AVALIAÇÃO NEUROCOGNITIVA NA INTERAÇÃO HUMANO-IA: UM ESTUDO EXPERIMENTAL SOBRE O BALANÇO ENTRE CARGA COMPUTACIONAL E COGNITIVA

Hélio Craveiro Pessoa Júnior[1].

Universidade de Brasília (UnB), Brasília, Distrito Federal.

http://lattes.cnpq.br/1415349950533370

RESUMO: Este estudo propõe um design experimental inovador para investigar a interseção crítica entre a eficiência computacional de assistentes de Inteligência Artificial (IA) e a carga cognitiva (CL) imposta ao usuário. O problema central reside no balanço entre a fidelidade de Grandes Modelos de Raciocínio (LRMs) e sua acessibilidade, onde modelos de alta performance são computacionalmente caros e modelos comprimidos (quantizados) podem degradar a experiência do usuário. O objetivo é avaliar empiricamente como a fidelidade do modelo (Q4 vs. Q8) e a complexidade da interação (agente único vs. multiagente) afetam a CL e o desempenho de aprendizes (N=40) com diferentes níveis de expertise. A metodologia de métodos mistos emprega uma triangulação de dados subjetivos (NASA-TLX, SEQ), psicofisiológicos (EEG, POG, PPG, SpO2) e de desempenho, coletados através de uma plataforma customizada. Espera-se encontrar uma relação inversa onde modelos de baixa fidelidade (Q4) aumentam a CL, e que a expertise modere este efeito. A análise visa identificar um "ponto ideal" cognitivo-computacional, fornecendo diretrizes para o design de sistemas de IA centrados no humano que otimizem a aprendizagem ao invés de apenas a eficiência técnica. A plataforma experimental utilizada está disponível em http://crossdebate.com, permitindo replicação e extensão futura do estudo.

PALAVRAS-CHAVE: Fluxos de trabalho de análise de Dados. LRMs agenticamente relacionados. Computação neurofisiológica.

NEUROCOGNITIVE ASSESSMENT IN HUMAN-AI INTERACTION: AN EXPERIMENTAL STUDY ON THE BALANCE BETWEEN COMPUTATIONAL AND COGNITIVE LOAD

ABSTRACT: This study proposes an innovative experimental design to investigate the critical intersection between the computational efficiency of Artificial Intelligence (AI) assistants and the cognitive load (CL) imposed on the user. The core problem lies in the trade-off between the fidelity of Large Reasoning Models (LRMs) and their accessibility, where high-performance

models are computationally expensive, and compressed models (quantized) may degrade the user experience. The objective is to empirically assess how model fidelity (Q4 vs. Q8) and interaction complexity (single-agent vs. multi-agent) affect the CL and performance of learners (N=40) with varying levels of expertise. The mixed-methods methodology employs a triangulation of subjective (NASA-TLX, SEQ), psychophysiological (EEG, POG, PPG, SpO2), and performance data collected via a custom platform. An inverse relationship is expected, where low-fidelity models (Q4) increase CL, with expertise moderating this effect. The analysis aims to identify an optimal cognitive-computational "sweet spot," providing guidelines for designing human-centered AI systems that optimize learning rather than just technical efficiency. The experimental platform used is available at http://crossdebate.com, enabling replication and future extension of the study.

KEYWORDS: Data analysis workflows. Agentically-related LRMs. Neurophysiological computing.

INTRODUÇÃO

A incorporação da Inteligência Artificial (IA) generativa na educação e no trabalho do conhecimento inaugura avanços inéditos, mas também paradoxos complexos. Ferramentas baseadas em Grandes Modelos de Raciocínio (LRMs), como Gemini 2.5 Pro, Claude 4 Opus e MiniMax M1, atuam como assistentes de aprendizagem, prometendo ganhos substanciais de produtividade e qualidade. Pesquisas recentes mostram que a colaboração humano-IA pode resultar em produtos mais eficientes e, por vezes, superiores em relação ao trabalho exclusivamente humano. No entanto, essa eficiência esconde custos psicológicos relevantes e frequentemente negligenciados. Um estudo de Liu et al. (2025) revelou que, embora a IA melhore o desempenho imediato, ela reduz a motivação intrínseca e aumenta o tédio em tarefas subsequentes sem assistência, devido à diminuição do "senso de controle" sobre o próprio trabalho.

Esse paradoxo entre eficiência e custo motivacional é o cerne do chamado "paradoxo da eficácia pedagógica computacional". Modelos de IA mais potentes exigem mais recursos computacionais, restringindo seu uso a instituições privilegiadas e ampliando desigualdades de acesso. Uma solução tecnológica comum – a quantização de modelos para maior eficiência (Q4 vs. Q8) – introduz o "custo cognitivo da eficiência computacional": ao reduzir a precisão, arrisca-se degradar a qualidade das respostas, impactando negativamente a cognição do aprendiz.

A análise desse fenômeno vai além da Teoria da Carga Cognitiva (CLT) de Sweller (1988), que distingue Carga Intrínseca (ICL), Germânica (GCL) e Extrínseca (ECL). A hipótese central é que a fidelidade do assistente de IA modula diretamente a ECL: alta fidelidade (Q8) reduz a carga improdutiva, enquanto baixa fidelidade (Q4) aumenta a necessidade de verificação e depuração, desviando o foco da aprendizagem genuína.

Além do aumento da carga, surge uma mudança fundamental no engajamento cognitivo: o chamado "cognitive offloading". Estudo do MIT Media Lab (Chow, 2025) mostrou, via EEG, que usuários do ChatGPT apresentaram menor engajamento cerebral e maior tendência ao copiar e colar, indicando afastamento do pensamento crítico e da elaboração profunda. Assim, assistentes de baixa fidelidade não só impõem ECL elevada, mas também incentivam o descarregamento cognitivo, prejudicando o investimento na GCL e minando o senso de agência e competência do aprendiz.

Diante desse cenário, métricas de desempenho isoladas são manifestamente insuficientes. É necessária uma abordagem multimodal, combinando medidas subjetivas (questionários), de desempenho e psicofisiológicas (EEG, pupilometria, PPG). A plataforma experimental CrossDebate integra esses dados de forma sincronizada, permitindo uma análise mecanicista e detalhada da interação humano-IA, indo além das métricas tradicionais de resultado e possibilitando uma compreensão mais profunda dos mecanismos subjacentes ao uso da IA na aprendizagem.

OBJETIVO

O objetivo geral deste estudo é investigar os mecanismos neurocognitivos e afetivos que governam a interação entre aprendizes e assistentes de IA. Propõe-se modelar como a fidelidade do modelo (Q4 vs. Q8) e a complexidade da arquitetura de interação (agente único vs. multiagente) modulam a carga cognitiva, o engajamento motivacional e o desempenho da aprendizagem em indivíduos com diferentes níveis de expertise.

Os objetivos específicos são: *1.* Quantificar o impacto da fidelidade do assistente de IA na carga cognitiva extrínseca (ECL) e germânica (GCL) através de uma triangulação de medidas subjetivas (SEQ, NASA-TLX), psicofisiológicas (EEG, POG, HRV) e de desempenho (qualidade da tarefa, tempo de conclusão); *2.* Analisar como a colaboração com IA de baixa fidelidade afeta a motivação intrínseca e o senso de controle do aprendiz, buscando replicar e estender os achados de Liu et al. (2025) para o contexto educacional; *3.* Identificar os correlatos neurais do "cognitive offloading" induzido pela IA, especificamente testando a hipótese de uma supressão da atividade nas bandas teta e alfa durante a interação com modelos de baixa fidelidade, conforme sugerido pelo estudo do MIT reportado por Chow (2025); *4.* Isolar e medir a "carga cognitiva de gerenciamento" como uma nova faceta da ECL, imposta por arquiteturas de interação multiagente que exigem a coordenação e integração de múltiplas fontes de informação; e, *5.* Mapear a função de custo cognitivo-computacional, uma relação não linear entre a eficiência computacional do sistema e o custo cognitivo-afetivo imposto ao usuário, para identificar um "ponto ideal" que otimiza a ergonomia cognitiva da interação.

FUNDAMENTAÇÃO TEÓRICA

Este estudo experimental mergulha no paradoxo central da interação humano-IA na educação e no trabalho do conhecimento: o ganho aparente de produtividade frequentemente mascara custos cognitivos e motivacionais significativos. Pesquisas seminais, como a de Liu et al. (2025), já demonstraram que a colaboração com IA generativa, embora melhore o desempenho imediato, pode corroer a motivação intrínseca e o senso de controle do indivíduo. De forma ainda mais alarmante, o trabalho reportado por Chow (2025) sobre um estudo do MIT fornece evidências neurofisiológicas de que o uso de IA pode induzir um "descarregamento cognitivo" (cognitive offloading), resultando em menor engajamento cerebral e na supressão do pensamento crítico. O objetivo desta pesquisa é, portanto, dissecar o balanço entre a eficiência computacional de modelos de IA e a carga cognitiva imposta ao aprendiz, visando um design que otimize a aprendizagem genuína.

O alicerce teórico para esta investigação é a Teoria da Carga Cognitiva (CLT), proposta por Sweller (1988). A CLT postula que nossa memória de trabalho é um recurso limitado e que a aprendizagem é otimizada quando a carga imposta a ela é gerenciada eficazmente. A teoria distingue três fontes de carga: a Carga Intrínseca (ICL), inerente à complexidade do conteúdo; a Carga Extrínseca (ECL), uma carga improdutiva gerada por um design instrucional deficiente; e a Carga Germânica (GCL), o esforço mental produtivo dedicado à construção de esquemas mentais e à aprendizagem profunda.

Aprofundando essa teoria, Sweller e Chandler (1994) argumentam que a ECL é frequentemente o resultado direto de um design instrucional ou de interface inadequado, que força o aprendiz a despender esforço em atividades irrelevantes para o objetivo da aprendizagem. No contexto deste estudo, a fidelidade do assistente de IA é posicionada como uma variável instrucional crítica. A hipótese central é que um assistente de baixa fidelidade (quantizado em Q4), ao gerar respostas imprecisas ou ambíguas, atua como uma fonte massiva de ECL, obrigando o usuário a um ciclo constante de verificação, depuração e busca por confirmação externa.

A inovação central do estudo é, portanto, tratar a fidelidade da IA não como um detalhe técnico, mas como um modulador direto da carga cognitiva. Um assistente de alta fidelidade (Q8) deve, teoricamente, minimizar a ECL, fornecendo um andaime cognitivo claro e confiável que libera recursos da memória de trabalho para serem investidos na GCL. Em contraste, o assistente de baixa fidelidade (Q4) não apenas aumenta a ECL, mas também desvia o foco da tarefa principal, impedindo o engajamento nos processos de compreensão profunda que caracterizam a GCL e são essenciais para a construção de conhecimento duradouro e para o desenvolvimento de um senso de competência e agência.

Contudo, o impacto de um assistente de baixa fidelidade transcende a simples sobrecarga. A frustração e a desconfiança geradas por uma ferramenta errática podem fomentar comportamentos mal-adaptativos, como o "cognitive offloading" reportado por

Chow (2025). Neste cenário, o aprendiz, em vez de lutar contra a ECL, opta por aceitar passivamente as saídas da IA sem o devido escrutínio, abdicando do pensamento crítico. Este comportamento, embora reduza o esforço imediato, sabota a aprendizagem e explica a queda na motivação intrínseca e no senso de controle documentada por Liu et al. (2025), pois o indivíduo se torna um mero executor em vez de um agente cognitivo ativo.

Para capturar empiricamente um construto tão complexo, uma única medida é insuficiente. Por isso, o estudo adota uma abordagem de triangulação, conforme defendido por Paas e Merriënboer (1994), que combina sistematicamente três tipos de dados: medidas subjetivas de carga de trabalho percebida (NASA-TLX), medidas de desempenho (qualidade da tarefa) e, crucialmente, medidas psicofisiológicas objetivas e contínuas. Esta abordagem multimodal permite uma avaliação robusta e nuançada da experiência do usuário, onde os dados fisiológicos podem validar ou contextualizar os relatos subjetivos.

A abordagem psicofisiológica é particularmente inovadora, utilizando uma gama de sensores para obter uma janela em tempo real para os processos cognitivos. A eletroencefalografia (EEG) medirá a potência nas bandas teta e alfa, correlatos neurais robustos da carga da memória de trabalho e do engajamento atencional, conforme estabelecido por Gevins e Smith (2000) e Klimesch (1999). A pupilometria (POG) avaliará a dilatação pupilar, um indicador sensível do esforço cognitivo (Beatty e Lucero-Wagoner, 2000). Adicionalmente, a variabilidade da frequência cardíaca (HRV), que reflete o estado do sistema nervoso autônomo, será monitorada, pois uma HRV reduzida está associada a maior estresse e carga mental (Hjortskov et al., 2004). Todos esses dados são coletados na plataforma CrossDebate, cuja arquitetura (frontend em React 18.3.1 e backend em FastAPI 0.115.6) é otimizada para minimizar a ECL do próprio sistema.

A síntese destes resultados visa mapear uma relação não linear, em forma de U, entre a eficiência computacional e o “Custo Cognitivo-Afetivo” imposto ao usuário. Esta relação é conceitualmente análoga à Lei de Yerkes-Dodson (1908), que descreve um desempenho ótimo em níveis intermediários de excitação. O objetivo final é identificar um “ponto ideal cognitivo-computacional” que otimize a ergonomia da interação. Ao compreender esses mecanismos, o estudo busca fornecer diretrizes para o desenvolvimento de sistemas de IA genuinamente centrados no humano, que equilibrem a capacidade da máquina com a necessidade de manter o aprendiz engajado, motivado e em um estado de esforço cognitivo produtivo e sustentável.

METODOLOGIA

Adota-se um desenho experimental de métodos mistos convergente, que permite a triangulação sistemática de dados quantitativos (fisiológicos, comportamentais, desempenho) e qualitativos (protocolos de pensamento em voz alta). O núcleo do estudo é um desenho fatorial misto 2x3x2, que examina os efeitos de três variáveis independentes:

Fidelidade do Assistente de IA (Fator Intra-sujeitos): *Nível 1: Baixa Fidelidade (Q4):* Interação com assistentes de IA baseados em modelos LRM com quantização de 4 bits (formato GGUF Q4), caracterizados por maior eficiência computacional, mas com maior propensão a gerar respostas semanticamente imprecisas ou factualmente incorretas; e, *Nível 2: Alta Fidelidade (Q8):* Interação com assistentes baseados em modelos com quantização de 8 bits (formato GGUF Q8), que oferecem maior confiabilidade e coerência nas respostas, ao custo de maior demanda computacional.

Complexidade da Interação (Fator Intra-sujeitos): *Nível 1: Agente Único:* Interação com um único assistente de IA generalista; *Nível 2: Multiagente Sequencial:* A tarefa é decomposta em uma sequência linear de subtarefas, cada uma assistida por um agente de IA especializado; e, *Nível 3: Multiagente Paralelo:* A tarefa envolve a coordenação e integração de informações de múltiplos agentes de IA operando concorrentemente.

Expertise do Aprendiz (Fator Entre-sujeitos): *Nível 1: Iniciantes:* Participantes com pouca ou nenhuma experiência prévia no domínio da tarefa; e, *Nível 2: Especialistas:* Participantes com conhecimento e experiência substanciais no domínio.

O uso de fatores intra-sujeitos para Fidelidade e Complexidade aumenta significativamente o poder estatístico do estudo, pois cada participante atua como seu próprio controle, reduzindo a variância de erro associada a diferenças individuais. Este desenho é particularmente poderoso para elucidar as interações de segunda e terceira ordem (ex: Fidelidade x Expertise), que são teoricamente mais informativas do que os efeitos principais isolados.

Será recrutada uma amostra de N=40 estudantes universitários, estratificada para garantir um número igual de Iniciantes (n=20) e Especialistas (n=20). A expertise será determinada por meio de um questionário de triagem e um pré-teste de conhecimento específico do domínio. Os critérios de inclusão exigem visão normal ou corrigida para contato (essencial para a pupilometria) e ausência de distúrbios neurológicos ou psiquiátricos conhecidos.

O estudo será conduzido em um laboratório com ambiente controlado na Universidade do Distrito Federal, para minimizar artefatos nos dados fisiológicos, como ruído elétrico (para EEG) e variações de luminosidade (para POG).

Cada **sessão experimental**, com duração aproximada de 60 minutos, seguirá um procedimento padronizado: *Preparação (10 min):* Assinatura do termo de consentimento livre e esclarecido, preenchimento do questionário de triagem e demográfico. Colocação e calibração dos sensores fisiológicos. Uma linha de base fisiológica será gravada em estado

de repouso, permitindo a análise de mudanças relativas (normalização por z-score) e controlando a variabilidade interindividual; *Familiarização (5 min):* Um tutorial interativo sobre a interface da plataforma CrossDebate e a natureza das tarefas. *Blocos Experimentais (40 min):* O participante completará 6 blocos de tarefas (resultantes do cruzamento 2 Fidelidade x 3 Complexidade), apresentados em ordem contrabalançada para mitigar efeitos de ordem e fadiga. Durante cada bloco, o participante realizará uma tarefa de aprendizagem complexa enquanto verbaliza seus pensamentos (protocolo de pensamento em voz alta). *Avaliações Pós-Bloco (5 min):* Após cada bloco, o participante preencherá os questionários NASA-TLX e Single Ease Question (SEQ); e, *Debriefing:* Ao final da sessão, os objetivos do estudo serão explicados ao participante e suas dúvidas, esclarecidas.

Instrumentos e Coleta de Dados: *Plataforma CrossDebate:* Um ambiente experimental web customizado, com frontend em React e backend em FastAPI. A plataforma foi projetada para apresentar os estímulos, registrar todas as interações do usuário (logs de cliques, edições de texto, tempo na tarefa), orquestrar os modelos de IA e, crucialmente, sincronizar temporalmente todos os fluxos de dados (interação, áudio, fisiologia) com precisão de milissegundos, um requisito fundamental para a análise de eventos relacionados.

Medidas Subjetivas: *NASA-TLX:* Questionário multidimensional para avaliar a carga de trabalho percebida em seis escalas, incluindo Demanda Mental, Esforço e Frustração; *Single Ease Question (SEQ):* Escala de 7 pontos para uma avaliação rápida e direta da dificuldade percebida da tarefa.

Medidas Psicofisiológicas: *Eletroencefalografia (EEG):* Utilizando o dispositivo portátil InteraXon Muse 2, a análise se concentrará em correlatos neurais robustos da carga cognitiva. Especificamente, um aumento na potência da banda *Téta frontal* (4−8 Hz) é um indicador validado da carga da memória de trabalho. Simultaneamente, uma supressão da potência da banda *Alfa parietal* (8−12 Hz) reflete o engajamento atencional e a alocação de recursos corticais para a tarefa. Será também calculado o *Theta/Alpha Ratio (TAR)* como um índice integrado e potencialmente mais sensível da carga cognitiva.

Pupilometria (POG): Com um rastreador ocular de alta frequência (Tobii Eye Tracker 5), a métrica principal será a *Resposta Pupilar Evocada pela Tarefa (TEPR)*. A dilatação pupilar é um proxy fidedigno da atividade do sistema locus coeruleus-norepinefrina (LC-NE), um neuromodulador central para o esforço cognitivo. A metodologia incluirá procedimentos rigorosos para corrigir o reflexo pupilar à luz (PLR), isolando a dilatação fásica relacionada à cognição daquela induzida por variações na luminância da tela, garantindo a validade da

medida. *Fotopletismografia (PPG) e Variabilidade da Frequência Cardíaca (HRV):* Medida com um smartwatch (Samsung Watch5), a HRV avalia a atividade do sistema nervoso autônomo (SNA). Serão analisadas métricas do domínio do tempo, como o *RMSSD* (raiz quadrada da média das diferenças sucessivas ao quadrado entre intervalos R-R), que reflete o tônus parassimpático (vagal), e do domínio da frequência, como a *razão LF/HF*, que indica o balanço simpato-vagal. Uma HRV reduzida (baixo RMSSD) indica dominância simpática, uma resposta fisiológica ao estresse e à carga mental elevada. Estudos longitudinais associam um baixo tônus parassimpático a um pior desempenho em funções executivas, posicionando a HRV como um preditor do potencial cognitivo.

Análise Quantitativa: Os dados fisiológicos serão pré-processados para remoção de artefatos. A principal ferramenta estatística será a modelagem por *Modelos Lineares Generalizados Mistos (GLMMs)*. Esta abordagem é ideal para analisar os efeitos das variáveis independentes (Fidelidade, Complexidade, Expertise) e suas interações sobre as diversas medidas dependentes (subjetivas, fisiológicas, de desempenho), controlando adequadamente as medidas repetidas dentro dos sujeitos e a variabilidade individual. A força desta metodologia reside na capacidade de realizar uma análise orientada a eventos, modelando a resposta fisiológica em janelas de tempo curtas após interações específicas com a IA (ex: geração de uma resposta), o que permite inferir causalidade momento a momento.

Análise Qualitativa: As transcrições dos protocolos de pensamento em voz alta serão submetidas à Análise Temática para identificar padrões nas estratégias cognitivas, dificuldades percebidas e reações afetivas dos usuários.

Triangulação: A integração das análises quantitativa e qualitativa permitirá que os achados de um método contextualizem e expliquem os do outro. Por exemplo, um aumento significativo na potência Téta (quantitativo) na condição Q4 será explicado por temas como “frustração com a imprecisão” e “esforço de verificação” emergentes da análise qualitativa.

RESULTADOS ESPERADOS E DISCUSSÃO

A principal hipótese é que a redução da fidelidade do assistente (Q4 vs. Q8) aumentará o custo cognitivo-afetivo do aprendiz. Espera-se um efeito principal significativo da Fidelidade em todas as medidas. Os participantes na condição Q4 relatarão maior carga de trabalho percebida (NASA-TLX) e dificuldade (SEQ). Crucialmente, espera-se que também relatem menor motivação intrínseca e maior tédio, replicando os achados de Liu et

al., (2025) para o ambiente de aprendizagem.

Fisiologicamente, a condição Q4 induzirá um aumento na potência Téta frontal, uma supressão da Alfa parietal, um aumento na TEPR e uma redução na HRV (menor RMSSD), indicando maior carga na memória de trabalho, esforço atencional e estresse fisiológico. O desempenho na tarefa (qualidade da solução) será inferior na condição Q4. A análise qualitativa revelará que este efeito não é apenas devido a uma sobrecarga, mas a uma estratégia de "cognitive offloading". Surgirão temas de "Desconfiança na IA", "Sobrecarga de Verificação" e, paradoxalmente, "Aceitação Passiva", onde os aprendizes, frustrados, começam a incorporar o feedback da IA sem escrutínio crítico, um comportamento mencionado por Chow (2025).

A segunda hipótese postula que o aumento da complexidade da interação (de Agente Único para Multiagente) aumentará a carga cognitiva de gerenciamento. Espera-se um efeito principal da Complexidade em todas as medidas de carga. As condições Multiagente, especialmente a Paralela, imporão uma carga significativamente maior do que a de Agente Único. A análise orientada a eventos identificará picos de carga fisiológica durante as fases de transição entre agentes e de integração de informações potencialmente conflitantes. A análise qualitativa fornecerá o contexto, com temas como "Perdido no Fluxo de Trabalho" e "Conflito de Feedback" explicando a origem desta carga adicional.

A hipótese mais rica em nuances é a interação entre Fidelidade e Expertise. O impacto negativo da baixa fidelidade (Q4) será drasticamente mais pronunciado para Iniciantes do que para Especialistas. Os Especialistas, possuindo esquemas de conhecimento robustos, identificarão e descartarão rapidamente o feedback incorreto do assistente Q4. Fisiologicamente, isso se manifestará como picos de carga mais curtos e uma recuperação mais rápida da linha de base. Este fenômeno é uma manifestação clássica do "efeito de expertise reversa", onde um suporte instrucional que é útil para iniciantes se torna redundante ou prejudicial para especialistas. Em contraste, os Iniciantes, sem uma base de conhecimento para ancorar sua verificação, ficarão mais confusos e sobrecarregados, exibindo respostas de estresse fisiológico mais pronunciadas e duradouras. A análise qualitativa revelará estratégias de pensamento distintas: especialistas descreverão heurísticas de verificação eficientes, enquanto iniciantes expressarão maior dependência e confusão.

A tabela a seguir sintetiza as predições, conectando cada hipótese às suas manifestações esperadas nos diferentes canais de dados. Esta matriz serve como um roteiro conceitual, demonstrando a coerência do desenho experimental e como a convergência das evidências fortalecerá a validade das conclusões.

Hipótese	Condição Experimental Chave	Medida Subjetiva (NASA-TLX, SEQ)	Medida Fisiológica (EEG, POG, HRV)	Medida de Desempenho/ Qualitativa
H1: Paradoxo da Fidelidade	Baixa Fidelidade (Q4) vs. Alta (Q8)	↑ Carga Percebida, ↑ Frustração, ↓ Motivação Intrínseca	↑ Téta Frontal, ↓ Alfa Parietal, ↑ TEPR, ↓ RMSSD (HRV)	↓ Qualidade da Tarefa; Temas de "Desconfiança", "Sobrecarga de Verificação", "Aceitação Passiva"
H2: Custo da Complexidade	Multiagente vs. Agente Único	↑ Demanda Mental (Gerenciamento)	Picos de carga nos pontos de integração/ transição	↓ Eficiência; Temas de "Conflito de Feedback", "Perdido no Fluxo"
H3: Moderação pela Expertise	Baixa Fidelidade (Q4) x Iniciantes vs. Especialistas	Efeito negativo da Q4 amplificado para Iniciantes	Respostas de estresse (↓ HRV) mais pronunciadas e duradouras em Iniciantes	Especialistas demonstram "verificação eficiente"; Iniciantes mostram "confusão e dependência"

A síntese destes resultados levará à discussão de uma relação não linear, em forma de U, entre a carga computacional e o que se redefine como "Custo Cognitivo-Afetivo". No extremo de baixa carga computacional (Q4), o custo é alto devido à imprecisão, que gera ECL e custos afetivos (frustração, perda de controle, tédio). No extremo de carga computacional muito alta (um sistema lento), o custo também aumenta devido à latência que quebra o fluxo cognitivo. Entre estes extremos, existe um "ponto ideal cognitivo-computacional" – representado pela condição Q8 responsiva – onde a qualidade do assistente minimiza a ECL sem impor uma carga sistêmica que prejudique a fluidez da interação. Este conceito se alinha funcionalmente à Lei de Yerkes-Dodson (1908), que descreve uma relação semelhante entre excitação e desempenho.

CONSIDERAÇÕES FINAIS

Os resultados esperados deste estudo experimental sugerem que a integração bem-sucedida da IA na educação depende fundamentalmente da otimização da interação cognitiva entre o aprendiz e a máquina. A principal contribuição teórica é a proposição de um **Modelo Integrado de Interação Cognitiva com IA**, que estende a Teoria da Carga Cognitiva ao incorporar formalmente: (1) a dimensão **motivacional** do senso de controle e motivação intrínseca, conforme destacado por Liu et al. (2025), e (2) o mecanismo de **cognitive offloading** como um comportamento adaptativo (mas muitas vezes mal-

adaptativo) à qualidade da IA, evidenciado por Chow (2025). Este modelo postula que a eficácia de um sistema de IA para a aprendizagem não é uma função monotônica de sua precisão técnica, mas sim de sua capacidade de manter o aprendiz em um estado de engajamento cognitivo e afetivo ótimo.

As implicações práticas são diretas e impactantes. Para **desenvolvedores de IA e designers de interface**, a prioridade deve ser a **confiabilidade e a previsibilidade** sobre a velocidade bruta ou a eficiência computacional. É preferível um assistente confiável (Q8), mesmo que marginalmente mais lento, a um rápido (Q4) que gera respostas erráticas e aumenta a frustração do usuário. O desempenho do backend (FastAPI) e o design da interface (React) são cruciais para minimizar a ECL relacionada ao sistema e abstrair a complexidade dos fluxos multiagente.

Para **educadores e designers instrucionais**, a IA deve ser enquadrada não como uma ferramenta para "fazer o trabalho", mas como um **parceiro de andaime cognitivo**. As tarefas de aprendizagem devem ser projetadas para incentivar o uso da IA para ideação, exploração ou verificação, mas exigir que o aprendiz realize a síntese crítica, a argumentação e a integração final. Esta abordagem preserva o investimento na Carga Germânica (GCL), que é o motor da aprendizagem profunda e do desenvolvimento de habilidades de pensamento crítico.

A visão de futuro aponta para o desenvolvimento de **sistemas co-adaptativos**. A realização da promessa da IA na educação reside na criação de sistemas que utilizam o monitoramento neurofisiológico em tempo real para se adaptarem dinamicamente ao estado cognitivo do usuário. Um sistema que detecta sinais de sobrecarga cognitiva (ex: aumento sustentado da potência teta, queda na HRV) poderia intervir proativamente: simplificando a tarefa, oferecendo um andaime mais estruturado ou ajustando a fidelidade do modelo de IA para reencontrar o "ponto ideal" da curva ergonômica. Ao projetar sistemas que monitoram e gerenciam a carga cognitiva em tempo real, é possível criar ferramentas de IA que não apenas entregam informação, mas que capacitam os aprendizes a pensar de forma mais clara, engajar-se mais profundamente e construir conhecimento de forma significativa e sustentável. A resposta não está na maximização da máquina, mas no equilíbrio e na adaptação contínua que respeitam os limites e potencializam as capacidades da mente humana.

REFERÊNCIAS

BEATTY, J.; LUCERO-WAGONER, B. The pupillary system. In: CACIOPPO, J. T.; TASSINARY, L. G.; BERNTSON, G. G. (Eds.). **Handbook of psychophysiology**. 2. ed. Cambridge: Cambridge University Press, 2000. p. 142-162.

CHOW, Andrew R. **ChatGPT may be eroding critical thinking skills, according to a**

new MIT study. Time, [2025]. Disponível em: https://time.com/7295195/ai-chatgpt-google-learning-school/. Acesso em: 20 jun. 2025.

GEVINS, A.; SMITH, M. E. Neurophysiological measures of working memory and individual differences in cognitive ability and cognitive style. **Cerebral Cortex**, v. 10, n. 9, p. 829-839, 2000.

HJORTSKOV, N. et al. The effect of mental stress on heart rate variability and blood pressure during computer work. **European Journal of Applied Physiology**, v. 92, n. 1-2, p. 84-89, 2004.

KLIMESCH, W. EEG alpha and theta oscillations reflect cognitive and memory performance: a review and analysis. **Brain Research Reviews**, v. 29, n. 2-3, p. 169-195, 1999.

LIU, Yukun et al. **Research: Gen AI makes people more productive—and less motivated**. Harvard Business Review, 13 maio 2025. Disponível em: https://hbr.org/2025/05/research-gen-ai-makes-people-more-productive-and-less-motivated. Acesso em: 20 jun. 2025.

PAAS, F.; VAN MERRIËNBOER, J. J. G. Variability of worked examples and transfer of geometrical problem-solving skills: A cognitive-load approach. **Journal of Educational Psychology**, v. 86, n. 1, p. 122–133, 1994.

SWELLER, J. Cognitive load during problem solving: Effects on learning. **Cognitive Science**, v. 12, n. 2, p. 257-285, 1988.

SWELLER, J.; CHANDLER, P. Why some material is difficult to learn. **Cognition and Instruction**, v. 12, n. 3, p. 185-233, 1994.

YERKES, R. M.; DODSON, J. D. The relation of strength of stimulus to rapidity of habit□ formation. **Journal of Comparative Neurology and Psychology**, v. 18, n. 5, p. 459-482, 1908.