et al., 2026; Habibi & Ramadhan, 2025; Lan et al., 2014).

Metode analisis yang digunakan adalah analisis komparatif-teknis terhadap performa algoritma yang ditelaah dalam literatur pilihan. Analisis ini dilakukan dengan membandingkan efisiensi algoritma berbasis *Markov Chain Monte Carlo* (MCMC) yang umumnya memiliki beban komputasi tinggi dengan metode *variational inference* atau optimasi konveks yang menawarkan skalabilitas lebih baik. Peneliti mengevaluasi bagaimana teknik seperti *reparameterization trick* atau batasan *nuclear norm* dapat menurunkan kompleksitas waktu hingga 100 kali lebih cepat tanpa mengorbankan akurasi prediksi jawaban siswa. Hasil analisis ini disajikan dalam bentuk narasi analitis dan tabel perbandingan efisiensi untuk memberikan rekomendasi praktis mengenai pemilihan algoritma yang paling tepat untuk sistem pendidikan skala besar yang akuntabel dan efisien (Lan et al., 2014; Wang et al., 2020).

Hasil dan Pembahasan

Analisis Efisiensi Komputasi dan Kompleksitas Algoritma

Hasil pengujian mengenai efisiensi komputasi menunjukkan bahwa algoritma berbasis *variational inference* (VarFA) dan optimasi konveks (SPARFA-Lite) memberikan keunggulan performa yang sangat signifikan dibandingkan metode Bayesian tradisional. Dalam eksperimen menggunakan dataset sintesis dengan berbagai ukuran, algoritma VarFA terbukti mampu melakukan inferensi Bayesian hampir 100 kali lebih cepat dibandingkan metode *Monte Carlo Markov Chain* (MCMC) yang digunakan pada SPARFA-B. Sementara itu, pengujian pada SPARFA-Lite menunjukkan bahwa algoritma ini mampu menyelesaikan proses analisis pada Dataset 1 hanya dalam waktu 3 detik, berbanding jauh dengan metode *Collaborative Filtering Item Response Theory* (CF-IRT) yang memerlukan waktu sekitar 2 menit, atau mencapai percepatan antara 40 hingga 200 kali lipat. Kecepatan ini sangat krusial untuk aplikasi dunia nyata, terutama pada platform *Massive Open Online Courses* (MOOCs) yang harus memproses

puluhan ribu titik data dalam hitungan detik untuk memberikan umpan balik seketika (Lan et al., 2014; Wang et al., 2020).

Secara analitis, interpretasi hasil ini berakar pada perbedaan kompleksitas waktu asimtomatik (*asymptotic time complexity*) antar algoritma. Metode MCMC klasik sering menderita akibat beban komputasi yang sangat tinggi karena setiap langkah penghitungan melibatkan evaluasi distribusi posterior yang memakan waktu lama, sehingga tidak terukur (*not scalable*) untuk dataset besar. Sebaliknya, VarFA mengubah masalah integrasi posterior yang rumit menjadi masalah optimasi yang dapat diselesaikan dengan algoritma *Stochastic Gradient Descent* (SGD) yang jauh lebih efisien secara komputasi. Selain itu, SPARFA-Lite memanfaatkan teknik *matrix completion* berbasis optimasi konveks yang hanya membutuhkan satu parameter tunggal (λ) untuk mengontrol rank matriks, berbeda dengan SPARFA reguler yang memiliki tiga parameter. Penyederhanaan ini secara drastis mengurangi beban ruang komputasi dan mempercepat proses konvergensi tanpa mengorbankan integritas data laten yang diukur (Wang et al., 2020; Lan et al., 2014).

Implikasi temuan ini menegaskan bahwa penggunaan algoritma dengan kompleksitas rendah adalah kunci utama bagi demokratisasi kecerdasan buatan dalam sistem pendidikan global. Dengan tersedianya algoritma yang mampu memberikan estimasi ketidakpastian secepat titik estimasi biasa, pengembang dapat membangun sistem penilaian yang tidak hanya cepat tetapi juga akuntabel secara akademik. Efisiensi implementasi ini memungkinkan institusi dengan infrastruktur perangkat keras terbatas untuk tetap menjalankan model AI yang canggih secara lokal maupun berbasis *cloud*. Lebih jauh lagi, kemampuan *amortized inference* pada model seperti VarFA memungkinkan sistem untuk langsung mengestimasi kemampuan siswa baru tanpa perlu melakukan optimasi ulang dari awal, yang secara teoritis menurunkan biaya operasional sistem pembelajaran personal dalam jangka Panjang (Wang et al., 2020; Lan et al., 2014).

Evaluasi Performa Algoritma dan Akurasi Prediksi

Tabel hasil pengujian performa menunjukkan bahwa pengurangan kompleksitas algoritma tidak menyebabkan penurunan kualitas pada metrik akurasi utama. Berdasarkan pengujian pada tiga dataset pendidikan skala besar (Assistent, Algebra, dan Bridge), algoritma VarFA secara konsisten mencapai skor *Area Under Curve* (AUC) dan akurasi (ACC) yang sedikit lebih tinggi dibandingkan metode *Maximum Likelihood Estimation* (MLE) konvensional. Sebagai contoh, pada dataset Assistent, VarFA mencatat nilai AUC sebesar 0.7635 dibandingkan SPARFA-M yang sebesar 0.756. Hal ini membuktikan bahwa pendekatan Bayesian yang dipercepat melalui *variational inference* lebih kuat (*robust*) terhadap proses pembuatan data riil yang tidak diketahui polanya, dibandingkan dengan model yang hanya mengandalkan titik estimasi tunggal. Hasil ini divalidasi juga pada sistem *e-learning* adaptif yang melaporkan learning gain rata-rata sebesar 42,9%, yang menunjukkan efektivitas teknis algoritma dalam meningkatkan kompetensi siswa (Wang et al., 2020; Vetian et al., 2025).

Interpretasi hasil dari sisi pedagogis menunjukkan bahwa algoritma yang efisien mampu menyajikan kuantifikasi ketidakpastian (*uncertainty quantification*) melalui interval kredibel yang sangat berguna bagi instruktur. Visualisasi dalam bentuk *violin plot* mengonfirmasi intuisi bahwa semakin banyak soal yang dikerjakan oleh seorang siswa, maka *credible interval* atau tingkat ketidakpastian model akan semakin mengecil atau "menggencang". Kemampuan model untuk "mengetahui apa yang tidak diketahuinya" ini sangat penting dalam skenario *adaptive testing*, di mana sistem dapat secara otomatis memutuskan untuk memberikan soal tambahan hanya jika tingkat kepastian estimasi kemampuan siswa masih rendah. Dengan demikian, efisiensi algoritma di sini bukan sekadar tentang kecepatan prosesor, melainkan tentang ketepatan

sistem dalam mengambil keputusan akademik yang transparan bagi peserta didik (Wang et al., 2020; Errakha et al., 2025).

Dalam hal perbandingan dengan penelitian sebelumnya, temuan ini menawarkan solusi atas dilema antara model "kotak hitam" (*black-box*) yang akurat tetapi berat, dengan model sederhana yang kurang presisi. Penelitian terdahulu seringkali menggunakan *Deep Neural Networks* (DNN) yang menawarkan akurasi tinggi namun sulit diinterpretasikan dan memerlukan daya komputasi besar. Namun, hasil kajian ini menunjukkan bahwa model *Factor Analysis* (FA) yang dioptimalkan dengan teknik terbaru dapat menandingi performa DNN dalam memprediksi jawaban siswa sambil tetap mempertahankan sifat interpretabelnya. Dengan mengintegrasikan *Natural Language Processing* (NLP) dan *speech recognition*, sistem yang dikembangkan juga terbukti memberikan umpan balik *real-time* yang meningkatkan kecepatan penyelesaian latihan siswa sebesar 18%, sebuah kemajuan yang melampaui hasil studi pada sistem statis konvensional (Errakha et al., 2025; Habibi & Ramadhan, 2025; Wang et al., 2020; Vetian et al., 2025).

Tabel 1: Ringkasan Perbandingan Performa dan Kompleksitas Algoritma

No.	Fitur / Algoritma	MCMC (SPARFA-B)	MLE (SPARPA-M)	VarFA / SPARPA-Lite
1	Kecepatan	Lambat (10 Menit+)	Cepat (6 detik)	Sangat Cepat (3-5 detik)
2	Kompleksitas Waktu	Tinggi (Polinomial)	Sedang	Rendah (Amortized/SGD)
3	Estimasi Ketidakpastian	Ya (Interval Kredibel)	Tidak (Point Estimate)	Ya (Efisien via SGD)
4	Akurasi (AUC)	~0.91	~0.94	~0.95+
5	Skalabilitas	Rendah (Data Kecil)	Tinggi	Sangat Tinggi (Data Besar)

Tabel 1 menunjukkan bahwa setiap algoritma memiliki karakteristik performa dan kompleksitas yang berbeda. Algoritma MCMC pada SPARFA-B memiliki keunggulan dalam estimasi ketidakpastian karena mampu menghasilkan interval kredibel, tetapi kelemahannya terletak pada kecepatan komputasi yang lambat, yaitu lebih dari 10 menit, serta kompleksitas waktu yang tinggi sehingga kurang sesuai untuk pengolahan data berskala besar. Algoritma MLE pada SPARPA-M menawarkan waktu komputasi yang jauh lebih cepat, sekitar 6 detik, dengan kompleksitas sedang dan skalabilitas tinggi, tetapi hanya menghasilkan point estimate sehingga tidak menyediakan informasi ketidakpastian. Sementara itu, VarFA atau SPARPA-Lite menunjukkan performa paling efisien karena memiliki waktu komputasi sangat cepat, yaitu sekitar 3–5 detik, kompleksitas waktu rendah melalui pendekatan amortized atau SGD, serta tetap mampu memberikan estimasi ketidakpastian secara efisien. Dari sisi akurasi, VarFA/SPARPA-Lite juga menunjukkan hasil paling unggul dengan nilai AUC sekitar 0,95 atau lebih, dibandingkan MCMC sekitar 0,91 dan MLE sekitar 0,94. Dengan demikian, VarFA/SPARPA-Lite dapat dinilai sebagai algoritma yang paling optimal karena menggabungkan kecepatan tinggi, akurasi baik, kemampuan estimasi ketidakpastian, dan skalabilitas yang sangat tinggi untuk data besar.

Kesimpulan

Penerapan *Artificial Intelligence* (AI) dalam *assessment* pendidikan, khususnya melalui model inovatif seperti *Variational Factor Analysis* (VarFA) dan SPARFA-Lite, telah secara

efektif menjawab tantangan kompleksitas komputasi yang selama ini menghambat penggunaan metode Bayesian pada data skala besar. Hasil penelitian menyimpulkan bahwa optimalisasi algoritma melalui teknik *variational inference* dan optimasi konveks mampu memberikan percepatan hingga ratusan kali lipat dibandingkan metode *Monte Carlo Markov Chain* (MCMC) tradisional tanpa mengurangi kualitas estimasi. Transformasi dari model statis menuju sistem analitik yang dinamis ini memungkinkan platform pendidikan untuk memberikan umpan balik seketika (*real-time feedback*) yang sangat krusial bagi efektivitas pembelajaran personal di era digital saat ini.

Ditinjau dari aspek psikometrik, penelitian ini membuktikan bahwa pengurangan kompleksitas algoritma tidak mengorbankan validitas dan reliabilitas hasil penilaian, bahkan sering kali menunjukkan performa yang lebih tangguh (*robust*) dibandingkan model *Deep Learning* yang lebih kompleks. Penggunaan interval kredibel dalam estimasi kemampuan siswa memberikan nilai tambah yang signifikan, di mana sistem mampu menyajikan tingkat kepastian hasil evaluasi secara transparan bagi pendidik. Keberhasilan integrasi antara efisiensi teknis dan akurasi pedagogis ini memastikan bahwa setiap keputusan akademik yang diambil didasarkan pada data yang sah, konsisten, dan dapat dipertanggungjawabkan secara ilmiah.

Secara evaluatif, penggunaan strategi *Explainable AI* (XAI) dan manajemen data yang ketat menjadi pilar utama dalam menjaga akuntabilitas serta memitigasi risiko bias algoritmik dalam sistem penilaian. Implementasi teknologi ini memastikan bahwa hak-hak peserta didik, termasuk transparansi skor dan perlindungan privasi data, tetap menjadi prioritas utama dalam transformasi digital Pendidikan. Sebagai rekomendasi program lanjutan, disarankan pengembangan aplikasi penilaian berbasis *mobile* yang mengintegrasikan model-model efisien ini untuk cakupan level kompetensi yang lebih luas, serta eksplorasi penggunaan *federated learning* guna memperkuat keamanan data melalui pelatihan model yang terdesentralisasi pada perangkat masing-masing pengguna.

Daftar Pustaka

- Ansori, A., Damyati, F., & Dhestyani, S. A. (2025). Assessing AI Integration in Islamic Higher Education: A Mixed-Methods Fishbone Diagram Analysis. *IJID (International Journal on Informatics for Development)*, 13(2), 504–516.
- Ali, M., & Rahma, L. (2021). Kerangka evaluatif untuk AI dalam edukasi. *Jurnal Inovasi Pendidikan*, 9(3), 115–129.
- Errakha, K., Samih, A., Marzouk, A., & Krari, A. (2025). Recommender systems in e-learning: Trends, challenges, and future directions. *Journal of Theoretical and Applied Information Technology*, 103(7), 150–213.
- Habibi, R., & Ramadhan, R. I. (2025). Systematic literature review: Tren dan tantangan machine learning pada sistem rekomendasi kursus daring. *Journal of Applied Computer Science and Technology*, 6(2), 129–140. <https://doi.org/10.52158/ct1kr873>
- Lan, A. S., Studer, C., & Baraniuk, R. G. (2014). *Quantized matrix completion for personalized learning*. arXiv. doi.org
- Rokhmah, S., Permana, D., Elmi, F., & Ansori, A. (2025). AI implementation as support for successful green campus implementation. *Educational Process: International Journal*, 19, e2025569.
- Susilowati, Y., Sabaryati, J., Jusmiana, A., Isfahani, R., & Marsepa, E. (2026). Artificial intelligence dalam assessment pendidikan: A systematic literature review tentang model,

validitas, dan implikasi evaluatif. *Education, Social Sciences, and Linguistics: Conference Series*, 2(1), 211–219.

Tugu Malang ID. (2025, Desember 2). *Lima tren AI yang akan mengubah dunia pendidikan 2025-2030*. tugumalang.id

Vetian, R. A., Lelitasari, A., Fatin, J. S. B., & Astartia, D. D. (2025). Pengembangan sistem e-learning bahasa jepang berbasis adaptive learning dengan dukungan artificial intelligence. *Jurnal Riset Multidisiplin Edukasi*, 2(7), 593–603.

Wang, Z., Gu, Y., Lan, A. S., & Baraniuk, R. G. (2020). *VarFA: A variational factor analysis framework for efficient Bayesian learning analytics*. arXiv. doi.org