JASIMA: JURNAL KOMUNIKASI KORPORASI DAN MEDIA
Vol. VII, No. 2, Desember 2026 | E-ISSN: 2723-0562

Lebih dari Sekadar Keterampilan Teknis: Dampak Pelatihan Literasi AI terhadap Verifikasi Kritis dan Kesadaran Etis Pengguna di Indonesia

Beyond Technical Skill: The Impact of AI Literacy Training on User's Critical Verification and Ethical Awareness in Indonesia

Nuril Hidayah[1], Rutmalem Atania Tarigan[2], Loina Lalolo Krina Perangin-angin[3], Raundoh Tul Jannah[4], Priska Nur Safitri[5], Violita Siska Mutiara[6]
STAI Miftahul 'Ula, Nganjuk, Indonesia[1]
Mafindo, Jakarta, Indonesia[2]
Swiss German University, Tangerang, Indonesia[3]
Mafindo, Jakarta, Indonesia[4]
Mafindo, Jakarta, Indonesia[5]
STIKES Tri Mandiri, Bengkulu, Indonesia[6]

Riwayat Artikel : Submission 13/09/2026, **Revision** 12/10/2026, **Accepted** 13/10/2026

Abstrak

Adopsi kecerdasan artifisial generatif yang luas mengharuskan adanya transisi dari sekadar keterampilan teknis menuju literasi AI kritis. Studi ini menelaah pengaruh program pelatihan AI Ready terhadap verifikasi kritis dan kesadaran etis di kalangan pengguna di Indonesia. Data dianalisis menggunakan desain retrospektif *pretest* dan *posttest* berdasarkan Model Evaluasi Kirkpatrick, dari sebuah kohort nasional skala besar yang terdiri dari berbagai kelompok yang diukur pada satu titik waktu. Evaluasi tingkat 1 menunjukkan hasil positif terhadap kepuasan keseluruhan peserta terkait kejelasan instruktur dan relevansi materi. Analisis non-parametrik mengoperasionalkan berbagai elemen kompetensi kritis, khususnya terkait verifikasi informasi, identifikasi bias, perlindungan privasi, dan pencegahan plagiarisme, menunjukkan peningkatan yang nyata pada semua indikator. Secara khusus, perempuan mengungguli laki-laki dalam rata-rata peningkatan kompetensi kecerdasan buatan. Peserta juga menunjukkan niat yang kuat untuk menggunakan verifikasi silang secara berkala dan protokol privasi harian. Hasil studi ini menunjukkan bahwa dengan intervensi pendidikan yang terstruktur dengan baik, kesenjangan dalam kesadaran tentang risiko dan praktik dapat dikurangi. Ini berarti literasi AI kritis merupakan seperangkat keterampilan yang diperlukan untuk membekali komunitas dengan alat guna melindungi diri misinformasi mesin otomatis maupun dilema moral yang bersifat algoritmis.

Kata kunci: Etika AI[1], kecerdasan artifisial[2], literasi kritis[3], verifikasi informasi[5]

Abstract

The widespread adoption of generative AI required a transition from technical expertise to developing critical AI literacy. This research evaluated how the Artificial Intelligence Ready training program affected Indonesian users' critical verification skills and ethical awareness. Using a quantitative retrospective pretest-posttest design based on the Kirkpatrick Evaluation Model,

researchers collected data from a large national cohort representing different user groups. Results of the Level 1 evaluation indicated that most participants were satisfied with the trainer's ability to clearly communicate the content and its relevance to their work. Non-parametric analysis showed substantial gains in all areas of critical competency, such as verification of information, identifying bias, protecting user privacy and preventing plagiarism. It was found that females had a greater gain in overall AI competency than males. Also, participants reported that they would likely use private protocols and cross-verify information every day. The results indicate that structured education programs provide a way to connect users' awareness of risks associated with using AI with actual practice. Therefore, developing critical AI literacy provides a safeguard for developing communities to prevent automated misinformation and manage complex ethical decisions related to algorithms.
Keywords**:** *AI ethics[1]; artificial intelligence[2]; critical literacy[3]; information verification[4]*

PENDAHULUAN

Saat ini AI telah menjadi bagian kehidupan sehari-hari banyak kalangan masyarakat di berbagai belahan dunia. Integrasi AI telah menjadi suatu tren global. Namun, antusiasme yang ada sering kali tidak disertai dengan pemahaman kritis. Perangkat AI seperti ChatGPT dan Gemini memang membantu proses demokratisasi akses terhadap informasi, namun perangkat-perangkat ini juga memunculkan kompleksitas baru dalam ekosistem digital, khususnya berkaitan dengan integritas informasi. Pengguna perlu dapat memisahkan gandum dari sekam dan apa yang benar-benar menjadi kebenaran versus halusinasi. Ha ini membutuhkan lebih dari sekadar keterampilan teknis. Terutama ketika kesenjangan literasi digital belum terpecahkan. Di sini dibutuhkan benteng pertahanan yang lebih komprehensif terhadap misinformasi.

Penelitian tentang *generative* AI dan adopsi teknologi dalam skala besar di negara-negara berkembang telah menunjukkan bahwa penggunaan perangkat teknologi sering kali melampaui pemahaman mendalam tentang perangkat tersebut, tata kelolanya, atau dasar buktinya, dengan selisih beberapa kali lipat. Dalam bidang pendidikan, lebih banyak sekolah mulai mengadopsi alat akademik berbasis *cloud* AI dan *generative* AI. Dalam hal kebijakan dan tata kelola, banyak negara berkembang telah bereksperimen dengan AI generatif di bidang bisnis, layanan publik, dan proyek pembangunan. Akan tetapi strategi AI nasional, regulasi sektoral, dan kerangka etika masih berada di tahap awal (Folorunso et al., 2024; Khan et al., 2024; Mannuru et al., 2023).

Ada beberapa faktor struktural yang dapat menjelaskan mengapa hal ini terjadi. Pertama, pengguna umumnya berinteraksi dengan AI melalui platform digital sehari-hari tanpa memahami mekanisme yang bekerja di baliknya. Di satu sisi, penyediaan bukti mengenai kerentanan tersebut yang akan diperparah oleh kesenjangan infrastruktur serta literasi digital yang sudah ada (Folorunso et al., 2024; Kumar, 2025a, 2025b). Kedua, terjadi difusi teknologi secara *top-down*.

Kecepatan teknologi ini hadir dan berkembang jauh lebih cepat daripada yang dapat dikelola melalui upaya penguatan kapasitas lokal atau percakapan publik; sebaliknya, teknologi-teknologi tersebut dimungkinkan oleh platform teknologi global dan organisasi internasional (Khan et al., 2024; Mannuru et al., 2023). Akibatnya, logika belajar *learning-by-doing* menjadi dominan. Individu dan institusi cenderung mengadopsi teknologi tersebut terlebih dahulu dan baru beradaptasi kemudian. Mereka baru menyadari kompleksitas etisnya ketika muncul kegagalan tak terduga atau terjadi risiko misinformasi (Greenhalgh et al., 2017).

Bahaya yang jelas di sini berupa sesuatu yang disebut "halusinasi AI", yang secara tata bahasa benar dan terdengar sangat meyakinkan, tetapi tidak faktual (Shao, 2025). Di berbagai bidang, pemalsuan tersebut berlangsung secara sistematis, dan menjadi ancaman terhadap validitas informasi (Melnyk, 2025). Dengan kata lain, ini bukan sekadar persoalan teknis yang sederhana. Risikonya terletak pada seberapa besar hasil AI itu tampak meyakinkan (tingkat plausibilitas). Karena LLM dilatih untuk menghasilkan prediksi teks kata demi kata, mereka sering memberikan rekayasa lengkap tentang hal-hal yang bisa tampak sangat nyata. Hal ini membuat pengguna sangat sulit mendeteksi kesalahannya, dan kadang bahkan mempercayai informasi yang dihasilkan oleh sistem AI yang tidak benar ketimbang sumber yang otoritatif. Saat bertumpu pada jalan pintas kognitif, kerentanannya meningkat secara signifikan. Halusinasi AI ini dapat dengan mudah membangun keyakinan jika upaya verifikasi aktif tidak dilakukan, sehingga mencemari ekosistem informasi digital (Hwang & Jeong, 2025; Joshi, 2025).

Terlepas dari risiko-risiko di atas, diperlukan upaya membekali pengguna supaya mampu menavigasi lanskap yang masih belum seimbang ini. Tinjauan terhadap kerangka kerja literasi AI saat ini mengungkap adanya kecenderungan pada kecakapan operasional, yang berfokus pada cara berinteraksi dan cara menerapkan teknologi tersebut, dan belum mencakup cara mengevaluasinya secara kritis (Almatrafi et al., 2024; Casal-Otero et al., 2023). Sementara model kerangka kerja saat ini menyoroti pentingnya konstruksi etis dan sosio-teknis (misalnya, deteksi bias, kesadaran risiko, dan evaluasi kontekstual), sebagai dimensi intrinsik dalam sebuah kurikulum, perlakuannya sering kali lebih bersifat aspiratif daripada mendasar (Allen & Kendeou, 2023; Ng et al., 2023). Akibatnya, aspek fungsional masih lebih ditekankan daripada refleksi kritis dalam beragam intervensi pendidikan (Atias & Mawasi, 2025; Biagini, 2025). Kesenjangan ini mengungkap dimensi lain yang bertanggung jawab terhadap kesiapan pengguna. Mereka dilatih untuk membuat konten dengan cepat, tetapi tidak membekali mereka dengan keterampilan yang mereka

perlukan untuk bersikap kritis terhadap implikasi etis atau faktual dari apa pun yang dihasilkan (Pinski & Benlian, 2024).

Lebih parahnya lagi, ketidakseimbangan kurikulum ini juga dibarengi dengan evaluasi program literasi digital dan AI sering kali sempit dan informal. Banyak inisiatif yang mengukur keberhasilan program hanya berdasarkan jumlah peserta atau umpan balik simbolis, serta tidak memiliki penilaian pra- dan pasca-intervensi yang diperlukan untuk membuktikan adanya peningkatan (Choudhary & Bansal, 2022; Julien et al., 2022). Untuk menanggulanginya, banyak akademisi sudah memulai menyuarakan pentingnya kerangka evaluasi secara menyeluruh, yakni beralih pada pengukuran perubahan perilaku sebenarnya (Pan et al., 2025; Suri et al., 2025).

Model evaluasi kerja *multi-tiered* ala Kirkpatrick menawarkan kerangka yang menyeluruh. Model ini juga telah digunakan secara luas dalam pelatihan *EdTech*. Biasanya, para peneliti akan mengoperasionalkan model tersebut dengan terlebih dahulu mengidentifikasi sasaran institusional. Memajukan capaian peserta didik, atau penggunaan teknologi yang lebih luas mungkin; kemudian menguraikannya kembali melalui keempat levelnya (Achuthan et al., 2023; Carnell et al., 2022). Level 1 (Reaksi) adalah tingkat terendah yang mencatat kepuasan pengguna dan kegunaan yang dirasakan. Tingkat ini dapat memberikan umpan balik yang diperlukan untuk membuat konten menjadi lebih baik. Level 2 (Pembelajaran) menilai pembelajaran nyata, bukan persepsi, menggunakan penilaian pra dan pasca yang berfokus pada keterampilan teknis atau pengetahuan pedagogis. Level 3 (Perilaku) menggeser fokus dari memantau kompetensi ke praktik dan melihat apakah peserta benar-benar menggunakan apa yang telah mereka pelajari dalam metode pengajaran mereka sehari-hari. Level 4 (Hasil) adalah ketika perubahan perilaku ini berkorelasi dengan keberhasilan organisasi atau hasil pendidikan yang lebih baik. Dengan menyusun evaluasi melewati empat dimensi yang berbeda ini, model Kirkpatrick berusaha memastikan bahwa inisiatif literasi digital diukur bukan hanya dari kepuasan peserta, melainkan dari penerapannya di dunia nyata serta dampaknya secara sistemik (Praslova, 2010; Rimal, 2025).

Penelitian ini dimulai karena kurangnya literasi AI di Indonesia, sehingga sangat penting untuk menerapkan intervensi pendidikan. Oleh karena itu, dikembangkan sebuah program bernama pelatihan AI Ready. Program ini bertujuan menjangkau total 10.000 orang, termasuk siswa pendidikan dasar dan menengah (K-12), mahasiswa, pendidik, serta orang tua mereka. Dengan memahami bahwa literasi AI melampaui aspek teknis, dalam artian melibatkan etika, keamanan, dan kemampuan berinteraksi, rangkaian pelajaran ini disusun berdasarkan lima kompetensi utama, yaitu: penulisan prompt secara iteratif, mengidentifikasi plagiarisme dan mencantumkan sumber, memverifikasi

informasi/mitigasi risiko bias terhadap privasi, serta menggunakan AI sebagai bantuan untuk belajar atau rekreasi.

Dengan menyesuaikan kerangka teoritisnya pada model evaluasi Kirkpatrick, penelitian ini mengkaji tiga tingkat pertama (reaksi, pembelajaran, dan perilaku) yang berkaitan dengan efek jangka pendek langsung dari program *AI Ready*. Lebih spesifik, penelitian ini menilai seberapa baik peserta meningkatkan keterampilan-keterampilan kritis sebagai hasil dari keterlibatan dengan intervensi (praktik verifikasi; pencegahan plagiarisme; kesadaran privasi). Penelitian ini juga mengeksplorasi bagaimana peserta mentransfer keterampilan etis dan kritis baru tersebut ke konteks akademik, profesional, dan personal. Makalah ini berupaya memetakan perubahan perilaku yang benar-benar terjadi, sehingga memberikan bukti empiris yang bermanfaat bagi penyelenggara program dan turut berkontribusi pada bidang kajian ilmiah yang masih berada pada tahap awal terkait evaluasi literasi AI secara menyeluruh, khususnya dalam konteks pendidikan di Indonesia.

Literasi AI Kritis (CAIL) dan Etika AI memperluas paradigma pendidikan dari sekadar penguasaan teknis operasional menuju kapasitas sosio-teknis untuk mengevaluasi bias sistemik, ketidakadilan algoritmik, risiko privasi, serta dilema akuntabilitas dan dampak sosial AI secara mendalam (Al-Kfairy et al., 2024; Roe et al., 2025; Sperling et al., 2024; Yim, 2024). Komponen inti CAIL dan etika AI ini diintegrasikan ke dalam pengalaman pengguna sebagai titik alih. Pengguna harus beralih dari konsumen pasif menjadi penilai reflektif yang didukung oleh praktik verifikasi yang ketat, *lateral reading* (yakni pemeriksaan silang) dan pengawasan HITL, garis pertahanan utama terhadap misinformasi yang dibuat-buat (Hatem et al., 2023; Joshi, 2025; Melnyk, 2025). Pengoperasionalan kerangka teoretis ini ke dalam praktik dunia nyata memerlukan intervensi terstruktur seperti program pelatihan AI Ready, yang bertujuan untuk menanamkan kompetensi kritis dan kesadaran etis ini secara langsung, demi menanggapi kesenjangan yang dialami sekaligus berupaya memberdayakan ketahanan komunitas terhadap krisis integritas informasi di era kompleksitas digital.

METODE

Penelitian ini menggunakan desain *retrospective pretest-posttest* kuantitatif. Berbasis tiga level, yaitu: Model Evaluasi Kirkpatrick yang mencakup Reaksi (Level 1), Pembelajaran (Level 2), dan Perilaku (Level 3). Pendekatan ini digunakan dalam mengukur efektivitas pelatihan dan dampak terhadap kompetensi peserta. Beralih dari format *pretest-posttest* tradisional ke evaluasi retrospektif, penelitian ini berupaya memitigasi bias pergeseran respons (*response-shift bias*), sebuah fenomena umum dalam evaluasi literasi digital yang

dilaporkan secara mandiri (*self-reported*) di mana peserta pada awalnya melebih-lebihkan kompetensi mereka sebelum memahami kompleksitas materi. Desain ini dianggap dapat mengisolasi dan mengkuantifikasi pergeseran terukur dalam literasi AI kritis.

Program literasi digital AI Ready menargetkan kohort nasional berskala besar. Sampel evaluasi yang dipilih menggunakan *purposive sampling* dan *stratified random sampling* untuk memastikan keterwakilan di seluruh segmen demografi. Kriteria inklusi yang ketat mengharuskan peserta telah terdaftar yang telah menyelesaikan penuh modul pelatihan 12 jam dan telah diberi izin untuk menggunakan data secara resmi, Peserta yang tidak menyelesaikan semua modul atau modul evaluasi akhir. Berdasarkan data pascapenerapan akhir, kumpulan data evaluasi ini terdiri dari 17.207 individu, sebagian besar adalah kaum muda dan pendidik atau orang tua.

Data dikumpulkan secara *online* melalui kuesioner terstruktur dan terintegrasi berdasarkan kerangka kerja Kirkpatrick. Level 1 (reaksi) dinilai dengan skala Likert 4 poin (1 = Sangat Tidak Setuju hingga 4 = Sangat Setuju) untuk menilai kecukupan kesesuaian materi instruksional, durasi, format pelatihan, dan kejelasan fasilitator. Kompetensi perilaku yang dirasakan (level 2: pembelajaran (dijelaskan lebih rinci di bawah) dinilai baik secara retrospektif maupun dalam evaluasi setelah pelatihan dengan menggunakan skala yang terdiri dari 10 butir. Partisipan menilai keterampilan yang mereka rasakan (sebelum dan sesudah program) menggunakan skala yang mencakup indikator inti, termasuk memeriksa informasi, mengidentifikasi informasi yang keliru dan bias, serta mengambil langkah untuk mencegah plagiarisme atau risiko dengan privasi. Level 3 (Perilaku) diukur menggunakan skala yang menilai niat untuk melakukan perilaku etis tertentu dalam aktivitas bulan depan, termasuk verifikasi silang, pembuatan prompt yang efektif, sitasi, atau perlindungan privasi.

Pemrosesan *dataset* dilakukan secara sistematis untuk melakukan analisis deskriptif dan inferensial. Untuk kepuasan (Tingkat 1) dan penilaian proyeksi tindakan perilaku (Tingkat 3), digunakan statistik deskriptif, termasuk persentase dan frekuensi. Statistik inferensial diterapkan pada skor komposit (perolehan kompetensi AI kritis, Tingkat 2). Karena skor kompetensi berasal dari data ordinal dan skor perubahan tidak terdistribusi secara normal, digunakan Uji Peringkat Bertanda Wilcoxon untuk menentukan apakah peningkatan pengetahuan literasi peserta antara penilaian pra-pelatihan retrospektif dan penilaian pasca-pelatihan mencapai signifikansi statistik. Selanjutnya, uji Mann-Whitney U dilakukan untuk menganalisis peningkatan kompetensi berdasarkan perbedaan demografis.

HASIL DAN PEMBAHASAN

Berdasarkan tiga level pertama Model Evaluasi Kirkpatrick, pembahasan ini dimulai dengan tinjauan mengenai profil demografis peserta yang berhasil menyelesaikan seluruh modul pelatihan. Selanjutnya, hasil evaluasi memerinci tingkat kepuasan peserta terhadap penyampaian materi dalam program, serta mengukur peningkatan yang signifikan dalam pemerolehan kompetensi kritis literasi AI.

1. Profil Peserta dan Tingkat Penyelesaian

Survei pasca-implementasi, yang berfungsi sebagai set data akhir untuk evaluasi, diselesaikan oleh 17.207 peserta. Peserta perempuan merupakan mayoritas responden dengan persentase 63,28% (n=10.889), diikuti oleh peserta laki-laki sebesar 36,36% (n=6.257). Sebagian kecil responden memilih untuk tidak menyebutkan gender mereka (0,28%, n=49) atau mengidentifikasi diri sebagai gender lain (0,07%, n=12). Terkait hasil inklusivitas menunjukkan bahwa mayoritas responden bukan penyandang disabilitas (94,58%), sementara 2,28% sebagai penyandang disabilitas, dan 3,14% tidak mengungkapkan statusnya.

a. Level 1: Reaksi dan Kepuasan terhadap Program

Evaluasi di tingkat Reaksi adalah menilai tanggapan langsung dari peserta terhadap desain instruksional dengan skala pengukuran Likert 4 poin. Hasil tertinggi yaitu kejelasan pelatih (*trainer clarity*) (M = 3,39, SD = 0,69), dengan 91,16% peserta setuju/ sangat setuju mengenai pelatih menjelaskan konsep AI dengan jelas serta mudah dipahami. Relevansi pelatihan menerima skor rata-rata tertinggi kedua (M = 3,37, SD = 0,70), dengan 90,70% peserta menyatakan persetujuan positif. Format pelatihan juga dievaluasi secara positif (M = 3,33, SD = 0,69), dengan 90,09% sepakat bahwa formatnya ideal. Skor terendah secara komparatif terpantau pada durasi pelatihan (M = 3,27, SD = 0,74), meskipun 87,26% peserta tetap setuju bahwa durasi tersebut cukup, yang mengindikasikan penerimaan positif secara keseluruhan.

b. Level 2: Perolehan Literasi AI Kritis

Penilaian kompetensi retrospektif pre-post sebanyak 10 butir dilakukan untuk menilai perolehan kompetensi kritis dan etis. Rata-rata skor kompetensi keseluruhan meningkat dari 4,55 (SD = 1,54) sebelum program menjadi skor pascaprogram sebesar 4,96 (SD = 1,52), yang mencerminkan peningkatan absolut pada skor sebesar MD:0,41 atau setara dengan kurang lebih +9·10·%). Sebelum dan sesudah program, kompetensi AI yang dirasakan berbeda secara signifikan (Uji Peringkat Bertanda *Wilcoxon*; V = 16,066354, p <0,001). Dan menunjukkan peningkatan yang cukup besar untuk dimensi-dimensi kritis tertentu. Rata-rata

peningkatan kemampuan untuk menghasilkan prompt AI yang fungsional adalah 10,35% (*Mean Difference* = +0,46), dan ketika mereka mampu mengidentifikasi plagiarisme dalam konten buatan AI meningkat sebesar 9,72% (Selisih Rata-rata = +0,44). Metrik etis dan kritis yang sangat penting juga mengalami peningkatan signifikan: kemampuan untuk mengutip informasi buatan AI dengan tepat membaik sebesar 9,29% (Selisih Rata-rata = +0,42), mengidentifikasi risiko privasi meningkat sebesar 9,00% (Selisih Rata-rata = +0,41), dan kapasitas untuk mengidentifikasi misinformasi serta bias pada luaran AI meningkat sebesar 8,61% (Selisih Rata-rata = +0,39).

Perolehan Kompetensi Berbasis Gender*:* Uji Mann-Whitney U (uji jumlah peringkat Wilcoxon) dilakukan untuk menilai secara komparatif perubahan kompetensi antar kelompok berdasarkan gender. Hasil menunjukkan bahwa dibandingkan dengan pria (N = 6.257), partisipan perempuan memperlihatkan peningkatan yang secara signifikan lebih besar pada skor kompetensi AI keseluruhan untuk semua III item yang digabungkan. Skor kenaikan rata-rata untuk perempuan (M = 0,456, SD = 1,13) sangat dekat dengan laki-laki (M = −0,331, SD = 1,08). Selain itu, seluruh kompetensi AI meningkat lebih besar pada partisipan perempuan dibandingkan laki-laki; area Literasi AI Kritis menunjukkan peningkatan terbesar (0,131 poin) dan Penggunaan AI yang Aman & Bertanggung Jawab berada pada urutan kedua (+0,128).

c. Level 3: Niat dan Perilaku

Evaluasi ini mengukur niat keterlibatan dan penerapan keterampilan AI peserta di masa depan dalam kurun waktu satu bulan berikutnya. Secara keseluruhan peserta memperoleh skor rata-rata Niat Perilaku AI (*AI Behavioral Intention*) 3,89 dengan SD = 0,89 pada skala pengukuran 5 poin, hal ini mengindikasikan kesiapan positif untuk terlibat serta merekomendasikan AI dalam pembelajaran. Mengenai penerapan spesifik dari keterampilan kritis, peserta mengekspresikan niat yang kuat untuk mempraktikkan tindakan etis. Sekitar 84% peserta mengindikasikan niat yang tinggi atau sangat tinggi untuk menerapkan pengetahuan privasi saat menggunakan AI (M = 3,23). Niat untuk memverifikasi informasi buatan AI mencatatkan rata-rata 3,11, sedangkan komitmen untuk mengutip sumber buatan AI dengan tepat mencatatkan rata-rata 3,06. Capaian-capaian ini mengindikasikan bahwa peserta secara aktif berniat untuk menerjemahkan pengetahuan teoretis mereka ke dalam kebiasaan digital yang praktis dan bertanggung jawab.

2. Bergerak Melampaui Keterampilan Teknis

Hasil dari program AI Ready menunjukkan alasan yang jelas bahwa edukasi AI harus melampaui pengembangan keterampilan fungsional dalam

kerangka pendidikan AI menuju *Critical AI Literacy* (CAIL). Integrasi tiga tingkat ini yang dirancang dengan cermat menggunakan metrik Level 1, Level 2, dan Level 3 menunjukkan bahwa hasil yang bermakna dalam membangun ketahanan digital dapat dicapai berdasarkan Model Evaluasi Kirkpatrick.

Respons awal pada tingkat Reaksi mengindikasikan kepuasan yang tinggi, khususnya mengenai kejelasan pelatih (M = 3,39). Penelitian terdahulu menekankan bahwa dalam lingkungan pembelajaran daring dan *blended learning*, kehadiran fasilitator yang responsif merupakan penentu utama keterlibatan pengguna dan hasil belajar (Wang et al., 2023). Penyampaian yang efektif ini secara langsung memfasilitasi peningkatan kognitif substansial yang diamati pada level Pembelajaran.

Metrik utama yang berkaitan dengan deteksi bias, verifikasi informasi yang diberikan, serta perlindungan privasi telah menunjukkan peningkatan yang sangat signifikan ($p < 0,001$), sehingga memberikan harapan untuk menjembatani kesenjangan ini dalam lingkungan *online* yang aman saat ini. Kompetensi fungsional memiliki peningkatan tertinggi, yaitu kemampuan membuat *prompt* meningkat sebesar 10,35% yang sangat besar (Selisih Rata-rata = +0,46). Ini mengonfirmasi studi-studi sebelumnya bahwa keterampilan fungsional dan operasional lebih mudah diserap ke dalam hasil pelatihan dan sering kali menjadi pusat perhatian (Almatrafi et al., 2024; Casal-Otero et al., 2023).

Namun demikian, titik penting dari temuan ini terletak pada peningkatan keterampilan evaluatif yang terukur. Peserta menunjukkan peningkatan yang signifikan dalam mengidentifikasi risiko privasi (9,00%) dan mengidentifikasi misinformasi serta bias dalam luaran AI (8,61%). Dalam konteks adopsi AI di negara-negara berkembang, di mana integrasi teknologi sering kali didorong secara *top-down* dan melampaui pemahaman kritis pengguna (Khan et al., 2024; Mannuru et al., 2023), hasil ini merupakan ukuran keberhasilan yang krusial. Kemampuan program dalam merumuskan kompetensi pengguna secara umum ($V = 16.066.354$, $p < 0,001$) menunjukkan bahwa intervensi pendidikan yang singkat dan terstruktur dapat menanamkan sikap skeptis yang cukup untuk menangkal pemalsuan algoritmik. Selain itu, meningkatnya deteksi plagiarisme yang dihasilkan AI sebesar 9,72% dan meningkatnya kepatuhan dalam mengutip sumber secara etis sebesar 9,29% menunjukkan bahwa pengguna memang sedang beralih dari sekadar konsumen keluaran mesin yang pasif menjadi para penilai.

Salah satu persyaratan utama ketahanan untuk informasi digital adalah kebutuhan untuk melakukan verifikasi silang dan juga mengenali di mana fabrikasi algoritmik muncul (Joshi, 2025; Melnyk, 2025). Karena LLM pada dasarnya dirancang untuk kelancaran menghasilkan teks yang canggih, bukan

untuk akurasi faktual, hal ini berarti bahwa peningkatan sebesar 9,72% dalam identifikasi plagiarisme yang dihasilkan AI dan peningkatan sebesar 9,29% dalam sitasi etis merupakan indikasi pengguna aktif yang beralih dari konsumen pasif atas keluaran mesin menjadi penilai kritis. Terakhir, peningkatan substansial dalam kompetensi etis dan kritis, yang selaras erat dengan metrik peningkatan teknis, menegaskan bahwa kurikulum yang holistik dapat berhasil menanamkan pengaman *human-in-the-loop* yang mendasar di tingkat pengguna individual.

Dalam negara berkembang seperti Indonesia, penyebaran lebih lanjut alat AI generatif ini dijalankan dengan logika belajar sambil praktik, di mana adopsi massal dari tingkat atas jauh mendahului pemahaman kritis dan pengembangan kapasitas lokal oleh pengguna (Greenhalgh et al., 2017; Khan et al., 2024; Mannuru et al., 2023). Pengadopsian model generatif tanpa verifikasi dapat menimbulkan bahaya informasi yang serius, seperti halusinasi dan misinformasi (Joshi, 2025; Melnyk, 2025). Mode operasinya mengarah pada optimasi sistematis untuk kelancaran linguistik, bukan akurasi faktual. Keberhasilan program AI Ready ini dalam meningkatkan keterampilan verifikasi pengguna yang berkembang menunjukkan bahwa bahkan pendidikan jangka pendek yang paling intens sekalipun dapat memberikan skeptisisme yang memadai untuk meminimalkan fabrikasi algoritmik seperti jenis yang kami temukan.

Terakhir dan yang paling penting, evaluasi Perilaku menunjukkan bahwa kurikulum dapat membantu menjembatani kesenjangan dalam pencapaian kinerja pada tingkat perilaku. Studi kualitatif baru mengindikasikan bahwa praktisi dalam kelompok pengguna umum hanya memiliki kesadaran yang dangkal tentang risiko AI tanpa mengembangkan kebiasaan pemeriksaan fakta atau perlindungan yang tepat saat menghadapi situasi (Grotlüschen et al., 2024). Namun, keberhasilan ini menandai pergeseran yang bersifat struktural menuju praktik yang dilakukan dengan kesadaran. Secara khusus, sekitar 84% peserta menyatakan tingkat kesediaan yang tinggi/sangat tinggi untuk menerapkan pengetahuan tentang privasi (M=3,23) saat berinteraksi dengan sistem AI. Hal ini juga menunjukkan niat yang tinggi untuk menggunakan teknik *prompting* yang baik (79% berniat tinggi/sangat tinggi), serta untuk memvalidasi silang informasi (M=3,11), yang mengindikasikan bahwa peserta berencana secara aktif mengintegrasikan perlindungan ini ke dalam alur kerja mereka.

Niat perilaku ini menggarisbawahi urgensi literasi etis dalam menavigasi teknologi *black-box*. Model AI generatif yang dilatih dengan kumpulan data masif secara inheren membawa risiko tinggi berupa kebocoran data sensitif dan pelanggaran kekayaan intelektual (Al-Kfairy et al., 2024; Surbakti, 2025). Niat perilaku yang kuat untuk mengimplementasikan protokol

privasi dan mengutip sumber secara etis (M = 3,06) merepresentasikan pergeseran kognitif yang vital. Peserta mulai menyadari posisi rentan mereka di dalam ekosistem ekstraksi data global dan kini telah dibekali kapasitas untuk mengambil langkah perlindungan proaktif.

Di luar persoalan privasi, kemampuan untuk mengidentifikasi bias algoritmik membawa implikasi sosiopolitik yang signifikan, khususnya bagi pengguna di negara berkembang. Kajian akademik terbaru memperingatkan bahwa *large language models* (LLM) sering kali mereproduksi bias representasional dan beroperasi melalui kacamata Barat atau penutur bahasa Inggris. Hal ini dapat secara sistematis memarjinalkan perspektif lokal dan menormalisasi stereotip budaya serta gender (Gautam et al., 2024; Wan et al., 2024). Sebagai tambahan, analisis sensitif gender dari intervensi ini mengungkapkan bahwa peserta perempuan menunjukkan peningkatan rata-rata yang lebih tinggi pada kompetensi AI secara keseluruhan (0,456 poin) dibandingkan dengan peserta laki-laki (0,331 poin), dengan perbedaan gender terbesar teramati secara spesifik pada dimensi Literasi AI Kritis (+0,131 poin). Temuan ini sangat penting, karena mengonfirmasi bahwa pelatihan CAIL terstruktur mampu menumbuhkan ketahanan sosiokultural dan secara aktif memberdayakan demografi yang paling rentan terhadap marjinalisasi algoritmik. Alih-alih mengonsumsi informasi secara pasif, peserta, khususnya perempuan, kini lebih siap untuk secara aktif mempertanyakan asumsi dan struktur kekuasaan yang tertanam di dalam respons algoritmik.

Keberhasilan intervensi program AI Ready memperkuat pandangan bahwa arah pendidikan AI di masa depan harus berpusat pada penguatan metakognisi dan regulasi diri. Fenomena bias otomatisasi, di mana pengguna terlalu memercayai luaran mesin, telah terbukti memicu degradasi pemikiran analitis independen (Helal et al., 2025; Hwang & Jeong, 2025). Temuan penelitian evaluatif ini mengonfirmasi bahwa intervensi pendidikan yang secara eksplisit membahas risiko halusinasi AI mampu menumbuhkan skeptisisme yang sehat (Hwang & Jeong, 2025). Tantangan sekaligus peluang utama untuk pengembangan kurikulum di masa depan terletak pada upaya mempertahankan kewaspadaan ini. Program literasi tingkat lanjut seyogianya tidak hanya mengajarkan cara mengoperasikan AI secara efektif, tetapi juga menumbuhkan kebijaksanaan untuk memutuskan secara sadar kapan harus tidak menggunakan AI demi menjaga otonomi intelektual dan integritas moral peserta.

Terakhir, keselarasan dari ketiga level evaluasi ini mengonfirmasi bahwa literasi AI kritis bertindak sebagai pilar fundamental bagi ketahanan digital. Dengan membekali kohort yang didominasi oleh kalangan muda dengan kemampuan spesifik untuk memverifikasi silang klaim dan menavigasi opasitas algoritmik *black-box*, intervensi ini membangun sistem pengaman *human-in-the-*

loop yang terlokalisasi (Chu-Ke & Dong, 2024; Guo et al., 2025). Hal ini selaras dengan urgensi sosioteknis yang lebih luas, yaitu memastikan bahwa seiring dengan meluasnya teknologi AI secara global, kapasitas untuk bernalar etis dan menjaga integritas informasi juga turut meluas secara serentak di tingkat pengguna individu.

Peningkatan signifikan yang diamati pada kompetensi AI kritis peserta menghadirkan implikasi mendalam bagi para pembuat kebijakan literasi digital, organisasi non-pemerintah (LSM), dan institusi pendidikan. Banyak data yang jelas menunjukkan bahwa orang-orang sangat terbuka untuk belajar secara etis selama mengintegrasikannya secara struktural dengan pelatihan keterampilan teknologi. Oleh karena itu, kerangka literasi digital nasional yang memprioritaskan akses internet dasar dan keterampilan fungsional harus segera beradaptasi dengan CAIL sebagai komponen inti pendidikan di sekolah.

Institusi pendidikan menyoroti perlunya perubahan pedagogis. Alih-alih menerapkan larangan menyeluruh terhadap kecerdasan buatan generatif untuk menghindari kekhawatiran terkait *ai-giarism* atau misinformasi, institusi perlu mempersiapkan pendidik untuk berperan sebagai fasilitator yang dapat membimbing siswa dalam menggunakan alat-alat ini secara kritis dan kreatif. Tingginya niat peserta untuk memverifikasi informasi yang dihasilkan oleh AI (M = 3,11) dan ketepatan dalam mencantumkan rujukan sumber (M = 3,06) menunjukkan bahwa risiko kecurangan dapat dikurangi secara signifikan melalui desain yang terstruktur. Selain itu, dalam konteks kerja mereka untuk melawan disinformasi digital bagi masyarakat sipil dan LSM, penelitian ini memberikan peta jalan yang tervalidasi dan dapat dikuantifikasi. Dengan mengintegrasikan verifikasi dan pendeteksian bias ke dalam kurikulum pelatihan publik, para pemangku kepentingan dapat mengambil langkah-langkah untuk secara proaktif menumbuhkan generasi yang tangguh secara digital, cukup mampu menghadapi hoaks atau propaganda berbasis AI yang dihasilkan secara otomatis dengan upaya mereka sendiri.

KESIMPULAN

Adopsi skala besar kecerdasan buatan mendorong perlunya peninjauan menyeluruh terhadap standar literasi digital. Studi Kasus Program AI Ready menunjukkan bahwa seseorang dapat secara sistematis mengembangkan kompetensi kritis dan etis melalui intervensi pendidikan yang terstruktur. Peningkatan statistik yang paling penting ($p < 0,001$) pada peserta dalam mengonfirmasi informasi, mengenali bias algoritmik, membela diri dari plagiarisme, serta keterampilan perlindungan data mengarah pada kesimpulan bahwa literasi AI teknis tidak dapat diaplikasikan sendirian. Selain itu, hasil ini menegaskan bahwa pelatihan CAIL dapat bekerja secara langsung untuk

memberdayakan populasi yang berisiko termarginalkan oleh algoritma. Hal ini selanjutnya dibuktikan melalui peningkatan kompetensi kritis yang lebih besar pada responden perempuan. Studi ini menerapkan Model Evaluasi Kirkpatrick untuk menunjukkan bagaimana pelatihan holistik, yang melampaui kepuasan pengguna terhadap metode evaluasi saat ini, juga memengaruhi peningkatan kognitif dan niat perubahan perilaku yang kuat. Literasi AI Kritis berfungsi sebagai mekanisme fundamental *human-in-the-loop*. Mekanisme yang membantu navigasi melalui era otomasi, bukan dengan mengubah individu menjadi konsumen, melainkan dengan ketahanan melalui etika dan evaluasi yang cermat. Meskipun temuan-temuan ini memberikan wawasan empiris yang kaya, penelitian ini menyoroti beberapa keterbatasan metodologis. Kendati desain *retrospective pretest-posttest* secara efektif memitigasi bias pergeseran respons sehingga memberikan *baseline* yang lebih akurat untuk kompetensi evaluatif yang dilaporkan secara mandiri, studi ini masih tidak memiliki kelompok kontrol, dan evaluasi perilaku (Kirkpatrick Level 3) utamanya mengandalkan niat yang dilaporkan secara mandiri. Penelitian masa depan hendaknya menggunakan uji coba teracak atau kuasi-eksperimen untuk memberikan inferensi kausal yang lebih kuat mengenai efek pelatihan. Selain itu, penelitian masa depan hendaknya menggunakan rancangan longitudinal dan penilaian terhadap perilaku digital dalam konteks yang autentik untuk menentukan apakah niat etis ini beralih menjadi pola perilaku *online* yang berkelanjutan di berbagai pengaturan tempat kerja atau pembelajaran.

UCAPAN TERIMA KASIH

Para penulis menyampaikan rasa terima kasih yang sebesar-besarnya kepada Mafindo (Masyarakat Antifitnah Indonesia) atas kolaborasi strategis, penyediaan akses data, serta dedikasi tanpa lelah dalam mengimplementasikan program literasi AI Ready di seluruh lapisan masyarakat Indonesia. Penghargaan yang mendalam juga ditujukan kepada ASEAN Foundation atas dukungan pendanaan serta komitmen berkelanjutan mereka dalam memperkuat ketahanan digital dan literasi teknologi di seluruh kawasan Asia Tenggara. Keberhasilan intervensi pendidikan ini, sejalan dengan evaluasi penelitian yang komprehensif ini, tidak akan mungkin terwujud tanpa kontribusi luar biasa dari para pelatih utama (*master trainers*), fasilitator lokal, tim manajemen program, serta ribuan peserta yang telah mendedikasikan diri mereka sepenuhnya di sepanjang proses pelatihan yang intensif.

REFERENSI

Abro, A., Rafiq, K., & Raza, M. (2025). The Cognitive Impact of Artificial Intelligence on the Spread of Fake News through Social Media: Public Perceptions and Democratic Resilience. *Journal of Asian Development Studies*. https://doi.org/10.62345/jads.2025.14.2.62

Achuthan, K., Kolil, V. K., & Jyothy, S. (2023). Effectiveness of Virtual Laboratory Teacher Training Workshops: A Kirkpatrick Model Analysis. *Int. J. Emerg. Technol. Learn.*, *18*, 225–239. https://doi.org/10.3991/ijet.v18i15.40037

Al-Kfairy, M., Mustafa, D., Kshetri, N., Insiew, M., & Alfandi, O. (2024). Ethical Challenges and Solutions of Generative AI: An Interdisciplinary Perspective. *Informatics*, *11*. https://doi.org/10.3390/informatics11030058

Allen, L. K., & Kendeou, P. (2023). ED-AI Lit: An Interdisciplinary Framework for AI Literacy in Education. *Policy Insights from the Behavioral and Brain Sciences*, *11*, 3–10. https://doi.org/10.1177/23727322231220339

Almatrafi, O., Johri, A., & Lee, H. (2024). A Systematic Review of AI Literacy Conceptualization, Constructs, and Implementation and Assessment Efforts (2019-2023). *Computers and Education Open*. https://doi.org/10.1016/j.caeo.2024.100173

Arkan, Z., & Bal, M. (2025). The Relationship between School-Age Students' Literacy Skills and Digital Well-Being: A Systematic Review. *BMC Psychology*, *13*. https://doi.org/10.1186/s40359-025-03573-4

Atias, O., & Mawasi, A. (2025). Conceptualizing AI Literacies for Children and Youth: A Systematic Review On The Design Of AI Literacy Educational Programs. *Computers and Education: Artificial Intelligence*. https://doi.org/10.1016/j.caeai.2025.100491

Bećirović, S., Polz, E., & Tinkel, I. (2025). Exploring students' AI Literacy and Its Effects on Their AI Output Quality, Self-Efficacy, and Academic Performance. *Smart Learning Environments*, *12*. https://doi.org/10.1186/s40561-025-00384-3

Biagini, G. (2025). Towards an AI-Literate Future: A Systematic Literature Review Exploring Education, Ethics, and Applications. *International Journal of Artificial Intelligence in Education*, *35*, 2616–2666. https://doi.org/10.1007/s40593-025-00466-w

Buchan, M., Bhawra, J., & Katapally, T. (2024). Navigating The Digital World: Development of An Evidence-Based Digital Literacy Program and Assessment Tool for Youth. *Smart Learning Environments*, *11*. https://doi.org/10.1186/s40561-024-00293-x

Carnell, S., Siqueira, A., Miles, A., & Lok, B. (2022). *Informing and Evaluating Educational Applications With the Kirkpatrick Model in Virtual Environments: Using a Virtual Human Scenario to Measure Communication Skills Behavior Change*. *3*. https://doi.org/10.3389/frvir.2022.810797

Casal-Otero, L., Catalá, A., Fernández-Morante, C., Taboada, M., Cebreiro, B., & Barro, S. (2023). AI literacy in K-12: A Systematic Literature Review. *International Journal of STEM Education*, *10*, 1–17. https://doi.org/10.1186/s40594-023-00418-7

Choudhary, H., & Bansal, N. (2022). Addressing Digital Divide through Digital Literacy Training Programs: A Systematic Literature Review. *Digital Education Review*. https://doi.org/10.1344/der.2022.41.224-248

Chu-Ke, C., & Dong, Y. (2024). Misinformation and Literacies in the Era of Generative Artificial Intelligence: A Brief Overview and a Call for Future Research. *Emerging Media*, *2*, 70–85. https://doi.org/10.1177/27523543241240285

Folorunso, A., Olanipekun, K., Adewumi, T., & Samuel, B. (2024). A Policy Framework on AI Usage in Developing Countries and Its Impact. *Global Journal of Engineering and Technology Advances*. https://doi.org/10.30574/gjeta.2024.21.1.0192

Gautam, S., Venkit, P. N., & Ghosh, S. (2024). From Melting Pots to Misrepresentations: Exploring Harms in Generative AI. *ArXiv*, *abs/2403.10776*. https://doi.org/10.48550/arxiv.2403.10776

Gong, J., & Popescu, A. (2025). The Impact of a Digital Literacy Intervention on Internet Addiction and Social Skills in Undergraduate Students. *Journal of Adolescent and Youth Psychological Studies*. https://doi.org/10.61838/kman.jayps.6.3.13

Greenhalgh, T., Wherton, J., Papoutsi, C., Lynch, J., Hughes, G., A'Court, C., Hinder, S., Fahy, N., Procter, R., & Shaw, S. (2017). Beyond Adoption: A New Framework for Theorizing and Evaluating Nonadoption, Abandonment, and Challenges to the Scale-Up, Spread, and Sustainability of Health and Care Technologies. *Journal of Medical Internet Research*, *19*. https://doi.org/10.2196/jmir.8775

Grotlüschen, A., Dutz, G., & Skowranek, K. (2024). Writing with Artificial Intelligence? Ad-hoc-Survey Findings Raise Awareness for Critical Literacy at The International Literacy Day. *International Journal of Lifelong Education*, *43*, 371–384. https://doi.org/10.1080/02601370.2024.2349657

Guo, S., Swire-Thompson, B., & Hu, X. (2025). Specific Media Literacy Tips Improve AI-generated Visual Misinformation Discernment. *Cognitive Research: Principles and Implications*, *10*. https://doi.org/10.1186/s41235-025-00648-z

Hatem, R., Simmons, B., & Thornton, J. (2023). A Call to Address AI "Hallucinations" and How Healthcare Professionals Can Mitigate Their Risks. *Cureus*, *15*. https://doi.org/10.7759/cureus.44720

Helal, M., Elgendy, I., Albashrawi, M., Dwivedi, Y., Al-Ahmadi, M., & Jeon, I. (2025). The Impact of Generative AI on Critical Thinking Skills: A Systematic Review, Conceptual Framework and Future Research Directions. *Information Discovery and Delivery*. https://doi.org/10.1108/idd-05-2025-0125

Hwang, Y., & Jeong, S. (2025). Generative Artificial Intelligence and Misinformation Acceptance: An Experimental Test of the Effect of Forewarning About Artificial Intelligence Hallucination. *Cyberpsychology, Behavior, and Social Networking*, *28*, 284–289. https://doi.org/10.1089/cyber.2024.0407

Joshi, S. (2025). Comprehensive Review of AI Hallucinations: Impacts and Mitigation Strategies for Financial and Business Applications. *International Journal of Computer Applications Technology and Research*. https://doi.org/10.7753/ijcatr1406.1003

Julien, H., Gerstle, D., Detlor, B., Rose, T., & Serenko, A. (2022). Digital Literacy Training in Canada, Part 2: Defining and Measuring Success. *The Library Quarterly*, *92*, 87–100. https://doi.org/10.1086/717233

Khan, M. S., Umer, H., & Faruqe, F. (2024). Artificial Intelligence for Low Income Countries. *Humanities and Social Sciences Communications*, *11*. https://doi.org/10.1057/s41599-024-03947-w

Kumar, D. (2025a). Deepfakes, Free Speech, and the Right to Truth: A Comparative Legal Study on Regulating Synthetic Media in the USA, UK, and India. *Advanced International Journal for Research*. https://doi.org/10.63363/aijfr.2025.v06i04.1117

Kumar, D. (2025b). Impact of Infrastructure and Digital Literacy on Cloud-AI EdTech Adoption in Rural India. *Edumania-An International Multidisciplinary Journal*. https://doi.org/10.59231/edumania/9131

Mannuru, N. R., Shahriar, S., Teel, Z., Wang, T., Lund, B., Tijani, S., Pohboon, C. O., Agbaji, D., Alhassan, J., Galley, J., Kousari, R., Ogbadu-Oladapo, L., Saurav, S. K., Srivastava, A., Tummuru, S. P., Uppala, S., & Vaidya, P. (2023). Artificial intelligence in developing countries: The impact of generative artificial intelligence (AI) technologies for development. *Information Development*, *41*, 1036–1054. https://doi.org/10.1177/02666669231200628

Melnyk, S. (2025). Problems of Data Unreliability When Using Artificial Intelligence in Educational Activities. *Cherkasy University Bulletin: Pedagogical Sciences*. https://doi.org/10.31651/2524-2660-2025-3-127-136

Ng, D., Wu, W., Leung, J., Chiu, T., & Chu, S. (2023). Design and validation of the AI Literacy Questionnaire: The Affective, Behavioural, Cognitive and Ethical Approach. *Br. J. Educ. Technol.*, *55*, 1082–1104. https://doi.org/10.1111/bjet.13411

Pan, Q., Reichert, F., Liang, Q., De La Torre, J., & Law, N. (2025). Measuring Digital Literacy Across Ages and Over Time: Development and Validation of A Performance-Based Assessment. *Education and Information Technologies*, *30*, 22065–22100. https://doi.org/10.1007/s10639-025-13592-8

Pinski, M., & Benlian, A. (2024). AI literacy For Users – A Comprehensive Review and Future Research Directions of Learning Methods, Components, and Effects. *Computers in Human Behavior: Artificial Humans*. https://doi.org/10.1016/j.chbah.2024.100062

Praslova, L. (2010). Adaptation of Kirkpatrick's Four Level Model of Training Criteria to Assessment of Learning Outcomes and Program Evaluation in Higher Education. *Educational Assessment, Evaluation and Accountability*, *22*, 215–225. https://doi.org/10.1007/s11092-010-9098-7

Rimal, I. (2025). Effectiveness of Instructional Skills Training for TVET Instructors in Nepal: Evaluation through the Kirkpatrick Model. *Journal of Technical and Vocational Education and Training*. https://doi.org/10.3126/tvet.v19i1.75724

Roe, J., Perkins, M., & Furze, L. (2025). Reflecting Reality, Amplifying Bias? Using Metaphors to Teach Critical AI Literacy. *Journal of Interactive Media in Education*. https://doi.org/10.5334/jime.961

Shao, A. (2025). New sources of inaccuracy? A conceptual framework for studying AI hallucinations. *Harvard Kennedy School Misinformation Review*. https://doi.org/10.37016/mr-2020-182

Sperling, K., Stenberg, C.-J., McGrath, C., Åkerfeldt, A., Heintz, F., & Stenliden, L. (2024). In Search of Artificial Intelligence (AI) Literacy in Teacher Education: A Scoping Review. *Computers and Education Open*. https://doi.org/10.1016/j.caeo.2024.100169

Surbakti, F. (2025). Systematic Literature Review on Generative AI: Ethical Challenges and Opportunities. *International Journal of Advanced Computer Science and Applications*. https://doi.org/10.14569/ijacsa.2025.0160530

Suri, N. A., Festiyed, Azhar, M., Yerimadesi, Ahda, Y., & Alberida, H. (2025). Measuring What Matters: A Systematic Review and Vosviewer-Based Bibliometric Approach to Digital Literacy Assessment Instruments, Competency Dimensions and Challenges in Education. *Research in Learning Technology*. https://doi.org/10.25304/rlt.v33.3413

Wan, Y., Subramonian, A., Ovalle, A., Lin, Z., Suvarna, A., Chance, C., Bansal, H., Pattichis, R., & Chang, K.-W. (2024). Survey of Bias in Text-to-Image Generation: Definition, Evaluation, and Mitigation. *ArXiv*, *abs/2404.01030*. https://doi.org/10.48550/arxiv.2404.01030

Wang, K., Liu, P., Zhang, J., Zhong, J., Luo, X., Huang, J., & Zheng, Y. (2023). Effects of Digital Game-Based Learning on Students' Cyber Wellness Literacy, Learning Motivations, and Engagement. *Sustainability*. https://doi.org/10.3390/su15075716

Yim, I. (2024). A Critical Review of Teaching and Learning Artificial Intelligence (AI) Literacy: Developing An Intelligence-Based AI Literacy Framework for Primary School Education. *Comput. Educ. Artif. Intell.*, *7*, 100319. https://doi.org/10.1016/j.caeai.2024.100319