

Inteligência artificial na educação e neuroplasticidade: contribuições, limites e desafios para a aprendizagem

Artificial intelligence in education and neuroplasticity: contributions, limits and challenges for learning

Daniela Hayoshi

Resumo

A literatura sobre inteligência artificial (IA) na educação mede quase exclusivamente desempenho, percepção e engajamento; mudanças cerebrais associadas à neuroplasticidade praticamente não foram medidas. Esta revisão narrativa, com critérios explícitos de busca e classificação, separa três tipos de evidência: (a) resultados de aprendizagem (notas, retenção, transferência), (b) medidas comportamentais de esforço cognitivo e (c) medidas neurofisiológicas diretas (EEG, fNIRS). Os estudos mais robustos de desempenho mostram que o efeito da IA depende do desenho da atividade: um ensaio randomizado com cerca de mil estudantes de ensino médio encontrou ganho de 48% a 127% durante a prática assistida e, no exame sem IA, queda de desempenho no grupo com acesso irrestrito ao modelo, neutralizada quando o sistema foi configurado para dar pistas em vez de respostas. A meta-análise mais citada a favor da IA ($g = 0,867$) foi retratada em 2026 por erros de análise e inclusão de estudos inválidos. Os poucos estudos com medidas cerebrais são pequenos (n entre 24 e 54), de curta duração, sem revisão por pares ou com resultados nulos, e registram engajamento momentâneo — não plasticidade estrutural ou funcional duradoura. Conclui-se que não há base para afirmar que a IA, por si, melhora ou prejudica a neuroplasticidade. O que a evidência sustenta é mais estreito: atividades que substituem o esforço de recuperação e elaboração tendem a reduzir a aprendizagem medida; atividades que preservam esse esforço tendem a manter ou ampliar ganhos. Propõe-se uma agenda com desenhos longitudinais, medidas de transferência e neuroimagem pré/pós com controle ativo.

Palavras-chave: inteligência artificial na educação; neuroplasticidade; esforço cognitivo; descarregamento cognitivo; sistemas tutores inteligentes; eletroencefalografia.

Abstract

The literature on artificial intelligence (AI) in education focuses almost exclusively on performance, perception, and engagement; brain changes associated with neuroplasticity have rarely been measured. This narrative review, using explicit search and classification criteria, distinguishes three types of evidence: (a) learning outcomes (grades, retention, and transfer), (b) behavioral measures of cognitive effort, and (c) direct neurophysiological measures (EEG and fNIRS). The most robust performance studies indicate that the effects of AI depend on task design: a randomized trial involving approximately one thousand high school students found gains ranging from 48% to 127% during AI-assisted practice, while performance declined on an exam taken without AI among students with unrestricted access to the model; this decline was eliminated when the system was configured to provide hints rather than answers. The most frequently cited meta-analysis supporting AI ($g = 0.867$) was retracted in 2026 due to analytical errors and the inclusion of invalid studies. The few studies using brain-based measures are small (n ranging from 24 to 54), short in duration, unpublished in peer-reviewed journals or reporting null findings, and assess momentary engagement rather than lasting structural or functional plasticity. It is therefore concluded that there is no basis for claiming that AI, by itself, either enhances or impairs neuroplasticity. The evidence supports a narrower conclusion: activities that replace retrieval and elaborative effort tend to reduce measured learning, whereas activities that preserve such effort tend to maintain or enhance learning gains. A research agenda is proposed incorporating longitudinal designs, transfer measures, and pre- and post-intervention neuroimaging with active control conditions.

Keywords: artificial intelligence in education; neuroplasticity; cognitive effort; cognitive offloading; intelligent tutoring systems; electroencephalography.

1. Introdução

Desde o lançamento público de modelos de linguagem de grande porte em novembro de 2022, a expressão "IA e neuroplasticidade" passou a circular em dois registros opostos e igualmente mal fundamentados. No primeiro, a IA "potencializa" a plasticidade cerebral ao personalizar o ensino; no segundo, ela "atrofia" o cérebro ao fazer o trabalho pelo aluno. Ambos tratam a neuroplasticidade como uma variável que as pesquisas educacionais mediriam rotineiramente. Não é o caso.

O problema é de categoria. A neuroplasticidade é um conjunto de mecanismos biológicos (sinápticos, celulares, estruturais) que só pode ser aferido por medidas neurais — imagem estrutural, eletrofisiologia, marcadores moleculares — ou inferido, com cautela, de mudanças comportamentais duradouras. A pesquisa sobre IA na educação, por sua vez,

mede notas, retenção, percepção e tempo de tarefa. Extrapolar de uma categoria para outra sem medida intermediária é o erro que John Bruer, em 1997, chamou de "ponte longa demais" entre neurociência e prática educacional (Bruer, 1997). A IA generativa reativou essa ponte.

Este artigo responde a uma pergunta deliberadamente restritiva: o que as pesquisas realmente demonstram sobre IA na aprendizagem e quais delas mediram algo relacionado à neuroplasticidade, em vez de apenas desempenho ou engajamento? A tese defendida é que a IA pode apoiar experiências de aprendizagem personalizadas, mas não há base para afirmar que usar IA, por si só, melhora ou prejudica a neuroplasticidade. O que a evidência mostra é que o efeito depende do tipo de atividade — especificamente, de quanto esforço de recuperação e elaboração a atividade preserva — e que a parte "cerebral" da pergunta ainda está, em grande medida, sem dados.

O texto segue cinco movimentos: define neuroplasticidade sem simplificações (seção 3); descreve os usos educacionais da IA (seção 4); separa resultados de aprendizagem de medidas cerebrais (seções 5 e 6); discute esforço cognitivo, papel do educador, privacidade e acesso (seção 7); e encerra com as lacunas que uma pesquisa futura precisaria fechar (seção 8).

2. Método

Trata-se de uma revisão narrativa crítica com critérios explícitos, não de uma revisão sistemática. A escolha é deliberada: o campo é recente (menos de quatro anos de literatura sobre IA generativa em educação), heterogêneo e marcado por um caso de retratação que exige leitura qualitativa dos estudos, não apenas agregação numérica.

Estratégia de busca. Foram consultadas, entre setembro e outubro de 2026, bases e repositórios de acesso aberto (arXiv, PubMed Central, SSRN, DOAJ, portais de editoras) e buscadores acadêmicos, com combinações dos termos em inglês e português: *artificial intelligence, generative AI, ChatGPT, intelligent tutoring system, learning outcomes, metaanalysis, neuroplasticity, EEG, fNIRS, cognitive offloading, cognitive load*. Foram rastreadas também as referências dos estudos incluídos e as críticas publicadas a eles (comentários, pré-prints de replicação, notas de retratação).

Critérios de inclusão. (1) Estudos empíricos com grupo de comparação ou meta-análises sobre IA em contexto de aprendizagem; (2) estudos que coletaram qualquer medida neural (EEG, fNIRS, imagem) durante ou após uso de IA em tarefa de aprendizagem; (3) trabalhos de referência em neuroplasticidade humana induzida por aprendizagem, necessários para definir o conceito. Foram excluídos estudos puramente de opinião, de percepção sem

medida de desempenho e trabalhos sobre IA como objeto de ensino (por exemplo, cursos de programação de IA).

Classificação da evidência. Cada estudo foi classificado em três níveis, que estruturam as seções 5 e 6:

Nível	O que mede	Exemplos de medida	Permite inferir plasticidade?
A — Resultado de aprendizagem	Conhecimento ou habilidade após a intervenção	Nota em exame sem IA, retenção tardia, transferência	Só indiretamente e só se durável
B — Esforço e comportamento	O que o aluno faz durante a tarefa	Tempo, tipo de pergunta, recordação do próprio texto, autoavaliação	Não; é proxy de processo
C — Medida neural	Atividade ou estrutura cerebral	Conectividade EEG, ativação pré-frontal por fNIRS, volume de substância cinzenta	Sim, se houver pré/pós, controle ativo e persistência

Limites do método. Não houve avaliação formal de risco de viés por instrumento padronizado, nem dupla extração de dados. Toda afirmação quantitativa no texto cita a fonte primária ou, quando a fonte primária não pôde ser aberta, a nota de retratação ou o comentário publicado que a reporta. Onde a evidência é um pré-print não revisado por pares, isso está indicado.

3. Neuroplasticidade: o que é e o que não é

Neuroplasticidade é a capacidade do sistema nervoso de alterar sua estrutura e seu funcionamento em resposta à experiência, ao uso e à lesão. O termo cobre fenômenos de escalas muito diferentes, e é a confusão entre essas escalas que alimenta as afirmações simplificadas sobre IA.

Três escalas. Na escala sináptica, a plasticidade de longo prazo (potenciação e depressão de longa duração, LTP e LTD) modifica a eficácia da transmissão entre neurônios em minutos a horas e é o mecanismo celular mais bem caracterizado da memória. Na escala de circuito, mudanças de conectividade funcional reorganizam quais regiões trabalham juntas numa tarefa; é o que EEG e fMRI de repouso ou de tarefa captam. Na escala estrutural, semanas ou meses de prática produzem alterações mensuráveis em substância cinzenta e branca — o volume do hipocampo posterior maior em taxistas de Londres (Maguire et al., 2000) e o

aumento transitório de substância cinzenta em áreas visual-motoras após três meses de treino de malabarismo (Draganski et al., 2004) são os exemplos canônicos.

Condições para que ocorra. A plasticidade estrutural induzida por aprendizagem não é automática. O modelo de Lövdén e colegas (2010) descreve-a como resposta a um descompasso prolongado entre a demanda da tarefa e a capacidade atual do sistema: a mudança acontece quando a tarefa exige mais do que o cérebro consegue entregar com os recursos existentes, por tempo suficiente, e cessa quando a demanda é atendida. Três implicações decorrem daí. Primeiro, a tarefa precisa ser difícil o bastante para gerar o descompasso. Segundo, a dificuldade precisa persistir; sessões isoladas produzem mudanças funcionais momentâneas, não reorganização duradoura. Terceiro, a mudança estrutural tende a regredir quando a prática para — o próprio estudo do malabarismo mostrou retorno parcial ao estado inicial três meses após o fim do treino (Draganski et al., 2004).

O que o termo não significa. Neuroplasticidade não é sinônimo de "aprender": todo aprendizado envolve plasticidade sináptica, mas nem todo envolve mudança estrutural detectável. Também não é um recurso que uma ferramenta "ativa" ou "desliga"; é uma propriedade constante do tecido nervoso, que a experiência direciona. E não é unidirecional: a plasticidade que consolida uma habilidade é a mesma que consolida um hábito de evitar esforço. A literatura sobre neuromitos em educação documenta como versões degradadas desses conceitos — "usamos só 10% do cérebro", "estilos de aprendizagem visuais/auditivos" — são aceitas por majorias de professores em vários países (Dekker et al., 2012). "A IA fortalece/enfraquece a neuroplasticidade" é candidata a entrar nessa lista.

Consequência metodológica. Para que um estudo diga algo sobre plasticidade relacionada ao uso de IA, ele precisa de, no mínimo, uma medida neural ou comportamental durável, coleta antes e depois de uma intervenção de duração suficiente, um grupo-controle ativo (que faz a mesma tarefa sem IA) e acompanhamento após o fim da intervenção. A seção 6 mostra que nenhum estudo localizado cumpre os quatro requisitos.

4. Usos educacionais da IA

"IA na educação" designa pelo menos três famílias de tecnologia com histórico, evidência e perfil de risco distintos. Tratá-las como um bloco é a primeira fonte de conclusões inválidas.

Sistemas tutores inteligentes (ITS). São programas que modelam o estado de conhecimento do aluno e ajustam a sequência de problemas e as dicas — sistemas como Cognitive Tutor ou ALEKS, em uso desde os anos 1990. A IA aqui é, em geral, regras e modelos estatísticos, não geração de linguagem. É a família com evidência mais longa: meta-análises cobrindo décadas de ensaios controlados encontram efeitos moderados a grandes sobre desempenho (seção 5), com a ressalva de que o ganho é muito menor em testes padronizados do que em testes locais alinhados ao sistema.

Assistentes generativos de uso geral. Modelos de linguagem (ChatGPT, Claude, Gemini e similares) usados pelo aluno sem configuração específica: tirar dúvidas, resumir, redigir, resolver exercícios. O traço decisivo é que o sistema responde ao que foi pedido — inclusive quando o pedido é "me dê a resposta". Quase toda a literatura pós-2022 e todo o debate público sobre "IA e cérebro" referem-se a esta família.

Assistentes generativos com salvaguardas pedagógicas. O mesmo modelo, configurado por instruções (prompts de sistema) para dar pistas em vez de soluções, pedir que o aluno tente primeiro, ou verificar compreensão. A diferença em relação à família anterior é só de configuração, mas os resultados empíricos divergem o suficiente para justificar a separação (seção 5).

Outros usos. Avaliação automatizada de redações, análise de dados de aprendizagem (*learning analytics*), geração de materiais pelo professor e sistemas neuroadaptativos que leem sinais fisiológicos do aluno em tempo real (seção 6). Cada um levanta questões próprias de privacidade e equidade (seção 7), mas poucos foram avaliados quanto a aprendizagem.

A tabela abaixo resume o que distingue as três famílias principais na dimensão que importa para a tese deste artigo: quanto esforço do aluno a ferramenta preserva por desenho.

Família	Quem controla a sequência	Entrega a resposta pronta?	Evidência de desempenho	Primeira avaliação controlada
ITS clássico	O sistema	Em geral não; dá passos e dicas	Décadas de ensaios; efeitos moderados	Anos 1980–1990
Família	Quem controla a sequência	Entrega a resposta pronta?	Evidência de desempenho	Primeira avaliação controlada
Generativo sem salvaguardas	O aluno	Sim, se pedido	Ganho durante a prática; perda ou neutro em exame sem IA	2023
Generativo com salvaguardas	Compartilhado	Por desenho, não	Ganho durante a prática; neutro em exame sem IA	2023

5. O que as pesquisas medem: desempenho e engajamento

A quase totalidade da evidência disponível está no nível A (resultado de aprendizagem) e no nível B (esforço e comportamento). Três conjuntos de estudos bastam para mostrar o padrão e seus limites.

5.1 Sistemas tutores inteligentes: efeito real, mas sensível à medida

Kulik e Fletcher (2016) meta-analisaram 50 avaliações controladas de ITS e encontraram efeito mediano de 0,66 desvio-padrão sobre a instrução convencional — o equivalente a mover o aluno mediano do percentil 50 para o 75. O achado mais importante, porém, está na heterogeneidade: o efeito foi de cerca de 0,73 em testes desenvolvidos localmente e de apenas 0,13 em testes padronizados. Ma e colegas (2014), com 107 tamanhos de efeito e 14.321 participantes, encontraram $g = 0,42$ contra instrução em grande grupo e nenhuma diferença contra tutoria humana individual ($g = -0,11$). VanLehn (2011), revisando estudos de 1975 a 2010, estimou a tutoria humana em 0,79 e os tutores computacionais passo a passo em valores próximos — bem abaixo dos "dois sigmas" atribuídos a Bloom (1984).

O que isso sustenta: ITS bem implementados melhoram desempenho em conteúdo alinhado ao sistema. O que não sustenta: ganhos amplos e transferíveis, e qualquer inferência sobre o cérebro. Nenhum desses estudos coletou medida neural.

5.2 IA generativa: ganho durante a tarefa, perda sem a ferramenta

O ensaio de campo mais rigoroso até aqui é o de Bastani e colegas (2025), publicado no PNAS. Cerca de mil estudantes de 9º a 11º ano de uma escola na Turquia foram alocados por turma a três condições durante sessões de prática de matemática no semestre de 2023-2024: acesso ao GPT-4 com interface padrão ("GPT Base"), acesso ao mesmo modelo com prompts que o instruíam a dar pistas e não resolver ("GPT Tutor"), ou nenhuma IA. Durante a prática assistida, o desempenho subiu 48% (Base) e 127% (Tutor).

No exame posterior, sem IA, o grupo Base teve desempenho inferior ao controle; o grupo Tutor ficou estatisticamente igual ao controle. Os autores descrevem o resultado como uma troca inerente: a ferramenta eleva o desempenho enquanto está disponível e pode degradar a aprendizagem quando o desenho permite usá-la como atalho. Dois detalhes importam para a tese deste artigo. Primeiro, a diferença entre Base e Tutor é só de configuração — mesmo modelo, prompts diferentes. Segundo, os alunos dos dois grupos com IA superestimaram o quanto haviam aprendido.

Um protocolo de ensaio randomizado registrado em 2025 (Trials, NCT06511102) resume a literatura adjacente no mesmo sentido: estudos randomizados encontraram que alunos desempenham melhor com IA de uso geral e pior quando ela é retirada, e um estudo

posterior encontrou ganho para quem usa a IA para conversar e pedir explicações e prejuízo para quem pede respostas diretas.

5.3 A meta-análise retratada: um caso sobre qualidade de evidência

Em maio de 2025, Wang e Fan publicaram em *Humanities and Social Sciences Communications* uma meta-análise de 51 estudos (nov/2022 a fev/2025) reportando efeito grande do ChatGPT sobre desempenho ($g = 0,867$) e moderado sobre percepção ($g = 0,456$) e pensamento de ordem superior ($g = 0,457$). O artigo foi citado mais de 500 vezes e amplamente usado por empresas de tecnologia educacional. Em junho de 2025, Ingebrigtsen e Lukic (Universidade Ártica da Noruega) publicaram um pré-print apontando que a própria *forest plot* do estudo implicava efeito próximo de 0,5, que a análise incluía um estudo já retratado, estudos mal classificados e amostras pequenas com efeitos implausíveis, e que muitos estudos primários não controlavam variáveis de confusão. O artigo foi retratado pela revista em 22/04/2026; a nota de retratação cita discrepâncias na meta-análise que minaram a confiança do editor na validade da análise e das conclusões. O caso é relevante aqui por dois motivos. O número que mais circulou como "prova" de que IA melhora a aprendizagem não resistiu a verificação. E mesmo que o efeito corrigido ($\sim 0,5$) fosse válido, continuaria sendo uma medida de nível A: nada nela fala de cérebro.

5.4 Descarregamento cognitivo: correlação, não causa

Gerlich (2025) aplicou questionário a 666 adultos e encontrou correlação forte entre uso frequente de IA e descarregamento cognitivo ($r = 0,72$) e correlação negativa entre descarregamento e pensamento crítico ($r = -0,75$), com participantes de 17 a 25 anos apresentando maior dependência e menores escores. O estudo é transversal, usa autorrelato para as três variáveis e, como o próprio autor reconhece, descreve correlação e não causa. Serve como hipótese, não como evidência de erosão.

5.5 Síntese do nível A/B

Estudo	Desenho	n	Medida	Resultado	Fala de cérebro?
Kulik & Fletcher (2016)	Meta-análise, 50 ensaios de ITS	—	Testes locais e padronizados	0,66 mediano; 0,13 em padronizados	Não
Ma et al. (2014)	Meta-análise, 107 efeitos	14.321	Desempenho	$g = 0,42$ vs. grande grupo; $\approx$ tutoria humana	Não
Bastani et al. (2025)	Ensaio randomizado por turma	≈ 1.000	Prática com IA e exame sem IA	+48%/+127% na prática; Base pior no exame, Tutor igual	Não

Wang & Fan (2025) — retratado	Meta-análise, 51 estudos	—	Desempenho, percepção	$g = 0,867$ reportado; $\sim 0,5$ recalculado; retratado 2026	Não
Gerlich (2025)	Transversal, questionário	666	Autorrelato	$r = -0,75$ entre descarregamento e pensamento crítico	Não

6. O que as pesquisas medem: cérebro

A busca localizou um número muito pequeno de estudos que coletaram medida neural durante o uso de IA em tarefa de aprendizagem. Nenhum deles mede neuroplasticidade no sentido da seção 3; todos medem atividade momentânea.

6.1 "Your Brain on ChatGPT" (Kosmyna et al., 2025)

O estudo mais citado é um pré-print do MIT Media Lab (arXiv:2506.08872, junho de 2025, versão 2 em dezembro de 2025, não revisado por pares). Cinquenta e quatro participantes escreveram ensaios em três sessões, divididos em grupo com LLM, grupo com buscador e grupo "só cérebro", com EEG de 32 canais; numa quarta sessão, parte dos participantes trocou de condição. Os resultados: conectividade EEG mais ampla e forte no grupo só cérebro, intermediária no grupo buscador e mais fraca no grupo LLM; 83% do grupo LLM não conseguiram citar uma frase do próprio ensaio logo após escrevê-lo, contra cerca de 11% nos outros grupos; e, na quarta sessão, quem passou do LLM para escrever sozinho mostrou menor engajamento neural, enquanto quem passou de escrever sozinho para o LLM manteve conectividade mais ampla. Os autores cunharam o termo "dívida cognitiva" para o primeiro padrão.

O que o estudo mostra: diferenças de conectividade durante a tarefa, consistentes com menor esforço quando a ferramenta faz parte do trabalho. O que não mostra: mudança estrutural, mudança duradoura, dano, ou "atrofia". A própria página do projeto pede que não se use linguagem como "brain rot" ou "dano cerebral", lembra que EEG mede engajamento durante uma atividade e que o estudo não mediu QI, e reconhece que os achados são dependentes de contexto e podem não generalizar. Um comentário publicado em janeiro de 2026 (Stanković et al., arXiv:2601.00856) aponta amostra pequena, problemas na análise de EEG, inconsistências no relato e baixa transparência, e recomenda interpretação mais conservadora.

6.2 NeuroChat (Baradari, Kosmyrna et al., 2025)

Do mesmo laboratório, um estudo intrassujeitos com 24 participantes testou um tutor generativo que ajusta suas respostas em tempo real ao engajamento medido por EEG. O sistema aumentou o engajamento medido por EEG e o autorrelatado em comparação com um chatbot não adaptativo, mas não houve diferença nos resultados de aprendizagem de curto prazo. É a demonstração mais clara, dentro da própria literatura de IA, de que sinal neural de engajamento e aprendizagem são variáveis diferentes.

6.3 Estudos com fNIRS e EEG em 2026

Um estudo publicado em *Computers & Education* (2026) coletou fNIRS do córtex pré-frontal e expressões faciais durante tarefas de pensamento criativo com e sem IA generativa, reportando melhor desempenho criativo com IA e, ao mesmo tempo, mais emoções negativas; os autores usam os sinais como indicadores classificatórios de desempenho, não como medida de plasticidade. Um estudo exploratório com estudantes de ensino médio (arXiv:2606.26579, pré-print) comparou modos de interação com IA (tutor, colaborador, sem IA) com EEG frontal e não encontrou diferença estatisticamente significativa na potência espectral ($p > 0,29$); os autores avisam que isso não prova ausência de efeito, apenas que o desenho não detectou mudanças grandes. Um ensaio randomizado com fNIRS em redação analítica, registrado em 2025, ainda não tem resultados publicados até a data desta revisão.

6.4 O que o nível C permite concluir

Estudo	Medida	n	Duração	Pré/pós?	Controle ativo?	Seguimento?	Revisado por pares?
Kosmyrna et al. (2025)	EEG, conectividade	54	4 sessões	Não	Sim	Não	Não (pré-print)
NeuroChat (2025)	EEG, engajamento	24	1 sessão	Não	Sim	Não	Sim (ACM CUI)
fNIRS criatividade (2026)	fNIRS, pré-frontal	n.i.	Sessões únicas	Não	Sim	Não	Sim
EEG ensino médio (2026)	EEG frontal	n.i.	Sessões únicas	Não	Sim	Não	Não (pré-print)

Nenhum estudo cumpre os quatro requisitos da seção 3 (medida durável, pré/pós, controle ativo, seguimento). Todos registram o que o cérebro faz enquanto a ferramenta está em uso — informação útil sobre esforço, não sobre plasticidade. A frase correta, portanto, não é "a IA reduz a neuroplasticidade", e sim: "em tarefas de escrita, delegar a produção do texto a um LLM está associado a menor conectividade EEG e pior recordação imediata do próprio texto". É uma afirmação sobre esforço e memória de trabalho na sessão, compatível com a

hipótese de que uso repetido nesse modo produziria menos consolidação — hipótese que ninguém testou.

7. Discussão: esforço cognitivo, educador, privacidade e acesso

7.1 Esforço cognitivo é a variável que une os três níveis

O modelo de descompasso oferta-demanda (seção 3) e os achados de Bastani et al. (2025) e Kosmyna et al. (2025) convergem num ponto: a aprendizagem — e, por extensão, qualquer plasticidade duradoura — depende de o aluno realizar o trabalho de recuperar, elaborar e produzir. A psicologia cognitiva chama isso de "dificuldades desejáveis" (Bjork & Bjork, 2011): condições que pioram o desempenho durante a prática e melhoram a retenção. A IA generativa sem salvaguardas faz o oposto por desenho: melhora o desempenho durante a prática e piora a retenção. O resultado de Bastani et al. é literalmente esse padrão, medido.

Isso reorganiza a pergunta. Em vez de "a IA faz bem ou mal ao cérebro", a pergunta operacional é "esta atividade com IA preserva ou elimina o esforço que produz consolidação?". A mesma ferramenta cai dos dois lados: pedir ao modelo que explique um conceito e depois reescrevê-lo com as próprias palavras preserva esforço; pedir a resposta elimina. O descarregamento cognitivo em si não é patológico — calculadoras, anotações e buscadores são descarregamento (Risko & Gilbert, 2016). O problema é descarregar exatamente a operação que a tarefa existia para treinar.

7.2 O educador define o tipo de atividade

A diferença entre GPT Base e GPT Tutor em Bastani et al. foi um prompt de sistema com cerca de 500 palavras escrito por professores, contra cerca de 50 palavras no modo padrão. Essa diferença bastou para transformar perda de aprendizagem em neutralidade. Nenhuma propriedade do modelo mudou; mudou o desenho da atividade. Isso sugere que o efeito da IA sobre a aprendizagem é, em boa medida, uma variável pedagógica sob controle do educador — e que delegar esse desenho à interface padrão do produto equivale a escolher a condição que mais prejudica.

Há um segundo papel, menos discutido. Nos dois estudos, os alunos que usaram IA superestimaram o próprio aprendizado. A IA produz uma sensação de competência que não corresponde ao desempenho sem ela. Calibrar essa percepção — mostrar ao aluno a diferença entre o que ele fez com a ferramenta e o que faz sem — é uma função que só o educador, ou uma avaliação desenhada para isso, pode cumprir.

7.3 Privacidade: o custo dos dados que medem o cérebro

A agenda de pesquisa necessária (seção 8) exige exatamente os dados mais sensíveis: EEG, fNIRS, registros detalhados de interação e, idealmente, imagem estrutural de menores de idade ao longo de meses. Sistemas neuroadaptativos como o NeuroChat levam essa coleta para o uso cotidiano, não só para o laboratório. A orientação da UNESCO sobre IA generativa em educação (Miao & Holmes, 2023) recomenda idade mínima, proteção de dados e supervisão humana, mas não trata de dados neurais. O campo precisará de regras específicas para "neurodados" educacionais — quem os guarda, por quanto tempo, com que finalidade e com que direito de exclusão — antes que a infraestrutura de coleta se normalize.

Há também um risco de uso indevido do próprio discurso da neuroplasticidade: produtos que prometem "otimizar o cérebro do aluno" com base em sinais de engajamento, quando a evidência (NeuroChat) mostra que engajamento neural subiu sem que a aprendizagem subisse.

7.4 Acesso e equidade

Bastani et al. encontraram que o acesso à IA reduziu a dispersão de notas — ajudou mais os alunos mais fracos — durante a prática. Se esse ganho é real na prática e se dissolve no exame, a IA sem salvaguardas pode criar uma falsa redução de desigualdade: os alunos que mais precisavam de prática efetiva são os que mais a perdem. Por outro lado, a versão com salvaguardas exige trabalho de configuração e supervisão que escolas com menos recursos têm menos condições de fazer. O risco não é o acesso desigual à IA; é o acesso desigual à IA bem configurada.

Os estudos neurais analisados incluem adultos universitários e adolescentes. Khurana e Liew (2026), por exemplo, investigaram 48 participantes de 14 a 18 anos. A literatura de nível A mais robusta vem de uma escola na Turquia. Não há dado de nível C em crianças, em contexto brasileiro ou em uso prolongado.

8. Lacunas e agenda de pesquisa

As lacunas abaixo são as que separam a evidência atual de uma afirmação legítima sobre IA e neuroplasticidade. Estão ordenadas da mais básica à mais exigente.

- 1. Duração.** Nenhum estudo de nível C passa de quatro sessões. A plasticidade estrutural requer semanas a meses de descompasso sustentado. Ensaios de um semestre com medida neural no início e no fim são o mínimo.

2. Controle ativo e pareamento de tarefa. O grupo-controle precisa fazer a mesma tarefa, com o mesmo tempo, sem a ferramenta. Comparar "usou IA" com "não fez nada" mede o efeito de praticar, não o da IA.

3. Transferência e retenção tardia. A medida de aprendizagem deve ser tomada sem a ferramenta, semanas depois, em problemas estruturalmente diferentes dos praticados. Bastani et al. mediram sem a ferramenta; ninguém mediu transferência tardia com medida neural.

4. Separar modo de uso da ferramenta. Base e Tutor são o mesmo modelo. Estudos devem registrar e classificar o tipo de interação (pedir resposta, pedir explicação, pedir verificação) como variável independente, não tratar "usou ChatGPT" como condição única.

5. Medida neural com hipótese prévia. EEG e fNIRS devem testar hipóteses pré-registradas (por exemplo, mudança em marcadores de consolidação ou de controle executivo), não varrer bandas espectrais em busca de diferenças.

6. Imagem estrutural. Nenhum estudo usou ressonância magnética estrutural ou de difusão. É a única medida que fala diretamente de plasticidade estrutural, e tem desenho conhecido (Draganski et al., 2004) para ser adaptado.

7. Populações. Crianças, adolescentes mais novos, contextos de baixa renda e países do hemisfério sul estão ausentes do nível C.

8. Replicação independente. Os dois estudos mais citados de nível C vêm do mesmo laboratório e um deles tem crítica metodológica publicada sem resposta. O campo precisa de replicação por grupos sem vínculo com os autores originais nem com fornecedores de IA.

9. Verificação da literatura secundária. O caso Wang e Fan mostra que meta-análises precoces podem incluir estudos retratados e inflar efeitos. Revisões futuras devem verificar status de retratação de cada estudo incluído e relatar efeitos com e sem os estudos de maior risco.

Uma pergunta de pesquisa que cumpriria essas condições: *Em estudantes de 14 a 17 anos, um semestre de prática de matemática com um assistente generativo configurado para dar pistas, comparado à mesma prática sem assistente, produz diferença em (a) desempenho sem a ferramenta oito semanas após o fim, (b) transferência para problemas novos e (c) marcadores estruturais ou funcionais pré-registrados?* Nenhum estudo localizado chega perto desse desenho.

9. Conclusão

A pergunta central deste artigo tem resposta curta. As pesquisas demonstram que a IA na educação altera o desempenho de maneira que depende do desenho da atividade: sobe quando a ferramenta está disponível e cai, fica neutra ou se mantém no exame sem ela conforme o quanto a atividade preservou o esforço do aluno. As pesquisas que mediram algo no cérebro são quatro ou cinco, pequenas, de sessão única ou curta, e registram engajamento momentâneo — não plasticidade. A evidência mais citada a favor da IA foi retratada; a evidência mais citada contra ela é um pré-print cujos autores pedem que não se fale em dano.

A tese se sustenta na forma proposta. A IA pode apoiar aprendizagem personalizada; os sistemas tutores inteligentes têm décadas de evidência moderada e os assistentes generativos com salvaguardas evitam a perda vista sem elas. Mas não há base para afirmar que usar IA melhora ou prejudica a neuroplasticidade. O que a literatura permite dizer é mais útil que isso: a aprendizagem — e qualquer mudança cerebral duradoura que ela produza — exige que o aluno faça o trabalho de recuperar e elaborar, e a IA pode tanto preservar quanto eliminar esse trabalho, por decisão de quem desenha a atividade. A neuroplasticidade não é o que está em jogo; o esforço é. E esforço, ao contrário de plasticidade, é algo que educadores podem observar e desenhar hoje, sem esperar pelos estudos que ainda não existem.

Referências

Baradari, D., Kosmyna, N., Petrov, O., Kaplun, R., & Maes, P. (2025). NeuroChat: A neuroadaptive AI chatbot for customizing learning experiences. *Proceedings of the 7th ACM Conference on Conversational User Interfaces (CUI '25)*.

<https://doi.org/10.1145/3719160.3736623>

Bastani, H., Bastani, O., Sungu, A., Ge, H., Kabakçı, Ö., & Mariman, R. (2025). Generative AI without guardrails can harm learning: Evidence from high school mathematics.

Proceedings of the National Academy of Sciences, 122(26).

<https://doi.org/10.1073/pnas.2422633122>

Bjork, E. L., & Bjork, R. A. (2011). Making things hard on yourself, but in a good way: Creating desirable difficulties to enhance learning. In M. A. Gernsbacher et al. (Eds.), *Psychology and the real world* (pp. 56–64). Worth.

Bloom, B. S. (1984). The 2 sigma problem: The search for methods of group instruction as effective as one-to-one tutoring. *Educational Researcher*, 13(6), 4–16.

- Bruer, J. T. (1997). Education and the brain: A bridge too far. *Educational Researcher*, 26(8), 4–16.
- Dekker, S., Lee, N. C., Howard-Jones, P., & Jolles, J. (2012). Neuromyths in education: Prevalence and predictors of misconceptions among teachers. *Frontiers in Psychology*, 3, 429. <https://doi.org/10.3389/fpsyg.2012.00429>
- Draganski, B., Gaser, C., Busch, V., Schuierer, G., Bogdahn, U., & May, A. (2004). Changes in grey matter induced by training. *Nature*, 427, 311–312. <https://doi.org/10.1038/427311a>
- Gerlich, M. (2025). AI tools in society: Impacts on cognitive offloading and the future of critical thinking. *Societies*, 15(1), 6. <https://doi.org/10.3390/soc15010006>
- Ingebrigtsen, M., & Lukic, M. (2025). Substantial errors invalidate the conclusion from Wang and Fan (2025) about ChatGPT and learning outcomes. *PsyArXiv* (pré-print). https://osf.io/preprints/psyarxiv/wgu56_v2
- Kosmyna, N., Hauptmann, E., Yuan, Y. T., Situ, J., Liao, X.-H., Beresnitzky, A. V., Braunstein, I., & Maes, P. (2025). Your brain on ChatGPT: Accumulation of cognitive debt when using an AI assistant for essay writing task. *arXiv* 2506.08872 (pré-print, v2 dez. 2025). <https://arxiv.org/abs/2506.08872>
- Kulik, J. A., & Fletcher, J. D. (2016). Effectiveness of intelligent tutoring systems: A metaanalytic review. *Review of Educational Research*, 86(1), 42–78. <https://doi.org/10.3102/0034654315581420>
- Lövdén, M., Bäckman, L., Lindenberger, U., Schaefer, S., & Schmiedek, F. (2010). A theoretical framework for the study of adult cognitive plasticity. *Psychological Bulletin*, 136(4), 659–676.
- Ma, W., Adesope, O. O., Nesbit, J. C., & Liu, Q. (2014). Intelligent tutoring systems and learning outcomes: A meta-analysis. *Journal of Educational Psychology*, 106(4), 901–918.
- Maguire, E. A., Gadian, D. G., Johnsrude, I. S., Good, C. D., Ashburner, J., Frackowiak, R. S. J., & Frith, C. D. (2000). Navigation-related structural change in the hippocampi of taxi drivers. *PNAS*, 97(8), 4398–4403. <https://doi.org/10.1073/pnas.070039597>
- Miao, F., & Holmes, W. (2023). *Guidance for generative AI in education and research*. UNESCO.
- Retraction Note (2026). The effect of ChatGPT on students' learning performance, learning perception, and higher-order thinking: insights from a meta-analysis. *Humanities and Social Sciences Communications*. <https://www.nature.com/articles/s41599-026-07310z>
- Risko, E. F., & Gilbert, S. J. (2016). Cognitive offloading. *Trends in Cognitive Sciences*, 20(9), 676–688.

Stanković, M., Hirche, E., Kollatzsch, S., & Doetsch, J. N. (2026). Comment on: Your brain on ChatGPT. *arXiv* 2601.00856 (pré-print). <https://arxiv.org/pdf/2601.00856>

VanLehn, K. (2011). The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems. *Educational Psychologist*, 46(4), 197–221.

Wang, J., & Fan, W. (2025). The effect of ChatGPT on students' learning performance, learning perception, and higher-order thinking: insights from a meta-analysis. *Humanities and Social Sciences Communications*, 12. Retratado em 22/04/2026.

Song, X., Zhang, Y., Lu, Z., Xu, L., & Shen, H. (2026). Generative AI: A double-edged sword for creative thinking learning — Evidence from facial expressions and fNIRS. *Computers & Education*, 247, 105578. <https://doi.org/10.1016/j.compedu.2026.105578>

Khurana, K., & Liew, A. (2026). An exploratory behavioral and electroencephalographic study of artificial intelligence-assisted learning modes in high school students. *arXiv* 2606.26579 (pré-print). <https://arxiv.org/abs/2606.26579>

Chen, Y., Wang, Y., Wüstenberg, T., Kizilcec, R. F., Fan, Y., Li, Y., Lu, B., Yuan, M., Zhang, J., Zhang, Z., Geldsetzer, P., Chen, S., & Bärnighausen, T. (2025). Effects of generative artificial intelligence on cognitive effort and task performance: study protocol for a randomized controlled experiment among college students. *Trials*, 26, 244. <https://doi.org/10.1186/s13063-025-08950-3>

Fontes secundárias consultadas para o caso da retratação: Plagiarism Today, 15 jul. 2026; página oficial do projeto Your Brain on ChatGPT (MIT Media Lab); Knowledge at Wharton sobre Bastani et al.