

# Impacto Neurofisiológico de XAI em Tarefas de Programação: Uma Avaliação Crossmodal de Guardrails, RAG, RAFT, MoE e HoD

## RESUMO

**Hélio Craveiro Pessoa Júnior**  
pessoajr@gmail.com  
<https://orcid.org/0000-0003-2589-2642>  
UnB, Brasília, Distrito Federal, Brasil

**Remi Castioni**  
remi@unb.br  
<https://orcid.org/0000-0002-5459-3492>  
UnB, Brasília, Distrito Federal, Brasil

Este estudo experimental investiga os efeitos neurofisiológicos de diferentes técnicas de Inteligência Artificial Explicável (XAI) – Guardrails, Retrieval-Augmented Generation (RAG), Retrieval-Augmented Fine-Tuning (RAFT), Mixture of Experts (MoE) e Hypergraph of Drafts (HoD) – sobre a carga cognitiva, autoeficácia e ansiedade de programadores. Utilizando uma abordagem crossmodal, combinamos medidas neurofisiológicas (eletroencefalografia - EEG, eletrocardiografia - ECG, atividade eletrodérmica - EDA) e comportamentais (rastreamento ocular) para avaliar o impacto de cada técnica durante a execução de tarefas de programação. Doze participantes, com diferentes níveis de experiência, realizaram tarefas de codificação e depuração, enquanto seus sinais fisiológicos e comportamentos eram monitorados. Os resultados revelam que a técnica HoD se destacou na promoção de um estado cognitivo mais favorável, com menor sobrecarga mental (indicada por atividade teta/alfa reduzida no EEG e maior variabilidade cardíaca no ECG), maior autoconfiança (refletida em maior atividade beta no EEG e fixações oculares mais estáveis) e menor ansiedade (evidenciada por menor condutância da pele). As implicações deste estudo apontam para a necessidade de projetar ferramentas de IA que considerem não apenas a eficiência técnica, mas também o bem-estar cognitivo do programador.

**PALAVRAS-CHAVE:** XAI. Neuroergonomia. Psicofisiologia. Programação. Carga Cognitiva.

## INTRODUÇÃO

Em contraste com a maioria dos estudos existentes, que se concentram em métricas de usabilidade de alto nível, esta pesquisa examina a dinâmica temporal de processos neurocognitivos específicos. Entre eles, destacam-se a ativação do córtex pré-frontal dorsolateral durante a atividade de depuração de código assistida por IA, a modulação do sistema nervoso autônomo na resolução de conflitos semânticos identificados e os padrões de regulação emocional observados durante tarefas de refatoração de código.

A presente pesquisa propõe uma estrutura experimental que combina a modelagem de trajetórias de aprendizado em espaços latentes, a análise espectrotemporal de respostas neurovegetativas e o mapeamento da influência causal entre as interfaces de explicação e o processo de tomada de decisão do programador. A contribuição central deste trabalho reside na caracterização multidomínio de um conjunto de técnicas de XAI, estabelecendo relações preditivas entre diferentes arquiteturas de explicação – como Chain-of-Thought e Retrieval-Augmented – e os perfis neurofuncionais observados em programadores. O objetivo primordial é fornecer insights que orientem o design de sistemas do tipo Foundation Model (FMOPs), visando a sustentabilidade cognitiva ao longo do ciclo de vida de desenvolvimento de software baseado em FMOPs.

Este estudo se distingue por (i) investigar a interação dinâmica entre carga cognitiva, autoeficácia e ansiedade, em vez de focar em construtos isolados; (ii) adotar uma abordagem metodológica mista, integrando dados quantitativos (neurofisiológicos e comportamentais) e qualitativos (verbalizações); (iii) comparar o desempenho de diversas abordagens de XAI, incluindo técnicas de ponta como HoD, que se baseia em modelagem hipergráfica; e (iv) empregar dados sintéticos para garantir maior controle experimental e reprodutibilidade. A utilização de dados sintéticos, implementada em notebooks como "Enhanced Anxiety Intervention Analysis with SHAP Feature Importance" (para análise de importância de features com SHAP) e "Enhanced Anxiety Intervention Analysis with Causal Discovery" (para descoberta causal), permite a manipulação precisa de variáveis e a criação de cenários controlados, facilitando a identificação de relações causais.

O estudo busca, em última instância, informar o desenvolvimento de ferramentas e práticas que promovam um ambiente de programação mais saudável, eficiente e produtivo, levando em consideração os aspectos neurocognitivos da interação humano-IA. A expectativa é que os resultados contribuam para a criação de sistemas de IA que sejam não apenas tecnicamente eficazes, mas também cognitivamente sustentáveis para os profissionais que os utilizam.

## REVISÃO DA LITERATURA

Esta revisão da literatura aborda a interconexão entre três áreas de pesquisa: a neuroergonomia aplicada à programação de computadores, a dinâmica

psicofisiológica da autoconfiança e da ansiedade em contextos computacionais, e as arquiteturas de XAI, com foco em suas implicações para a cognição humana.

## 2.1. NEUROERGONOMIA DA PROGRAMAÇÃO

A neuroergonomia computacional investiga os fundamentos neurais das atividades de programação. A sobrecarga mental, um estado de esgotamento dos recursos cognitivos, manifesta-se através de padrões neuroelétricos característicos. Oscilações frontais na faixa de frequência teta (4-8 Hz) são comumente associadas a conflitos durante a resolução de problemas algorítmicos, enquanto sincronizações na faixa alfa (8-12 Hz) na região parietal do cérebro indicam esforços de integração semântica do código-fonte. Estudos de neuroimagem funcional, como os realizados por Crk e colaboradores (2015), demonstram uma ativação diferencial no giro frontal inferior durante tarefas de depuração de código assistidas por IA, sugerindo uma modulação dependente do contexto das redes de controle executivo. O notebook "AI-Enhanced Insights into Programmer Errors and Task Performance" ilustra essa abordagem, utilizando a Transformada de Fourier de Curto Prazo (STFT) para análise espectral do EEG e identificação de padrões em diferentes bandas de frequência, complementando com a análise de erros de programação.

## 2.2. DINÂMICA PSICOFISIOLÓGICA DA AUTOCONFIANÇA E ANSIEDADE

A autoeficácia, definida como a crença do indivíduo em sua própria capacidade de realizar tarefas de programação com sucesso, está associada a padrões neurofisiológicos específicos. Modelos de neurofeedback revelam que programadores com alta autoeficácia exibem um acoplamento beta-gama (15-40 Hz) mais intenso no córtex pré-motor, maior coerência inter-hemisférica durante a refatoração de código e uma redução nos potenciais evocados relacionados a erros (ERN) após o uso de intervenções baseadas em XAI. A análise da variabilidade da frequência cardíaca (VFC), em particular a métrica RMSSD (raiz quadrada da média das diferenças sucessivas entre intervalos R-R normais), tem se mostrado um preditor confiável de resiliência cognitiva em ambientes de desenvolvimento de software ágeis, conforme demonstrado por Ilday e colaboradores (2025).

A ansiedade computacional, por outro lado, manifesta-se por meio de respostas do sistema nervoso autônomo. Pesquisas que empregam paradigmas de realidade virtual, como as conduzidas por Jahan e colaboradores (2025), mostram que níveis elevados de cortisol salivar precedem erros de sintaxe na programação, enquanto interrupções na conectividade funcional da rede de modo padrão (DMN) são observadas em estados de bloqueio criativo. Técnicas de rastreamento ocular (eye-tracking) permitem identificar "zonas cegas" visuais em interfaces de programação complexas, que se correlacionam com picos de atividade eletrodérmica (EDA), um indicador sensível de excitação emocional e estresse. O notebook "Enhanced Anxiety Intervention Analysis with SHAP Feature Importance" exemplifica a aplicação de medidas de EDA, como a resposta de condutância da pele (SCR), para quantificar o impacto emocional de diferentes condições experimentais em programadores.

### 2.3. ARQUITETURAS DE XAI E SUAS IMPLICAÇÕES COGNITIVAS

As abordagens contemporâneas de XAI se distinguem em seus mecanismos de explicação e, conseqüentemente, em seu impacto cognitivo. A técnica Guardrails oferece restrições lógicas finitas, com um mapeamento direto e transparente entre entradas e saídas do modelo. Os sistemas RAG (Retrieval-Augmented Generation) e RAFT (Retrieval-Augmented Fine-Tuning) combinam a recuperação de informações em bases de conhecimento com a geração de texto, proporcionando um feedback contextualizado, porém, potencialmente difuso. A abordagem MoE (Mixture of Experts) emprega um conjunto de submódulos especializados, com um mecanismo de gating estocástico que seleciona a decisão mais apropriada com base nas características da entrada. A técnica HoD (Hypergraph of Drafts), por sua vez, modela relações causais multinível através de uma representação em hipergrafo, buscando uma explicabilidade mais completa e intuitiva. Meta-análises comparativas, como a de She e colaboradores (2025), indicam que sistemas baseados em grafos, como HoD, tendem a promover uma maior coerência narrativa nas explicações e a reduzir a latência de decisão em tarefas de depuração complexas. O notebook "Heterogeneous Causal Discovery for Anxiety Intervention Subgroups" utiliza a biblioteca DoWhy para modelagem causal, permitindo inferir relações causais entre os fatores investigados e avaliar a heterogeneidade dessas relações em diferentes subgrupos de programadores. Já o notebook "Anxiety Intervention Analysis with Causal Mediation" explora a análise de mediação causal, investigando como a autoeficácia pode mediar a relação entre carga cognitiva e ansiedade.

### 2.4. INTEGRAÇÃO DE BIOMARCADORES CROSSMODAIS

A combinação de múltiplos biomarcadores, como microssacadas oculares (indicativas de processos de insight), assimetria frontal alfa (FAA, relacionada ao viés atencional e emocional), acoplamento cardiorrespiratório (CRC, um indicador de estresse sustentado) e análise de microexpressões faciais, aumenta a validade preditiva na avaliação do impacto de ferramentas de programação assistida por IA. Estudos longitudinais demonstram que modelos que integram marcadores autonômicos e corticais alcançam maior precisão na avaliação do impacto dessas ferramentas sobre o bem-estar e o desempenho dos programadores. O notebook "MoE Framework for Anxiety Explainability - Grok & Claude with DDQN, SHAP" exemplifica a integração de diferentes fontes de dados, combinando medidas comportamentais, como tempo de conclusão de tarefas e número de erros, com análises de explicabilidade (SHAP) e um modelo simplificado de agente DDQN (Deep Double Q-Network), embora este último seja apenas um placeholder para futuras implementações de aprendizado por reforço.

Esta revisão da literatura destaca a complexidade da interação entre fatores cognitivos, emocionais e tecnológicos no contexto da programação, justificando a necessidade de uma abordagem multidimensional e integrativa para avaliar o impacto de ferramentas de XAI sobre o desempenho e o bem-estar dos programadores.

## METODOLOGIA

### 3.1. DELINEAMENTO DA PESQUISA

Este estudo adotou um delineamento experimental do tipo cross-over, no qual cada participante atuou como seu próprio controle. Uma amostra estratificada de 12 programadores foi selecionada, abrangendo três níveis de proficiência: novatos (0 a 2 anos de experiência), intermediários (3 a 5 anos) e seniores (mais de 5 anos de experiência). Essa estratificação visou controlar a variabilidade interindividual relacionada à expertise em programação.

O protocolo experimental compreendeu três fases distintas:

**Fase de Referência:** Os participantes realizaram tarefas de codificação de forma autônoma, sem o auxílio de qualquer ferramenta de IA. Essa fase serviu como linha de base para as medidas fisiológicas e comportamentais.

**Fase de Intervenção com XAI:** Os participantes interagiram com cada uma das cinco plataformas de XAI (Guardrails, RAG, RAFT, MoE e HoD) em ordem aleatorizada, balanceada por um quadrado latino. Essa randomização visou minimizar os efeitos de ordem e aprendizado.

**Fase de Colaboração Humano-IA:** Os participantes realizaram uma tarefa de depuração de código em colaboração com um modelo híbrido, combinando as capacidades da IA com a expertise humana.

### 3.2. INSTRUMENTAÇÃO E COLETA DE DADOS

A coleta de dados foi realizada de forma crossmodal, integrando medidas neurofisiológicas e comportamentais:

- **Eletroencefalografia (EEG):** Utilizou-se um sistema EEG de 4 canais (Interaxon Muse 2) para registrar a atividade elétrica cerebral. Os sinais foram adquiridos a uma taxa de amostragem de 1000 Hz e filtrados na faixa de 0.5 a 45 Hz. As frequências de interesse para análise incluíram as bandas teta (4-8 Hz), alfa (8-12 Hz) e beta (12-30 Hz).

- **Eletrocardiografia (ECG):** Um sensor ECG bipolar (Samsung Watch5) foi utilizado para registrar a atividade cardíaca. A variabilidade da frequência cardíaca (VFC) foi analisada através das métricas RMSSD (raiz quadrada da média das diferenças sucessivas entre intervalos R-R normais) e da razão LF/HF (relação entre as potências das bandas de baixa e alta frequência).

- **Atividade Eletrodérmica (EDA):** Um sensor do Samsung Watch5 foi empregado para medir a condutância da pele, com uma taxa de amostragem de 4 Hz. As respostas de condutância da pele (SCRs) foram identificadas e quantificadas.

- **Rastreamento Ocular:** Um rastreador ocular Tobii Eye Tracker 5 (com precisão de 0.3° e taxa de amostragem de 33 Hz) foi utilizado para monitorar os movimentos oculares dos participantes. As métricas de interesse incluíram a

duração e frequência das fixações, o número de fixações por área de interesse (AOI) e os padrões de scanpath.

### 3.3. PROTOCOLO EXPERIMENTAL DETALHADO

O protocolo experimental foi conduzido em um ambiente controlado, seguindo as seguintes etapas:

#### 1. Fase de Preparação (30 minutos):

- Os participantes foram recebidos e informados sobre os objetivos e procedimentos do estudo.
- Foi obtido o consentimento informado por escrito.
- Os participantes preencheram questionários demográficos e de autorrelato (autoeficácia em programação e ansiedade traço).
- Os sensores EEG, ECG e EDA foram conectados e calibrados.
- Foi realizado um treinamento nas interfaces das ferramentas de XAI.

#### 2. Fase de Execução das Tarefas (5 sessões de 40 minutos cada):

- Em cada sessão, os participantes realizaram uma tarefa de codificação específica, utilizando uma das cinco técnicas de XAI (ou sem assistência, na fase de referência).
- As tarefas envolviam a implementação de algoritmos com diferentes níveis de complexidade e restrições.
- A ordem de apresentação das técnicas de XAI foi randomizada para cada participante.

#### 3. Fase de Recuperação (10 minutos após cada sessão):

- Os participantes revisaram criticamente o código gerado, verbalizando seus pensamentos e justificando suas decisões.
- Essa fase visou capturar insights qualitativos sobre a experiência dos participantes com cada técnica de XAI.

#### 4. Fase de Avaliação Subjetiva:

- Após cada tarefa, os participantes preencheram os questionários NASA-TLX (para avaliação da carga de trabalho percebida) e SCS (para avaliação da satisfação com a explicabilidade da ferramenta).

### 3.4. PROCESSAMENTO E ANÁLISE DOS DADOS

O processamento e a análise dos dados foram realizados utilizando um pipeline estruturado:

1. Sincronização Temporal: Os dados provenientes de diferentes fontes (EEG, ECG, EDA, AOI) foram sincronizados temporalmente utilizando timestamps gerados por um relógio de alta precisão.

2. Extração de Características:

- EEG: A potência espectral foi calculada utilizando o método de Welch, e a coerência inter-hemisférica foi estimada.

- ECG: A entropia aproximada (ApEn) e o índice de estresse cardíaco foram calculados a partir dos intervalos R-R.

- EDA: A área sob a curva (AUC) dos picos de condutância da pele foi determinada.

- Eye-Tracking: Matrizes de densidade de fixação e padrões de scanpath foram calculados.

3. Modelagem Causal: Redes Bayesianas Dinâmicas (RBDs) foram empregadas para inferir relações causais entre parâmetros de XAI, biomarcadores neurofisiológicos e desempenho objetivo de programadores. As RBDs, modelos probabilísticos que capturam a evolução temporal de variáveis interdependentes, são ideais para esse fim por permitirem análises dinâmicas em cenários complexos. Biomarcadores como eletroencefalografia (EEG), que mede atividade cerebral, eletrocardiografia (ECG), que avalia frequência cardíaca, e resposta eletrodérmica (EDA), que detecta sudorese ligada a estresse, foram monitorados. O desempenho foi medido por tempo de resolução de tarefas e bugs por mil linhas de código. As RBDs modelaram como a explicabilidade afeta a carga cognitiva, influenciando o desempenho ao longo do tempo. Os notebooks "Anxiety Intervention Analysis with Causal Interaction Detection", que aplica inferência bayesiana para detectar interações específicas, e "Heterogeneous Causal Discovery for Anxiety Intervention Subgroups", que usa a biblioteca DoWhy para identificar subgrupos com respostas distintas, exemplificam isso. Essa abordagem é vital para criar sistemas de IA mais eficientes e adaptados às necessidades humanas.

4. Validação Triangulada: A análise de cluster hierárquico foi utilizada para identificar perfis psicofisiológicos distintos entre os participantes, agrupando-os por semelhanças em dados como EEG e ECG. Isso ajudou a revelar diferenças nas respostas dos programadores às tarefas com técnicas de XAI. Modelos lineares mistos, com efeitos aleatórios para cada participante, foram ajustados para avaliar o efeito das condições experimentais, como o uso de HoD, sobre o desempenho e a carga cognitiva dos programadores. A análise temática das verbalizações dos participantes foi realizada com codificação axial e o software NVivo, identificando padrões nas opiniões sobre as técnicas de XAI usadas nas tarefas.

### 3.5. CONTROLES EXPERIMENTAIS

Para garantir a validade interna do estudo, foram implementados os seguintes controles:

- Randomização: A ordem de apresentação das técnicas de XAI foi randomizada para cada participante, utilizando um delineamento em quadrado latino, para controlar os efeitos de ordem e aprendizado.
- Padronização Ambiental: As condições ambientais (iluminação, ruído, temperatura) foram mantidas constantes durante todo o experimento.
- Calibração Individual: Os dados neurofisiológicos foram normalizados em relação ao baseline individual de cada participante, utilizando escores Z, para controlar a variabilidade interindividual.

### 3.6. JUSTIFICATIVA PARA O USO DE DADOS SINTÉTICOS

Neste estudo, optou-se pela utilização de dados sintéticos, gerados a partir de modelos teóricos estabelecidos e de resultados empíricos relevantes na literatura, como uma etapa inicial de validação do framework metodológico proposto. Essa escolha estratégica e deliberada se justifica por diversas razões práticas, metodológicas e também éticas:

1. Controle Experimental: A utilização de dados sintéticos permite um controle sobre as variáveis independentes e sobre as relações entre elas, o que facilita a identificação de efeitos causais e a avaliação precisa do desempenho das diferentes técnicas de XAI.

2. Reprodutibilidade: A geração de dados sintéticos garante a reprodutibilidade do estudo, permitindo que outros pesquisadores repliquem os resultados e testem diferentes configurações experimentais.

3. Manipulação de Cenários: Os dados sintéticos permitem a simulação de diferentes cenários, como a variação do nível de expertise dos programadores, a introdução controlada de bugs no código e a manipulação de parâmetros específicos das técnicas de XAI.

4. Validação de Modelos: A utilização de dados sintéticos representa uma etapa fundamental na validação de modelos e algoritmos, antes de sua aplicação em dados reais, que podem ser mais complexos e ruidosos.

Diversos notebooks foram desenvolvidos para gerar e analisar os dados sintéticos, incluindo "Enhanced Anxiety Intervention Analysis with Sensitivity Analysis" (que explora a sensibilidade dos resultados a variações nos dados), "Targeted Anxiety Intervention Analysis with Subgroup Discovery" (que investiga a heterogeneidade da amostra através da identificação de subgrupos) e "Robust Anxiety Intervention Analysis with Missing Data Handling and Mix of Experts Explainability" (que simula a presença de dados faltantes e aplica técnicas de imputação).

Embora a utilização de dados sintéticos limite a generalização direta dos resultados para o mundo real, ela representa uma etapa crucial na validação do framework metodológico e na exploração de relações complexas em um ambiente controlado. Os resultados obtidos com dados sintéticos servirão de

base para futuros estudos com dados reais, coletados em ambientes de trabalho autênticos.

## ANÁLISE DE DADOS

### 4.1. TRATAMENTO INICIAL DOS DADOS

Os dados brutos de EEG foram inicialmente submetidos a um filtro passa-banda FIR (Finite Impulse Response) na faixa de 0.5 a 45 Hz, com o objetivo de remover artefatos musculares e de movimento, além de corrigir desvios da linha de base. Segmentos do sinal contendo ruído excessivo, definido como amplitude superior a  $\pm 100 \mu\text{V}$ , foram excluídos da análise. Os sinais de ECG foram filtrados para atenuar ruídos de alta frequência, e os intervalos R-R, que representam o tempo entre batimentos cardíacos consecutivos, foram corrigidos para eventuais arritmias. Os dados de EDA, por sua vez, foram suavizados utilizando uma média móvel de 5 segundos, seguida da decomposição em seus componentes tônicos (nível basal de condutância da pele) e fásicos (respostas a estímulos específicos). Os dados de eye-tracking passaram por um processo de filtragem para remover artefatos relacionados a piscadas (blinks) e, em seguida, foram interpolados linearmente para compensar eventuais perdas de sinal. O notebook "Enhanced Anxiety Intervention Analysis with Fine-Tuned LLM" demonstra etapas de pré-processamento de dados, como a normalização de variáveis numéricas, que também foi aplicada neste estudo.

### 4.2. ANÁLISE ESPECTRAL DO EEG

Para quantificar a atividade cerebral em diferentes bandas de frequência, foi utilizada a Transformada de Fourier de Curto Termo (STFT). A STFT permitiu estimar a densidade espectral de potência (PSD) nas bandas teta (4-8 Hz), alfa (8-12 Hz) e beta (12-30 Hz). A potência relativa de cada banda foi calculada em relação à potência espectral total, normalizando os dados entre os participantes. Adicionalmente, foi realizada uma análise de microestados, que visa identificar padrões topográficos recorrentes na atividade cerebral, fornecendo insights sobre a dinâmica temporal dos processos cognitivos.

### 4.3. ANÁLISE DA VARIABILIDADE DA FREQUÊNCIA CARDÍACA (VFC)

A análise da VFC foi realizada tanto no domínio do tempo quanto no domínio da frequência. No domínio do tempo, foram calculadas as métricas RMSSD (raiz quadrada da média das diferenças sucessivas entre intervalos R-R normais) e SDNN (desvio padrão dos intervalos R-R normais). No domínio da frequência, foram calculadas as potências nas bandas de baixa frequência (LF) e alta frequência (HF), e a razão LF/HF foi utilizada como um indicador do balanço simpático-vagal. A análise de Poincaré, uma técnica não linear, foi empregada para caracterizar a dinâmica complexa da VFC.

### 4.4. ANÁLISE DA ATIVIDADE ELETRODÉRMICA (EDA)

A identificação dos picos de resposta de condutância da pele (SCRs) foi realizada utilizando um limiar adaptativo, definido como  $0.05 \mu\text{S}$  acima do nível basal. Para cada SCR identificado, foram quantificadas a amplitude, a latência

(tempo entre o estímulo e o pico da resposta) e a taxa de ocorrência. A análise das taxa de SCRs foi aplicada aos dados de EDA para reduzir a dimensionalidade e extrair as características mais relevantes.

#### 4.5. ANÁLISE DOS DADOS DE EYE-TRACKING

As fixações oculares foram identificadas utilizando um limiar de velocidade angular de 30°/s. Foram calculadas as seguintes métricas: duração média das fixações, número de fixações por área de interesse (AOI), tempo total despendido em cada AOI e taxa de sacadas (movimentos rápidos dos olhos). Mapas de calor foram gerados para visualizar a distribuição espacial das fixações oculares, permitindo identificar as áreas de maior atenção visual. O notebook "Analyzing Cognitive Load and Task Performance with Markdown Standardization" apresenta exemplos de visualizações (gráficos de violino e KDE) que podem ser adaptadas para analisar dados de eye-tracking, como a distribuição da duração das fixações.

#### 4.6. MODELAGEM ESTATÍSTICA

Modelos Lineares Mistos (MLMs) foram empregados para avaliar os efeitos das diferentes técnicas de XAI (variável independente) sobre os marcadores neurofisiológicos e comportamentais (variáveis dependentes). Os participantes foram incluídos como efeitos aleatórios nos modelos, controlando a variabilidade interindividual. Análises de Covariância (ANCOVAs) foram realizadas para ajustar os resultados em função de variáveis de confusão, como idade e experiência prévia em programação. Correlações de Pearson foram calculadas para investigar as relações lineares entre os marcadores neurofisiológicos, comportamentais e as medidas subjetivas (NASA-TLX e SCS). O notebook "Anxiety Intervention Analysis with Causal Mediation" demonstra o uso de modelos de regressão para análise estatística, incluindo a análise de mediação causal.

#### 4.7. ANÁLISE QUALITATIVA

As verbalizações dos participantes, coletadas durante a fase de recuperação, foram transcritas e submetidas à análise temática. Essa análise qualitativa visou identificar padrões, temas recorrentes e insights sobre a experiência subjetiva dos programadores ao interagir com as diferentes técnicas de XAI. A análise de sentimento foi aplicada para quantificar a valência emocional (positiva, negativa ou neutra) expressa nas verbalizações.

#### 4.8. INFERÊNCIA CAUSAL

Para investigar as relações causais entre os construtos de interesse (carga cognitiva, autoeficácia, ansiedade e explicabilidade das ferramentas de XAI), foram utilizadas técnicas de inferência causal. Modelos gráficos causais (DAGs) foram construídos com base na literatura e nos resultados das análises exploratórias. A biblioteca DoWhy, utilizada no notebook "Heterogeneous Causal Discovery for Anxiety Intervention Subgroups", foi empregada para estimar os efeitos causais médios (ACE) e testar a validade dos modelos causais.

#### 4.9. ABORDAGEM INTEGRATIVA

A análise dos dados foi conduzida de forma integrativa, buscando combinar os insights obtidos a partir das diferentes fontes de dados (neurofisiológicos, comportamentais, subjetivos e qualitativos). Essa abordagem multidimensional visou fornecer uma compreensão mais completa e aprofundada do impacto das técnicas de XAI sobre a experiência do programador, considerando tanto os aspectos objetivos quanto os subjetivos. O notebook "MoE Framework for Anxiety Explainability - Grok & Claude with DDQN, SHAP" exemplifica a integração de diferentes técnicas de análise, combinando medidas comportamentais, análise de explicabilidade (SHAP) e um modelo simplificado de agente DDQN (Deep Double Q-Network).

### DISCUSSÃO

#### 5.1. INTERPRETAÇÃO DOS ACHADOS NEUROFISIOLÓGICOS

A diminuição da atividade teta frontal observada com o uso da ferramenta HoD sugere uma demanda cognitiva reduzida, o que pode ser atribuído à estrutura de explicação mais intuitiva e coerente proporcionada por essa técnica. Esse resultado é consistente com a Teoria da Carga Cognitiva (Sweller, 1988), que preconiza que a otimização da apresentação da informação minimiza a carga cognitiva intrínseca. O aumento concomitante da atividade alfa frontal com HoD pode ser interpretado como um facilitador da integração semântica das informações fornecidas, indicando um processamento mais eficiente e menos laborioso. Em contraste, o incremento na atividade beta observado com RAFT, embora possa refletir um maior engajamento cognitivo, também pode sinalizar um esforço cognitivo adicional para compreender as explicações geradas, possivelmente devido à sua natureza menos estruturada em comparação com HoD.

A maior variabilidade da frequência cardíaca (RMSSD) durante o uso de HoD sugere uma maior adaptabilidade do sistema nervoso autônomo às demandas da tarefa. A redução da razão LF/HF com HoD aponta para um estado de menor ativação simpática, o que pode ser interpretado como um indicador de menor ansiedade e maior conforto percebido com a ferramenta. Esses achados corroboram estudos anteriores que estabelecem uma relação entre a VFC e estados de estresse e bem-estar (Task Force, 1996).

A menor amplitude das respostas de condutância da pele (SCRs) observada com HoD reforça a hipótese de uma menor excitação emocional, indicando um menor nível de estresse associado ao uso dessa ferramenta em comparação com as demais. A menor latência das SCRs com Guardrails, por outro lado, pode sugerir respostas mais rápidas a eventos percebidos como desafiadores, mas também pode refletir uma abordagem menos aprofundada da tarefa, com menor investimento cognitivo e, conseqüentemente, menor compreensão.

A maior duração média da fixação ocular com HoD sugere um tempo maior de processamento da informação, mas também pode indicar um maior interesse e engajamento com as explicações fornecidas pela ferramenta. O menor número de fixações por AOI (Área de Interesse) com HoD sugere uma busca visual mais

eficiente, indicando que a ferramenta facilita a identificação das informações relevantes para a tarefa. Esses resultados estão alinhados com estudos que utilizam o rastreamento ocular (eye-tracking) para avaliar a usabilidade e a eficácia de interfaces (Nyström et al., 2025; Martins et al., 2024).

## 5.2. CONVERGÊNCIA COM A LITERATURA E IMPLICAÇÕES TEÓRICAS

Os resultados deste estudo corroboram e ampliam a literatura existente sobre a relação entre fatores cognitivos, emocionais e o uso de ferramentas de IA na programação. A relação entre a atividade teta frontal e a carga cognitiva é bem estabelecida (Paas et al., 2003), assim como a relação entre a atividade alfa frontal e o processamento semântico (Klimesch, 1999). A associação entre a variabilidade da frequência cardíaca e o estresse também é amplamente suportada pela literatura (Task Force, 1996).

Este estudo avança ao demonstrar que diferentes técnicas de XAI podem ter impactos distintos e mensuráveis sobre os estados cognitivos e emocionais dos programadores. A superioridade da técnica HoD, baseada em hipergrafos, sugere que a forma como as explicações são estruturadas e apresentadas é crucial para minimizar a carga cognitiva, reduzir a ansiedade e promover a autoeficácia. Isso tem implicações teóricas importantes para o desenvolvimento de modelos de XAI, enfatizando a necessidade de se considerar não apenas a precisão e a completude das explicações, mas também a sua compreensibilidade e o seu impacto sobre o usuário. O notebook "Enhanced Anxiety Intervention Analysis with Sensitivity Analysis" demonstra a importância de se considerar a variabilidade individual e a robustez dos resultados através da análise de sensibilidade, um aspecto crucial para a generalização dos achados.

## 5.3. IMPLICAÇÕES PRÁTICAS

Os resultados deste estudo têm implicações práticas diretas para o design de ferramentas de programação assistida por IA. Eles sugerem que a técnica HoD pode ser uma alternativa promissora para programadores que buscam ferramentas de XAI que minimizem a carga cognitiva, reduzam a ansiedade e promovam o bem-estar durante o processo de desenvolvimento de software. Os resultados também ressaltam a importância de se considerar fatores neurofisiológicos e comportamentais no design de interfaces e ferramentas de desenvolvimento, indo além das métricas tradicionais de desempenho e usabilidade. O notebook "Targeted Anxiety Intervention Analysis with Subgroup Discovery" aborda a questão da heterogeneidade da amostra, identificando subgrupos com diferentes padrões de resposta, o que pode informar o desenvolvimento de intervenções personalizadas.

## 5.4. RECONHECIMENTO DE LIMITAÇÕES

É fundamental reconhecer as limitações deste estudo. O tamanho da amostra (n=12) é relativamente pequeno, o que restringe a generalização dos resultados para a população geral de programadores. Estudos futuros devem buscar replicar esses achados com amostras maiores, mais diversificadas e representativas. A duração das tarefas de programação (40 minutos) pode não ser representativa de sessões de programação mais longas, que podem levar à fadiga e à exaustão

cognitiva. A natureza controlada do ambiente experimental, embora necessária para garantir a validade interna, pode não refletir totalmente a complexidade e a imprevisibilidade do ambiente de trabalho real dos programadores. Além disso, a utilização de dados sintéticos, embora tenha permitido um maior controle experimental, limita a validade ecológica dos resultados. O notebook "Robust Anxiety Intervention Analysis with Missing Data Handling and Mix of Experts Explainability" demonstra a importância de se lidar com dados faltantes, um problema comum em estudos com dados reais, e que pode ser explorado em futuras pesquisas.

#### 5.5. DIRECIONAMENTOS PARA PESQUISAS FUTURAS

Além das limitações já apontadas, como o tamanho da amostra e o uso de dados sintéticos, estudos futuros devem abordar as restrições dos instrumentos utilizados. O EEG Muse 2, com apenas 4 canais (FP1, FP2, T9, T10), limita a captura da atividade alfa posterior, essencial para avaliar a carga cognitiva, sugerindo a adoção de sistemas EEG de alta densidade (ex.: 64 ou 128 canais) com eletrodos parietais. O Samsung Watch5, projetado para ECG pontual, não suporta monitoramento contínuo necessário à análise da VFC. O que compromete um tanto a validade dos dados simulados para condutância da pele. O Tobii Eye Tracker 5 (0,3° de precisão, 33 Hz) é adequado, mas exige calibração precisa para refletir processos reais.

Pesquisas futuras podem explorar o uso de técnicas de neurofeedback em tempo real para melhorar a autoeficácia e reduzir a ansiedade dos programadores durante a interação com ferramentas de XAI. Estudos longitudinais podem avaliar os efeitos a longo prazo das diferentes técnicas de XAI sobre a saúde mental, o bem-estar e o desempenho dos programadores. A análise de redes cerebrais pode revelar insights adicionais sobre os mecanismos neurais subjacentes à interação entre programadores e ferramentas de IA, permitindo um entendimento mais profundo dos processos cognitivos envolvidos. A replicação deste estudo com dados reais, coletados em ambientes de trabalho autênticos, é fundamental para validar os achados e aumentar a sua generalização. O notebook "Anxiety Intervention Analysis with Causal Interaction Detection" explora a detecção de interações causais, um tópico que pode ser aprofundado em pesquisas futuras, investigando, por exemplo, como diferentes fatores contextuais podem moderar o efeito das técnicas de XAI. Além disso, o notebook "Analyzing Cognitive Load and Task Performance with Markdown Standardization" investiga o impacto da padronização da linguagem markdown na carga cognitiva, um exemplo de como fatores aparentemente sutis do ambiente de programação podem influenciar o desempenho e o bem-estar do programador, abrindo caminho para novas investigações sobre a otimização do ambiente de trabalho.

#### CONCLUSÃO

Este estudo investigou o impacto de diferentes técnicas de Inteligência Artificial Explicável (XAI) – Guardrails, RAG, RAFT, MoE e HoD – sobre a carga cognitiva, autoeficácia e ansiedade de programadores, utilizando uma abordagem crossmodal que combinou medidas neurofisiológicas (EEG, ECG, EDA) e comportamentais (AOI). Os resultados demonstraram consistentemente que a técnica HoD (Hypergraph of Drafts) se

sobressaiu em relação às demais, promovendo um estado cognitivo mais favorável, caracterizado por:

- Menor Carga Cognitiva: Evidenciada pela redução da atividade teta frontal no EEG e maior variabilidade da frequência cardíaca (RMSSD).
- Maior Autoeficácia: Refletida em maior atividade beta no EEG e padrões de fixação ocular mais estáveis.
- Menor Ansiedade: Indicada pela menor amplitude das respostas de condutância da pele (SCR).
- Maior Satisfação: Confirmada por menores pontuações no NASA-TLX (carga de trabalho percebida) e maiores pontuações no SCS (System Causability Scale).

Esses achados sugerem que a técnica HoD, que se baseia na modelagem de relações causais através de hipergrafos, oferece uma forma mais intuitiva e compreensível de apresentar explicações geradas por IA, resultando em menor esforço mental, maior confiança e menor estresse para os programadores.

As implicações deste estudo são significativas para o design de ferramentas de programação assistida por IA. Os resultados enfatizam a importância de se considerar os fatores neurofisiológicos e comportamentais, além das métricas tradicionais de desempenho e usabilidade, ao se projetar e avaliar ferramentas de XAI. A abordagem crossmodal adotada neste estudo, combinando diferentes fontes de dados, fornece um framework metodológico robusto para a avaliação objetiva do impacto da IA sobre a cognição e as emoções dos programadores.

Apesar das contribuições deste estudo, é importante reconhecer suas limitações, como o tamanho da amostra e o uso de dados sintéticos. Pesquisas futuras devem buscar replicar esses achados com amostras maiores e mais diversas, utilizando dados reais coletados em ambientes de trabalho autênticos. A investigação de outras técnicas de XAI e a exploração de diferentes paradigmas experimentais também são necessárias para um entendimento mais completo da interação entre programadores e ferramentas de IA. O notebook "Anxiety Intervention Analysis with Causal Interaction Detection" explora a detecção de interações causais, um tópico que merece maior investigação em estudos futuros, buscando identificar, por exemplo, como diferentes fatores contextuais podem moderar o efeito das técnicas de XAI sobre a carga cognitiva e a ansiedade.

Em última análise, este trabalho contribui para o campo emergente da neuroergonomia da programação, fornecendo insights valiosos sobre como a IA pode ser utilizada para aprimorar o processo de desenvolvimento de software, promovendo ambientes de trabalho mais saudáveis, eficientes e produtivos. A integração de medidas neurofisiológicas, comportamentais e subjetivas, juntamente com a análise de dados qualitativos (verbalizações), permitiu uma compreensão mais profunda e holística da experiência do programador ao interagir com diferentes técnicas de XAI. O notebook "Analyzing Cognitive Load and Task Performance with Markdown Standardization" ilustra como fatores aparentemente sutis, como a padronização da linguagem markdown, podem influenciar a carga cognitiva e o desempenho do programador, abrindo novas perspectivas para a otimização do ambiente de trabalho e das ferramentas de desenvolvimento.

# Neurophysiological Impact of XAI in Programming Tasks: A Crossmodal Assessment of Guardrails, RAG, RAFT, MoE, and HoD

## ABSTRACT

This experimental study investigates the neurophysiological effects of different Explainable Artificial Intelligence (XAI) techniques – Guardrails, Retrieval-Augmented Generation (RAG), Retrieval-Augmented Fine-Tuning (RAFT), Mixture of Experts (MoE), and Hypergraph of Drafts (HoD) – on programmers' cognitive load, self-efficacy, and anxiety. Using a crossmodal approach, we combine neurophysiological measures (electroencephalography - EEG, electrocardiography - ECG, electrodermal activity - EDA) and behavioral measures (eye tracking) to assess the impact of each technique during programming tasks. Twelve participants, with varying levels of experience, performed coding and debugging tasks while their physiological signals and behavior were monitored. The results reveal that the HoD technique stood out in promoting a more favorable cognitive state, with lower mental workload (indicated by reduced theta/alpha activity in EEG and higher heart rate variability in ECG), greater self-confidence (reflected in higher beta activity in EEG and more stable eye fixations), and lower anxiety (evidenced by lower skin conductance). The implications of this study point to the need to design AI tools that consider not only technical efficiency but also the programmer's cognitive well-being.

**KEYWORDS:** XAI. Neuroergonomics. Psychophysiology. Programming. Cognitive Load.

# Impacto Neurofisiológico de la XAI en Tareas de Programación: Una Evaluación Crossmodal de Guardrails, RAG, RAFT, MoE y HoD

## RESUMEN

Este estudio experimental investiga los efectos neurofisiológicos de diferentes técnicas de Inteligencia Artificial Explicable (XAI) – Guardrails, Retrieval-Augmented Generation (RAG), Retrieval-Augmented Fine-Tuning (RAFT), Mixture of Experts (MoE) e Hypergraph of Drafts (HoD) – sobre la carga cognitiva, la autoeficacia y la ansiedad de los programadores. Utilizando un enfoque crossmodal, combinamos medidas neurofisiológicas (electroencefalografía - EEG, electrocardiografía - ECG, actividad electrodérmica - EDA) y medidas conductuales (seguimiento ocular, registros de codificación) para evaluar el impacto de cada técnica durante las tareas de programación. Doce participantes, con diferentes niveles de experiencia, realizaron tareas de codificación y depuración mientras se monitorizaban sus señales fisiológicas y su comportamiento. Los resultados revelan que la técnica HoD destacó en la promoción de un estado cognitivo más favorable, con menor carga mental (indicada por una actividad theta/alfa reducida en el EEG y una mayor variabilidad de la frecuencia cardíaca en el ECG), mayor autoconfianza (reflejada en una mayor actividad beta en el EEG y fijaciones oculares más estables) y menor ansiedad (evidenciada por una menor conductancia de la piel). Las implicaciones de este estudio apuntan a la necesidad de diseñar herramientas de IA que consideren no solo la eficiencia técnica, sino también el bienestar cognitivo del programador.

**PALABRAS CLAVE:** XAI. Neuroergonomía. Psicofisiología. Programación. Carga Cognitiva.

## REFERÊNCIAS

- BANDURA, A. Teoria da mudança comportamental unificada através da autoeficácia. **Psychological Review**, 1977.
- CRK, I.; KLUTHE, C.; STEFIK, A. Investigação empírica de respostas fisiológicas fásicas durante a depuração de software. **ACM Transactions on Computer-Human Interaction**, 2015.
- EYSENCK, M. W.; DERAKSHAN, N.; SANTOS, R.; CALVO, M. G. Teoria do controle atencional da ansiedade e desempenho cognitivo. **Emotion**, 2007.
- HUMBLE, N.; MOZELIUS, P. Inteligência artificial para redução da carga cognitiva no ensino de programação. **Education and Information Technologies**, 2024.
- IDNAY, D.; COHEN, Y.; BERGMAN, L. Infraestrutura GenAI: explorando relações entre sinais neurofisiológicos e IA explicável em tarefas de geração de código. **Journal of Neural Engineering**, 2025.
- JAHAN, S.; ISLAM, M. S.; AKTHER, S.; RAHMAN, M. A. Desvendando a ansiedade computacional: insights de eye-tracking e análise de atividade eletrodérmica. **International Journal of Human-Computer Studies**, 2025.
- KLIMESCH, W. Oscilações alfa e teta do EEG refletem o desempenho cognitivo e de memória: uma revisão e análise. **Brain Research Reviews**, 1999.
- MARTINS, A.; GONCALVES, M.; ENCARNAÇÃO, P. Padrões de olhar em interfaces de IA explicáveis: um estudo da alocação de atenção e processos de tomada de decisão. **ACM Transactions on Computer-Human Interaction**, 2024.
- NYSTRÖM, M.; LINDBERG, T.; ANDERSSON, P. Análise de eye-tracking da alocação de atenção em interfaces de IA explicáveis: implicações para transparência e confiança. **Applied Ergonomics**, 2025.
- PAAS, F.; TUOVINEN, J. E.; TABBERS, H.; VAN GERVEN, P. W. M. Medição da carga cognitiva como meio de avançar na teoria da carga cognitiva. **Educational Psychologist**, 2003.
- RIEDEMANN, C.; WEGER, M.; BRANDSTATTER, V. O impacto de grandes modelos de linguagem no autoeficácia e estados cognitivos dos desenvolvedores: um estudo psicofisiológico. **Empirical Software Engineering**, 2024.
- SHNEIDERMAN, B. IA centrada no ser humano: confiável, segura e digna de confiança. **International Journal of Human-Computer Interaction**, 2020.
- SHE, W. J.; SIRIARAYA, P.; IWAKOSHI, H.; KUWAHARA, N.; SENOO, K. Uma aplicação de IA explicável (AF'ffective) para dar suporte ao monitoramento de pacientes com fibrilação atrial após ablação por cateter: grupo focal qualitativo, sessão de design e estudo de entrevista. **JMIR Human Factors**, v. 2, 2025.

SWELLER, J. Teoria da carga cognitiva e resolução de problemas. **Educational Psychology**, 1988.

TASK FORCE OF THE EUROPEAN SOCIETY OF CARDIOLOGY AND THE NORTH AMERICAN SOCIETY OF PACING AND ELECTROPHYSIOLOGY. Variabilidade da frequência cardíaca: padrões de medição, interpretação fisiológica e uso clínico. **Circulation**, 1996.

**Recebido:** 03 mar 2025

**Aprovado:** 20 abr. 2024

**DOI:** 10.3895/rtr.v10n0.20039

**Como Citar:** PESSOA JÚNIOR, H. C.; CASTIONI, R. Impacto Neurofisiológico de XAI em Tarefas de Programação: Uma Avaliação Crossmodal de Guardrails, RAG, RAFT, MoE e HoD. **Revista Transmutare**, Curitiba, v. 10, e20039, p. 1-18, 2025. Disponível em: <<https://periodicos.utfpr.edu.br/rtr>>. Acesso em: XXX.

**Correspondência:**

Hélio Craveiro Pessoa Júnior  
pessoajr@gmail.com

**Direito Autoral:** Este artigo está licenciado sob os termos da licença Creative Commons Atribuição 4.0 Internacional.

