JURNAL WYDIA CLIMAGO
ISSN : 2797-0078 (Online)
Volume 7, Nomor 1, Edisi Juni 2025, Halaman 35-42
https://ejournal-pusdiklat.bmkg.go.id/index.php/climago

RETROSPEKTIF DAMPAK PENGGUNAAN KECERDASAN BUATAN GENERATIF TERHADAP KOMPETENSI TEKNIS DAN KONSEPTUAL PEGAWAI PADA PELATIHAN TEKNIS PPSDM MKG

Yulyanto Dwi Prastyo Nugroho
Pusat Pengembangan Sumber Daya Manusia Meteorologi Klimatologi dan Geofisika

Informasi Artikel

Sejarah Artikel:

Accepted Mei 30, 2025

Kata Kunci :
cognitive offloading; kecerdasan buatan generatif; evaluasi pelatihan; kompetensi teknis-prosedural; analisis ex post facto; PPSDM MKG

ABSTRAK

Pelatihan teknis di lingkungan korporat semakin banyak memanfaatkan kecerdasan buatan generatif (generative AI/GenAI) sebagai alat bantu peserta, namun institusi penyelenggara jarang memiliki cara sistematis untuk memastikan bahwa nilai akhir yang diperoleh peserta benar-benar mencerminkan kompetensi yang didapat, bukan sekadar keluaran AI yang disalin. Artikel ini menyajikan analisis retrospektif (ex post facto) terhadap data penilaian yang telah tersedia dari dua pelatihan teknis yang diselenggarakan Pusat Pendidikan dan Pelatihan Sumber Daya Manusia Meteorologi Klimatologi dan Geofisika (PPSDM MKG) bagi pegawai Badan Meteorologi, Klimatologi, dan Geofisika (BMKG), tanpa memerlukan pengumpulan data primer baru. Pada Inhouse Training Security Operation Center (n=39), analisis korelasi menunjukkan bahwa skor posttest tertulis tidak berhubungan signifikan dengan skor praktik cyberdrill ($r=0,09$; $p=0,57$) dan berhubungan negatif signifikan dengan skor paparan ($r=-0,36$; $p=0,03$); klasifikasi kuadran berbasis median split menunjukkan 30,8% peserta berada pada pola "kompetensi semu" (skor tulis tinggi, kinerja praktik dan paparan rendah). Pada Pelatihan Pemrograman Python dan Analisis Data Python ($n\approx40$), tinjauan pengajar terhadap laporan project dan pemaparan menunjukkan bahwa pemahaman konseptual peserta (90%) secara konsisten lebih tinggi dibanding pemahaman teknis-prosedural/coding (75%), mengindikasikan bahwa penggunaan AI untuk pembangkitan kode terkonsentrasi pada lapisan kompetensi teknis, bukan lapisan pemahaman domain. Kedua temuan digunakan untuk mengusulkan kerangka evaluasi dua-lapis (domain/analitis vs teknis-prosedural) sebagai pelengkap penilaian skor tunggal dalam pelatihan teknis berbasis AI, serta agenda verifikasi lanjutan berupa unplugged retention test dan survei pola penggunaan AI pada siklus pelatihan berikutnya.

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

Penulis Koresponden
Yulyanto Dwi Prastyo Nugroho
Pusat Pengembangan Sumber Daya Manusia MKG, BMKG, Indonesia,
Jl. Perhubungan I No.5 Pondok Betung, Bintaro, Kec. Pd. Aren, Kota Tangerang Selatan, Banten 15221.
Email:yulyanto.nugroho @gmail.com

1. PENDAHULUAN

Sebagai unit pengembangan kompetensi teknis di lingkungan BMKG, PPSDM MKG secara rutin menyelenggarakan pelatihan yang menuntut penguasaan keterampilan praktis—mulai dari

.
operasi keamanan siber, pemrograman, hingga analisis data meteorologi-klimatologi berbasis kode. Ketersediaan AI generatif yang mudah diakses menimbulkan persoalan praktis bagi penyelenggara pelatihan: peserta berpotensi menyelesaikan tugas dan memperoleh nilai yang baik dengan bantuan AI tanpa disertai penguasaan materi yang memadai. Persoalan ini penting karena kompetensi yang diukur dalam pelatihan teknis semacam ini pada akhirnya dipertaruhkan pada situasi kerja nyata—misalnya respons terhadap insiden siber atau pengolahan data cuaca ekstrem—yang menuntut kompetensi yang benar-benar melekat pada pegawai, bukan sekadar tercatat pada lembar nilai.

Kekhawatiran ini sejalan dengan konsep cognitive offloading, yaitu penggunaan alat eksternal untuk mengurangi tuntutan kognitif internal suatu tugas [1]—strategi yang adaptif dalam batas wajar, tetapi berisiko mengikis keterlibatan kognitif bila dilakukan berlebihan. Eksperimen terkontrol pada siswa sekolah menengah menemukan bahwa akses ke AI tutor generik meningkatkan skor latihan sebesar 48%, namun menyebabkan performa ujian tanpa AI 17% lebih buruk dibanding kelompok yang sama sekali tidak menggunakan AI [2]. Studi EEG menemukan konektivitas neural dan rasa kepemilikan tulisan terendah pada kelompok pengguna AI penuh dibanding kelompok yang menulis tanpa bantuan [3]. Fan dkk. menyebut penurunan pemantauan diri terhadap proses belajar akibat ketergantungan AI sebagai metacognitive laziness [4], sementara pada ranah pengambilan keputusan berbantuan AI secara umum, kecenderungan mengikuti keluaran otomatis tanpa evaluasi kritis dikenal sebagai automation bias [5], yang dapat diredam melalui cognitive forcing function yang mendorong pengguna berpikir sebelum menerima saran AI [6].

Perbedaan mendasar antara persoalan di PPSDM MKG dan sebagian besar literatur tersebut terletak pada karakteristik pesertanya: pegawai dewasa yang pada umumnya telah memiliki pemahaman domain yang memadai atas pekerjaan mereka, namun terkendala pada penguasaan bahasa pemrograman sebagai alat implementasi. Pertanyaan yang relevan bukan lagi sekadar "apakah AI menurunkan kompetensi peserta" secara umum, melainkan "pada lapisan kompetensi mana—domain/konseptual atau teknis-prosedural—dampak penggunaan AI benar-benar terjadi". Artikel ini menjawab pertanyaan tersebut dengan menganalisis ulang data penilaian yang telah dimiliki PPSDM MKG dari dua pelatihan yang berbeda karakter tugasnya, tanpa memerlukan pengumpulan data primer baru: (1) Inhouse Training Security Operation Center, yang menilai peserta melalui tes tulis dan praktik simulasi (cyberdrill) serta paparan; dan (2) Pelatihan Pemrograman Python dan Analisis Data Python, yang menilai peserta melalui laporan project berbasis kode dan pemaparan lisan.

Kontribusi artikel ini bersifat ganda. Pertama, secara metodologis, artikel ini menunjukkan bahwa institusi pelatihan dapat mendeteksi indikasi kesenjangan skor-kompetensi tanpa perlu penelitian eksperimental baru, cukup dengan menganalisis ulang data penilaian multi-komponen yang telah rutin dikumpulkan. Kedua, secara konseptual, artikel ini mengusulkan bahwa dampak cognitive offloading akibat AI generatif bersifat spesifik-lapisan (layer-specific)—dapat terkonsentrasi pada kompetensi teknis-prosedural tanpa serta-merta mengikis kompetensi domain—sebuah nuansa yang belum banyak dibahas eksplisit pada literatur cognitive offloading yang masih didominasi konteks pendidikan formal [7], [8].

2. METODE PENELITIAN

Penelitian ini menggunakan desain ex post facto (causal-comparative), yaitu analisis terhadap data yang telah terjadi atau tercatat tanpa manipulasi atau intervensi tambahan [9]. Desain ini dipilih karena data penilaian yang relevan—skor tes tulis, praktik, dan paparan—telah tersedia secara lengkap dari penyelenggaraan pelatihan reguler PPSDM MKG, sehingga tidak diperlukan pengumpulan data primer baru. Validitas pendekatan sampel terbatas/kasus institusional semacam ini bergantung pada kejelasan tujuan analisis dan transparansi pelaporan keterbatasannya, bukan semata pada besar sampel [10].

Sumber data pertama adalah 39 peserta Inhouse Training Security Operation Center, dengan empat komponen penilaian yang telah dilaksanakan oleh pengajar: pretest tertulis, posttest tertulis, penilaian praktik simulasi keamanan siber (cyberdrill), dan penilaian paparan/presentasi peserta. Analisis dilakukan dengan statistik deskriptif, uji korelasi Pearson antara skor posttest dan dua

ukuran kinerja aplikatif (cyberdrill dan paparan), serta klasifikasi kuadran berdasarkan median split terhadap skor posttest dan rata-rata skor praktik-paparan.

Sumber data kedua adalah sekitar 40 peserta Pelatihan Pemrograman Python dan Analisis Data Python, dengan penilaian pengajar terhadap laporan project berbasis kode serta pemaparan lisan. Berdasarkan tinjauan laporan dan sesi tanya-jawab paparan, pengajar mengategorikan peserta pada dua dimensi terpisah: pemahaman konseptual (logika dan tujuan analisis) dan pemahaman teknis-prosedural (kemampuan menjelaskan dan memodifikasi kode program yang disusun, termasuk saat dibantu AI). Kriteria kualitatif yang digunakan pengajar meliputi kesesuaian kode dengan konteks tugas, kualitas respons saat tanya-jawab, dan penanganan revisi/debugging kode. Karena kategorisasi ini bersifat penilaian retrospektif pengajar dan bukan instrumen psikometrik terstandardisasi, hasilnya dilaporkan sebagai proporsi deskriptif, dengan estimasi tiga kelompok gabungan (kompetensi utuh, gap coding-spesifik, dan kesulitan menyeluruh) dihitung berdasarkan asumsi nested antara kedua dimensi—sebuah keterbatasan yang perlu diverifikasi lebih lanjut apabila data tingkat-individu tersedia

3. HASIL DAN PEMBAHASAN

3.1 Divergensi Skor Tulis dan Kinerja Praktik pada Pelatihan SOC

Skor posttest menunjukkan efek plafon yang jelas (M=95,7; SD=5,1; median=97,0; rentang 80–100), jauh lebih sempit dibanding sebaran skor pretest (M=81,8; SD=14,2; rentang 20–100)—hampir seluruh peserta "terlihat" menguasai materi di akhir pelatihan meski titik keberangkatan pengetahuan mereka sangat beragam. Korelasi pretest terhadap posttest tergolong lemah-sedang (r=0,39; p=0,015), menunjukkan performa akhir tes tulis hanya sebagian kecil dijelaskan oleh pengetahuan awal peserta.

Temuan yang lebih substantif adalah bahwa skor posttest tidak berkorelasi signifikan dengan skor cyberdrill (r=0,09; p=0,57), dan berkorelasi negatif signifikan dengan skor paparan (r=−0,36; p=0,03). Pada kohort ini, skor tes tulis pascapelatihan pada dasarnya tidak memprediksi—dan untuk paparan, berbanding terbalik dengan—kinerja aplikatif peserta. Tabel 1 merangkum klasifikasi kuadran berdasarkan median split pada posttest (median=97,0) dan rata-rata cyberdrill-paparan (median=82,4).

Kuadran	Pola Skor	Interpretasi	n (%)
Kompetensi utuh	Tulis tinggi, praktik+paparan tinggi	Pemahaman konsisten di kedua instrumen	12 (30,8%)
Kompetensi semu	Tulis tinggi, praktik+paparan rendah	Kesenjangan kompetensi tersembunyi di balik skor tulis	12 (30,8%)
Kompetensi tersembunyi	Tulis rendah, praktik+paparan tinggi	Kemungkinan keterbatasan instrumen tes tulis	9 (23,1%)
Kesulitan menyeluruh	Tulis rendah, praktik+paparan rendah	Kemungkinan kendala di luar isu AI	6 (15,4%)

Tabel 1. Sebaran Peserta Berdasarkan Pola Divergensi Skor Tulis dan Kinerja Aplikatif (n=39)

Hampir sepertiga peserta (30,8%) berada pada kuadran "kompetensi semu"—proporsi yang setara dengan kelompok "kompetensi utuh". Sebagai ilustrasi, satu peserta dengan pretest 20 mencapai posttest 83 (kenaikan 63 poin, terbesar pada kohort), namun skor cyberdrill (80) dan paparan (85,1) hanya berada di kisaran rata-rata kohort—peningkatan skor tulis yang besar tidak serta-merta diikuti kompetensi aplikatif yang setara. Sebagai temuan tambahan yang relevan,

.
rekonstruksi statistik terhadap formula skor akhir yang digunakan PPSDM MKG (R^2=0,9997) menunjukkan bobot Cyberdrill dan Paparan (masing-masing ≈40%) jauh lebih besar dibanding Posttest (≈15%)—desain evaluasi institusional PPSDM MKG tampak telah secara implisit mengantisipasi persoalan yang diangkat artikel ini, meski belum dirumuskan secara eksplisit dalam kerangka cognitive offloading.

Konsisten dengan pola ini, model evaluasi pelatihan Kirkpatrick yang lazim dipakai institusi telah lama dikritik karena Level 2 (learning/recall)—yang paling dekat diwakili tes tulis—tidak cukup merefleksikan Level 3–4 (behavior dan results) yang diwakili kinerja praktik dan hasil kerja nyata [11]. Temuan pada Bagian ini memberi bukti empiris konkret bagi kritik tersebut: tes tulis pre-test/post-test tidak memadai sebagai satu-satunya atau penentu akhir kompetensi pelatihan teknis, dan perlu ditriangulasi dengan penilaian praktik serta idealnya unplugged assessment (penilaian tanpa akses AI) [2].

3.2 Disparitas Kompetensi Domain dan Teknis-Prosedural pada Pelatihan Python

Pada Pelatihan Pemrograman Python dan Analisis Data Python, pengajar melaporkan bahwa 90% peserta menunjukkan pemahaman konseptual yang memadai—memahami logika dan tujuan analisis yang ingin dicapai—namun hanya 75% yang menunjukkan pemahaman teknis-prosedural yang memadai (mampu menjelaskan dan memodifikasi kode program yang mereka susun). Selisih 15 poin persentase ini (Tabel 2) menjadi indikasi kuantitatif bahwa penggunaan AI untuk pembangkitan kode cenderung terkonsentrasi pada lapisan teknis-prosedural, sementara kompetensi domain/analitis peserta relatif terjaga

Hasil penelitian ini membuktikan bahwa model Project Based Learning (PBL) mampu diterapkan peserta dengan baik pada tahap akhir dalam proyek implementasi di tempat kerja. Keberhasilan awal tergambar dari hasil belajar di dalam kelas, dimana peserta mengadopsi proses belajar dengan baik sehingga mampu menyusun action plan dalam bentuk PBL. Keberhasilan ini juga tergambar dari peserta yang sebagian besar mampu mengeksekusi proyek action plan-nya. Karena terbatasnya jangka waktu eksekusi yang hanya satu bulan, proyek yang telah dilaksanakan belum sampai pada tahap final, namun kelanjutan proyek ini diserahkan kepada masing-masing peserta sejauh mana dapat diselesaikan sampai tahap akhir.

Berdasarkan hasil pengalaman implementasi yang dipresentasikan, 63% peserta mampu mengeksekusi action plan dengan membangun komunikasi dan kolaborasi bersama rekan kerja. Keberhasilan lain yang patut ditonjolkan adalah peserta mampu menjalankan kegiatan dalam bentuk program yang berkaitan dengan pemanfaatan informasi cuaca dan iklim sesuai karakteristik wilayah kerja masing-masing. Dengan menerapkan model PBL, pelatihan ini mampu menjawab tantangan bagi peserta untuk membuat proyek atau kegiatan dengan melibatkan berbagai pihak pendukung dalam implementasinya.

Dimensi Kompetensi	Kompetensi Memadai (%)	Kompetensi Terbatas (%)	Basis Penilaian
Konseptual/domain	90	10	Laporan & tanya-jawab paparan
Teknis-prosedural (coding)	75	25	Kode & tanya-jawab paparan

Tabel 2. Proporsi Pemahaman Konseptual vs Teknis-Prosedural (n≈40)

Dengan asumsi nested—peserta yang kurang memahami konsep juga cenderung kurang memahami implementasi teknisnya—populasi peserta dapat diestimasi terbagi tiga: kompetensi utuh (≈75%), gap coding-spesifik (≈15%, memahami konsep namun kurang memahami kode), dan kesulitan menyeluruh (≈10%). Kelompok gap coding-spesifik adalah yang paling relevan secara konseptual: kelompok ini menunjukkan bahwa AI dapat berfungsi murni sebagai jembatan pada

lapisan teknis tanpa mengorbankan kompetensi domain—pola yang berbeda secara kualitatif dari penggunaan delegatif menyeluruh yang banyak dilaporkan pada konteks pendidikan formal [2], [3].

Pola ini konsisten dengan riset interaksi manusia-AI pada konteks pemrograman. Studi pada programmer pemula menemukan dampak asisten kode berbasis AI generatif tidak seragam: peserta dengan kemampuan metakognitif kuat cenderung terbantu, sementara yang kesulitan menilai kapan saran AI benar/salah justru dirugikan—dengan dua pola yang teridentifikasi: mengetik kode yang meniru saran AI tanpa benar-benar memahaminya (shepherding), dan menerima kode keliru yang berujung pada debugging berkepanjangan [12]. Studi lain pada pembelajar pemrograman pemula turut menemukan variasi signifikan antara interaksi mendalam dengan generator kode berbasis LLM dan sekadar penyalinan keluaran [13]. Temuan ini memperkuat bahwa indikator delegatif yang relevan pada konteks coding bukan sekadar "menggunakan AI atau tidak", melainkan kemampuan peserta menjelaskan, memodifikasi, dan melakukan debugging kode yang dihasilkan AI—indikator yang secara alami tersedia pada laporan project dan sesi tanya-jawab paparan yang telah menjadi bagian penilaian PPSDM MKG.

3.3 Kerangka Evaluasi Dua-Lapis: Domain vs Teknis-Prosedural

Dua temuan di atas, meski berasal dari pelatihan dengan karakter tugas berbeda, mengarah pada kesimpulan yang sama: skor evaluasi akhir yang tinggi/seragam secara administratif dapat menyembunyikan kesenjangan kompetensi aplikatif, dan kesenjangan tersebut tidak menyebar merata ke seluruh aspek kompetensi peserta, melainkan terkonsentrasi pada lapisan tertentu. Berdasarkan pola ini, artikel ini mengusulkan kerangka evaluasi dua-lapis bagi pelatihan teknis berbasis AI: (a) kompetensi domain/analitis—pemahaman tentang apa yang perlu dilakukan dan mengapa—dinilai melalui paparan dan tanya-jawab lisan; dan (b) kompetensi teknis-prosedural—kemampuan mengimplementasikan pemahaman tersebut secara teknis—dinilai melalui praktik langsung (cyberdrill, penulisan/modifikasi kode) tanpa akses AI. Penggunaan AI generatif berpotensi bersifat augmentatif pada satu lapisan namun delegatif pada lapisan lain secara bersamaan, sebagaimana ditunjukkan pola gap coding-spesifik pada Bagian 3.2—sehingga penilaian "kompetensi" secara tunggal berisiko mengaburkan perbedaan penting ini. Kerangka ini juga selaras dengan indikator posisi kontinum delegatif-augmentatif berupa rasio override/koreksi mandiri peserta terhadap saran AI yang diusulkan pada penelitian interaksi manusia-AI terbaru [14], serta dengan pembedaan cognitive surrender—penyerahan kendali kognitif tanpa verifikasi—dari cognitive offloading klasik yang tetap mempertahankan struktur berpikir [15].

3.4 Implikasi Praktis dan Agenda Verifikasi Lanjutan

Temuan ini menghasilkan tiga implikasi praktis bagi PPSDM MKG. Pertama, mempertahankan dan memperluas praktik penilaian multi-komponen (tes tulis, praktik, paparan) yang telah berjalan, dengan pembobotan yang secara eksplisit mencerminkan kerangka dua-lapis di atas—sejalan dengan temuan bahwa formula skor PPSDM MKG saat ini telah memberi bobot lebih besar pada komponen praktik dan paparan. Kedua, tambahkan unplugged retention test—penilaian tanpa akses AI, idealnya 4–8 minggu pascapelatihan—untuk memverifikasi apakah kompetensi yang tampak di akhir pelatihan benar-benar bertahan, terutama bagi peserta pada kuadran "kompetensi semu". Ketiga, pada pelatihan berbasis kode, jadikan kemampuan menjelaskan dan memodifikasi kode saat tanya-jawab paparan sebagai indikator penilaian eksplisit, bukan sekadar bagian informal dari penilaian pengajar.

Untuk memverifikasi kausalitas yang saat ini masih bersifat indikatif, artikel ini merumuskan dua proposisi yang dapat diuji pada siklus pelatihan berikutnya: H1—pola penggunaan AI yang lebih delegatif (diukur melalui survei self-report singkat atau log penggunaan pada platform AI yang dikelola institusi) berhubungan negatif dengan skor retensi pada unplugged test tunda; dan H2—kesenjangan skor tulis-praktik ("kompetensi semu") lebih besar pada peserta dengan pola penggunaan AI delegatif dibanding augmentatif. Kedua proposisi ini dapat diuji tanpa mengubah struktur pelatihan yang telah berjalan, cukup dengan menambahkan satu instrumen survei singkat dan satu sesi retention test tunda.

Sebagai keterbatasan, kedua sumber data berasal dari satu institusi dengan jumlah peserta terbatas ($n=39$ dan $n\approx40$), tanpa desain kontrol maupun pencatatan perilaku penggunaan AI secara

.
langsung, sehingga generalisasi ke konteks pelatihan teknis korporat yang lebih luas memerlukan kehati-hatian dan idealnya replikasi pada pelatihan dan angkatan lain. Estimasi tiga kelompok pada Bagian 3.2 juga bergantung pada asumsi nested yang belum diverifikasi pada tingkat individu

4. KESIMPULAN

Analisis retrospektif terhadap data dua pelatihan teknis PPSDM MKG menunjukkan bahwa skor evaluasi akhir yang tinggi dan seragam secara administratif tidak selalu mencerminkan kompetensi aplikatif yang setara. Pada Inhouse Training SOC, hampir sepertiga peserta menunjukkan pola "kompetensi semu"—skor tes tulis tinggi namun kinerja praktik dan paparan di bawah median—dengan skor tes tulis yang bahkan berkorelasi negatif dengan skor paparan. Pada Pelatihan Pemrograman Python, disparitas antara pemahaman konseptual (90%) dan pemahaman teknis-prosedural (75%) menunjukkan bahwa dampak penggunaan AI generatif bersifat spesifik-lapisan—terkonsentrasi pada kompetensi teknis-prosedural tanpa serta-merta mengikis kompetensi domain/analitis peserta.

Kerangka evaluasi dua-lapis yang diusulkan—membedakan kompetensi domain/analitis dan teknis-prosedural—memberikan alat diagnostik konseptual sekaligus praktis bagi PPSDM MKG untuk menilai kualitas hasil pelatihan secara lebih presisi, tanpa memerlukan perombakan besar terhadap struktur penilaian yang telah berjalan. Studi lanjutan yang menguji proposisi H1 dan H2 melalui unplugged retention test dan survei pola penggunaan AI akan memungkinkan verifikasi kausal atas pola yang saat ini baru teridentifikasi secara deskriptif-korelasional, sekaligus menjadi dasar bagi pengembangan instrumen pengukuran pola penggunaan AI yang dapat diterapkan secara rutin pada siklus pelatihan teknis PPSDM MKG berikutnya

REFERENSI

[1] E. F. Risko and S. J. Gilbert, "Cognitive offloading," Trends Cogn. Sci., vol. 20, no. 9, pp. 676–688, 2016.

[2] H. Bastani, O. Bastani, A. Sungu, H. Ge, Ö. Kabakcı, and R. Mariman, "Generative AI without guardrails can harm learning: Evidence from high school mathematics," Proc. Natl. Acad. Sci. U.S.A., vol. 122, Art. no. e2422633122, 2025.

[3] N. Kosmyna, E. Hauptmann, Y. T. Yuan, J. Situ, X.-H. Liao, A. V. Beresnitzky, I. Braunstein, and P. Maes, "Your brain on ChatGPT: Accumulation of cognitive debt when using an AI assistant for essay writing task," arXiv:2506.08872, 2025.

[4] Y. Fan, L. Tang, H. Le, K. Shen, S. Tan, Y. Zhao, Y. Shen, X. Li, and D. Gašević, "Beware of metacognitive laziness: Effects of generative artificial intelligence on learning motivation, processes, and performance," Br. J. Educ. Technol., vol. 56, no. 2, pp. 489–530, 2025.

[5] R. Khera, M. A. Simon, and J. S. Ross, "Automation bias and assistive AI: Risk of harm from AI-driven clinical decision support," JAMA, vol. 330, no. 23, pp. 2255–2257, 2023.

[6] Z. Buçinca, M. B. Malaya, and K. Z. Gajos, "To trust or to think: Cognitive forcing functions can reduce overreliance on AI in AI-assisted decision-making," Proc. ACM Hum.-Comput. Interact., vol. 5, no. CSCW1, pp. 1–21, 2021.

[7] M. Gerlich, "AI tools in society: Impacts on cognitive offloading and the future of critical thinking," Societies, vol. 15, no. 1, Art. no. 6, 2025.

[8] Y. Hou et al., "The role of critical thinking on undergraduates' reliance behaviours on generative AI in problem-solving," Br. J. Educ. Technol., 2025, doi: 10.1111/bjet.13613.

[9] K. Ekuma, "Artificial intelligence and automation in human resource development: A systematic review," Hum. Resour. Dev. Rev., 2024, doi: 10.1177/15344843231224009.

[10] "Navigating sample size estimation for qualitative research," J. Med. Evid., vol. 5, no. 2, pp. 133–139, 2024, doi: 10.4103/JME.JME_59_24.

[11] B. Ambu-Saidi, C. Y. Fung, K. Turner, and A. S. S. Lim, "A critical review on training evaluation models: A search for future agenda," J. Cogn. Sci. Hum. Dev., vol. 10, no. 1, pp. 1–15, 2024.

[12] B. N. Reeves, S. Chen, N. Alkhouri, S. Fowler, B. A. Becker, P. Denny, and J. Prather, "The widening gap: The benefits and harms of generative AI for novice programmers," arXiv:2405.17739, 2024.

[13] M. Kazemitabaar, X. Hou, A. Henley, B. J. Ericson, D. Weintrop, and T. Grossman, "How novices use LLM-based code generators to solve CS1 coding tasks in a self-paced learning environment," arXiv:2309.14049, 2023.

[14] S. Baldeo, "Generative artificial intelligence reliance and executive function attenuation: Behavioral evidence of cognitive offload in high-use adults," Technol. Mind Behav., 2026, doi: 10.1037/tmb0000191.

[15] S. D. Shaw and G. Nave, "Thinking—fast, slow, and artificial: The rise of cognitive surrender," Wharton School Res. Pap., 2026.

[16] S. Rismanchian, H. Uzun, J. Matayoshi, E. Cosyn, and E. Kurd-Misto, "Faster completion, less learning: Generative AI reduced study time on math problems and the knowledge they build," arXiv:2605.21629, 2026.

[17] R. E. Wang, A. T. Ribeiro, C. D. Robinson, S. Loeb, and D. Demszky, "Tutor CoPilot: A human-AI approach for scaling real-time expertise," arXiv:2410.03017, 2024.

[18] E. Brynjolfsson, D. Li, and L. R. Raymond, "Generative AI at work," Q. J. Econ., vol. 140, no. 2, pp. 889–942, 2025.