AI SLOP E A CRISE DE CONFIANÇA DIGITAL: O CASO SAPPHOS E IMPLICAÇÕES PARA O DESIGN UX

Fernando Rizzaro de Almeida[1];

[1]Universidade Federal do Rio Grande do Sul (UFRGS), Porto Alegre, RS.

https://lattes.cnpq.br/2959653274126931

Victoria de Menezes Piffero[2];

[2]Universidade Federal do Rio Grande do Sul (UFRGS), Porto Alegre, RS.

https://lattes.cnpq.br/1801746511045984

Gabriele Sarmento da Silva[3];

[3]Universidade Federal do Rio Grande do Sul (UFRGS), Porto Alegre, RS.

https://lattes.cnpq.br/4705521996198251

Ralf Guenter Hennig[4];

[4]Universidade Federal do Rio Grande do Sul (UFRGS), Porto Alegre, RS.

https://lattes.cnpq.br/9852230535130165

Guilherme Souza Purri[5];

[5]Universidade Federal do Rio Grande do Sul (UFRGS), Porto Alegre, RS.

https://lattes.cnpq.br/2937182050702659

Ana Lorenzon[6];

[6]Universidade Federal do Rio Grande do Sul (UFRGS), Porto Alegre, RS.

Gilberto Balbela Consoni[7];

[7]Universidade Federal do Rio Grande do Sul (UFRGS), Porto Alegre, RS.

https://lattes.cnpq.br/2937182050702659

Vinicius Gadis Ribeiro[8].

[8]Universidade Federal do Rio Grande do Sul (UFRGS), Tramandaí, RS.

https://lattes.cnpq.br/2937182050702659

RESUMO: O rápido crescimento da Inteligência Artificial Generativa (IAG) introduziu o fenômeno do *"AI Slop"*, caracterizado pela proliferação de conteúdo de baixa qualidade e pela falta de rigor no uso de ferramentas de IA para geração de código. Este artigo examina como o *AI Slop* se manifesta no desenvolvimento de software e suas consequências para

a Experiência do Usuário (UX), exemplificado pelo caso do aplicativo de relacionamentos Sapphos. Projetado como um espaço seguro, o Sapphos saiu do ar em 2025, após três dias devido a uma falha crítica que expôs dados sensíveis dos usuários. A crise, atribuível ao conceito de *"vibe coding"* (desenvolvimento rápido e não testado), serve como um alerta: quando a saída da IA carece de qualidade e transparência, o resultado é a erosão da confiança e riscos severos de segurança. O design de experiência deve priorizar a mitigação dos riscos do *AI Slop*, focando na integridade, supervisão humana e segurança.

PALAVRAS-CHAVE: Inteligência Artificial. AI Slop. Design UX.

AI SLOP AND THE DIGITAL TRUST CRISIS: THE SAPPHOS CASE STUDY AND CRITICAL IMPLICATIONS FOR UX DESIGN

ABSTRACT: The rapid growth of Generative Artificial Intelligence (GAI) has introduced the phenomenon of "AI Slop," characterized by the proliferation of low-quality content and a lack of rigor in the use of AI tools for code generation. This article examines how AI Slop manifests itself in software development and its consequences for User Experience (UX), exemplified by the case of the dating app Sapphos. Designed as a secure space, Sapphos went offline in 2025 after three days due to a critical failure that exposed sensitive user data. The crisis, attributable to the concept of "vibe coding" (rapid and untested development), serves as a warning: when AI output lacks quality and transparency, the result is an erosion of trust and severe security risks. Experience design should prioritize mitigating the risks of AI Slop, focusing on integrity, human oversight, and security.

KEY-WORDS: Artificial Intelligence. AI Slop. UX Design.

INTRODUÇÃO

A rápida evolução e a ampla adoção da Inteligência Artificial Generativa (IAG) e dos Grandes Modelos de Linguagem (*LLMs*), como o GPT-2 (Radford et al., 2019) e o GPT-3.5 (Brown et al., 2020), revolucionaram a geração de textos, imagens e códigos. Contudo, essa eficiência sem precedentes trouxe consigo um desafio sistêmico significativo: o surgimento do *"AI Slop"*. Este fenômeno é definido pela criação extensiva de conteúdo de baixa qualidade, que carece do rigor intelectual tradicionalmente associado às produções geradas por humanos (Thompson et al., 2024).

As implicações dessa proliferação desregulada de conteúdo sintético estão muito além da simples saturação de informações, representando ameaças fundamentais à integridade dos dados, à segurança dos sistemas e à experiência do usuário (UX). O problema ganha proporções críticas quando o AI Slop se manifesta no processo de desenvolvimento de software, especialmente por meio de práticas emergentes como o *"vibe coding"* (Meske et

al., 2025), a criação rápida de aplicativos com o auxílio de IA, frequentemente sem revisões rigorosas de segurança ou testes exaustivos (Maiberg, 2025; Ropek, 2025).

O presente trabalho explora a etiologia desse fenômeno, desde a degradação da qualidade da IA por colapso de modelo e contaminação de dados (Shumailov et al., 2024; Georgopoulos et al., 2021) até as armadilhas cognitivas impostas aos desenvolvedores (Spitzer et al., 2024). Compreender essa dinâmica é fundamental para restabelecer a integridade nos processos de design e desenvolvimento digital.

OBJETIVO

Este estudo tem como objetivo principal analisar como o fenômeno do *"AI Slop"* e a prática do *"vibe coding"* impactam negativamente o desenvolvimento de software e a Experiência do Usuário (UX). Adicionalmente, busca-se investigar as consequências práticas e legais da negligência técnica e ética na adoção de códigos gerados por Inteligência Artificial, utilizando o colapso de segurança do aplicativo Sapphos (Camacho, 2025) como estudo de caso central para propor diretrizes de mitigação baseadas na supervisão humana e na transparência.

METODOLOGIA

A presente pesquisa classifica-se, quanto à sua natureza, como aplicada, quanto à abordagem, como qualitativa; e quanto aos seus objetivos, como exploratória e explicativa. Em relação aos procedimentos metodológicos, trata-se de um estudo de caso, fundamentado em Yin (2010), aliado a uma pesquisa bibliográfica.

O estudo de caso foca na catástrofe de segurança envolvendo o aplicativo de relacionamentos Sapphos, lançado e retirado do ar em setembro de 2025 no Brasil. A análise bibliográfica explora a literatura recente sobre Interação Humano-Computador (IHC), Inteligência Artificial Generativa, contaminação de dados (autofagia da IA) e vulnerabilidades de segurança cibernética (como a falha *IDOR* e ataques de *bad bots*) (Xiaomeng et al., 2018). A coleta de dados baseou-se na análise de artigos científicos, publicações de especialistas do setor de tecnologia e relatórios de segurança.

FUNDAMENTAÇÃO TEÓRICA

O fenômeno do *AI Slop* está intrinsecamente enraizado nos desafios relacionados à qualidade e à proveniência dos dados utilizados para treinar as sucessivas gerações de modelos de IA. O treinamento de IA em larga escala depende de conjuntos de dados massivos enviados às *APIs*. Uma *API* (Interface de Programação de Aplicações) pode ser definida, no âmbito da engenharia de software, como um conjunto padronizado de rotinas, protocolos e ferramentas que permite a comunicação e a integração entre diferentes

sistemas ou componentes de software.

De acordo com Sommerville (2018), uma *API* estabelece um contrato de interação que especifica as operações disponíveis, os parâmetros necessários e os formatos de dados esperados, abstraindo a complexidade interna do serviço. As *APIs* de *LLMs* utilizam *tokens* como a menor unidade de processamento, controle (janela de contexto, *max_tokens*), cobrança e medição de uso, convertendo automaticamente texto em tokens na entrada e tokens em texto na saída, por meio de um tokenizador determinístico associado a cada modelo (Vaswani et al.; 2017; Sennrich et al., 2016). Esse treinamento é frequentemente composto por centenas de bilhões de *tokens* extraídos da web (Sun et al., 2021; Chowdhery et al., 2023; Touvron et al., 2023; Almazrouei et al., 2023, apud Thompson et al., 2024).

Os *tokens* constituem as unidades fundamentais de texto empregadas por modelos de linguagem no processamento de informações, podendo representar palavras, segmentos subléxicos (fragmentos de palavras) ou até mesmo caracteres isolados (OpenAI, 2026). Conforme alertam Bender et al. (2021), a adoção de modelos de linguagem de grande escala sem controle adequado sobre os dados de treinamento e sem transparência pode gerar sistemas que repetem vieses, produzem informações falsas e amplificam danos sociais, um fenômeno que antecipa e agrava o quadro atual do *AI Slop*. O uso indiscriminado de dados gerados por modelos de IA para treinar gerações subsequentes cria defeitos irreversíveis nesses sistemas (Shumailov et al., 2024).

Esse processo de retroalimentação é denominado autofagia da IA (Alemohammad et al., 2023), que leva diretamente ao colapso do modelo, um processo degenerativo onde a IA perde gradualmente a capacidade de lembrar a verdadeira distribuição dos dados originais. Tal degradação afeta a diversidade e a qualidade do conteúdo, prejudicando especialmente a modelagem de eventos de baixa probabilidade e questões pertinentes a grupos marginalizados (Georgopoulos et al., 2021; Shumailov et al., 2024). Além da contaminação por dados sintéticos gerados por IA, a própria produção científica tem sido alvo de manipulações que deterioram a qualidade dos dados de treinamento. Heathers (2025) documentou casos em que pesquisadores inseriram mensagens em artigos com o objetivo de enganar ferramentas de IA utilizadas na revisão por pares, introduzindo ruído adicional e comprometendo a integridade dos corpora textuais que alimentam futuras gerações de *LLMs*.

A disseminação desregulada de dados sintéticos também causa decaimento linguístico (Thompson et al., 2024; Touvron et al., 2023). A presença massiva de traduções automáticas de baixa qualidade na web contamina o treinamento (Thompson et al., 2024), e estudos recentes documentam mudanças significativas no estilo da escrita acadêmica desde o lançamento do ChatGPT, com empobrecimento lexical e maior uniformidade textual (Bao et al., 2025). Da mesma forma, Kobak et al. (2025) demonstraram que a redação assistida por *LLMs* em artigos biomédicos introduz um excesso de vocabulário raro e padrões sintáticos não usuais, indicando que a influência da IA na produção textual científica já é mensurável

e potencialmente prejudicial à clareza e originalidade. Esse fenômeno resulta em *LLMs* menos fluentes e mais propensos a produzir alucinações, que são saídas de dados que são "falsidades plausíveis e excessivamente confiantes" (Kalai & Vempala, p.1, 2024).

Essa degradação abstrata ganha contornos práticos no desenvolvimento de software através do *"vibe coding"* (Maiberg, 2025; Ropek, 2025). Projetos baseados nessa abordagem são caracterizados por documentação mínima, ausência de testes abrangentes, lógica apressada e uma complexidade técnica que a própria equipe de desenvolvimento não compreende totalmente (Shumailov et al., 2024; Taleb, 2007). Frequentemente adotado por profissionais sem habilidades digitais profundas (como gerentes de produto), o *vibe coding* introduz vulnerabilidades severas (Sohni, apud Ropek, 2025; Maiberg, 2025).

Esses riscos surgem de dois fatores principais: a saída inerentemente vulnerável das ferramentas atuais de IA e a negligência ou excesso de confiança humana (Taeb et al., 2024). Os desenvolvedores frequentemente pulam revisões de segurança e não compreendem as implicações do código gerado, o que é desastroso quando o sistema lida com dados sensíveis de usuários (Maiberg, 2025; Camacho, 2025).

Além das vulnerabilidades técnicas, o *AI Slop* gera um "efeito de desinformação" (Spitzer et al., 2024). Pesquisas em IHC mostram que quando explicações incorretas acompanham os conselhos da IA, o conhecimento processual humano é prejudicado. Desenvolvedores que aceitam códigos gerados sem a devida revisão correm o risco de internalizar uma compreensão falha, reduzindo a eficiência e a capacidade de raciocínio autônomo (Maiberg, 2025; Dakhel et al., 2023, apud Spitzer et al., 2024). O uso indiscriminado de *LLMs* também causa a homogeneização do design e das funcionalidades. A prática do *vibe coding* frequentemente propaga padrões estéticos influenciados pela IA, reduzindo a inovação, a diversidade de componentes de interface e a genuinidade dos produtos digitais (Pimenova et al., 2025; Fawzy et al., 2025).

O aplicativo Sapphos foi concebido como um espaço seguro exclusivo para mulheres sáficas no Brasil (Liu, 2025; Camacho, 2025). Para garantir a segurança, o aplicativo exigia validação manual que incluía envio de e-mail, telefone, documentos pessoais e uma foto da usuária segurando a identidade oficial (RG) ao lado do rosto. Apesar dessa premissa, o Sapphos foi retirado do ar apenas três dias após o lançamento devido a uma vulnerabilidade crítica de *Insecure Direct Object Reference* (*IDOR*) (Camacho, 2025; Xiaomeng et al., 2018). A falha de *IDOR* é uma vulnerabilidade de controle de acesso que ocorre quando uma aplicação expõe referências internas a objetos (como chaves de banco de dados, números de identificação ou nomes de arquivos) sem verificar se o usuário autenticado está autorizado a acessá-los. Dessa forma, um agente malicioso pode modificar essas referências para visualizar ou manipular dados de outros usuários (Knivets, 2023). No caso Sapphos, Dados Altamente Sensíveis (PII – *Personally Identifiable Information*), incluindo as fotos segurando documentos, foram amplamente expostos, revelando a falta de criptografia básica e de cuidado com a transmissão de dados.

O problema técnico foi agravado pela má gestão da crise, evidenciando o que se define como *"regulatory slop"*. Em vez de reconhecer a falha apontada por um *ethical hacker*, a empresa atacou o usuário, erodindo completamente a confiança da comunidade. A ausência de medidas de segurança também configurou violação direta à Lei Geral de Proteção de Dados (LGPD) no Brasil, sujeitando os responsáveis a sanções severas (Camacho, 2025). Além disso, os Termos de Uso utilizavam modelos internacionais genéricos que continham cláusulas inválidas e irrenunciáveis perante o Código de Defesa do Consumidor brasileiro, demonstrando que a pressa no desenvolvimento sem expertise humana gera riscos sistêmicos (Maiberg, 2025; Camacho, 2025).

Vulnerabilidades geradas por *vibe coding* são um terreno fértil para *"bad bots"*, programas automatizados que realizam raspagem de dados (*scraping*) e ataques de negação de serviço. A raspagem de dados é um método para coletar dados não estruturados ou semi-estruturados de websites e convertê-los em um formato estruturado, permitindo análises quantitativas que seriam inviáveis manualmente. O processo consiste em utilizar programas (robôs ou *spiders*) que simulam a navegação humana, acessam o código-fonte (HTML, XML, JSON) da fonte alvo, localizam padrões ou elementos específicos (como tags, classes CSS ou XPath) e extraem os dados desejados, geralmente armazenando-os em formatos estruturados (CSV, JSON, bancos de dados) para análise posterior (Munzert et al., 2015; Mitchel, 2018).

A prática de *scraping* não é por si só ilegal, uma vez que se baseia no acesso público a informações disponíveis na web, sendo amplamente utilizada para fins legítimos como pesquisa acadêmica, monitoramento de preços e indexação por mecanismos de busca. No entanto, quando empregada por *bad bots*, a mesma ferramenta pode causar prejuízos significativos, incluindo negação de serviço (*DoS – Denial of Service*), violação de privacidade, concorrência desleal e comprometimento da segurança cibernética. Dessa forma, a legalidade e a eticidade do *scraping* dependem diretamente do contexto, do respeito a limites técnicos e legais e da intencionalidade do agente automatizado (Mitchel, 2018). Em 2024, o tráfego mundial de *bots* superou o humano, sendo que 44% desse tráfego era de *bad bots* avançados, que visavam a abusos de *API* (Imperva, 2025), exatamente como a falha *IDOR* no Sapphos. Dados extraídos dessas vulnerabilidades alimentam ataques de roubo de identidade digital. A escalada do débito técnico gerado por essas práticas forçou uma resposta do mercado: o surgimento dos *"vibe coding fixers"* ou especialistas em limpeza de código (Maiberg, 2025; Ropek, 2025). Profissionais experientes estão sendo contratados para estabilizar códigos frágeis gerados por IA, reestruturar arquiteturas, fechar brechas de segurança e arrumar interfaces (UI/UX) inconsistentes (Ulam Labs, 2023; Ropek, 2025; Maes, 2025). A existência dessa demanda prova empiricamente que códigos funcionais criados rapidamente por IA são insuficientes para garantir robustez.

Para combater a crise de confiança gerada pelo *AI Slop*, a UX e a engenharia de software devem integrar escrutínio técnico rigoroso (Spitzer et al., 2024). A IA não deve ser um sistema sem transparência, é necessário um aprofundamento cada vez maior das

pesquisas em Inteligência Artificial Explicável (*XAI – Explainable Artifical Intelligence*), que consiste em um campo interdisciplinar que visa tornar os modelos de aprendizado de máquina, tradicionalmente considerados caixas-pretas devido à sua opacidade interna, mais confiáveis e compreensíveis para os seres humanos. Conforme destacam Adadi e Berrada (2018), o objetivo central da *XAI* é permitir que sistemas baseados em IA expliquem suas decisões e predições, promovendo assim maior auditabilidade e responsabilização, especialmente em aplicações críticas como diagnóstico médico, justiça e finanças. Auditorias de segurança e testes contínuos em códigos gerados por IA devem ser obrigatórios; a IA deve ser vista apenas como uma ferramenta que necessita de verificação humana especializada (Maes, 2025; Meske et al., 2025). Paralelamente, plataformas devem desenvolver mecanismos para filtrar conteúdos sintéticos (*AIGT*) (Sun et al., 2024; Thompson et al., 2024), enquanto o design de IA Explicável (*XAI*) deve focar na precisão das explicações para não prejudicar o aprendizado cognitivo dos desenvolvedores em longo prazo (Spitzer et al., 2024).

CONSIDERAÇÕES FINAIS

O caso Sapphos atua como um conto de advertência definitivo sobre os perigos da proliferação do *AI Slop* no desenvolvimento de software. A rápida implementação, impulsionada pelo ethos do *vibe coding*, resultou em uma vulnerabilidade crítica que expôs uma comunidade vulnerável e destruiu a promessa central do aplicativo de ser um “espaço seguro”. Esse fracasso demonstra que a velocidade é insustentável quando desprovida de rigor técnico, jurídico e ético.

A ascensão da profissão de *“vibe coding cleanup specialist”* sublinha um imperativo fundamental para o design de UX moderno: o simples fato de um código funcionar não o torna adequado para produção. Aplicações reais precisam ser seguras, intuitivas, esteticamente consistentes e escaláveis. Conclui-se que assegurar que as experiências digitais sejam construídas sobre bases de confiança e segurança continua sendo uma responsabilidade inteiramente humana, dependente de diligência, habilidades especializadas e auditoria incansável. Pesquisas futuras devem se concentrar em aprimorar as métricas de avaliação de *LLMs*, valorizando a incerteza para reduzir alucinações e no desenvolvimento de mecanismos robustos de detecção de conteúdo sintético para frear o colapso dos modelos algorítmicos na internet.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001

REFERÊNCIAS

ADADI, A.; BERRADA, M. Peeking inside the black-box: a survey on explainable artificial intelligence (XAI). **IEEE Access**, New York, v. 6, p. 52138-52160, 17 set. 2018. DOI: 10.1109/ACCESS.2018.2870052. Disponível em: https://ieeexplore.ieee.org/document/8466590.

Acesso em: 12 maio 2026.

ALMAZROUEI, E. et al. **The falcon series of open language models.** 2023. Disponível em: https://arxiv.org/abs/2311.16867. Acesso em: 4 maio 2026.

BAO, T. et al. **Examining linguistic shifts in academic writing before and after the launch of ChatGPT. 2025.** Disponível em: https://arxiv.org/abs/2505.12218v1. Acesso em: 26 abril 2026.

BENDER, E. M. et al. On the dangers of stochastic parrots: can language models be too big? In: **Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency,** 2021. p. 610-623. Disponível em: https://doi.org/10.1145/3442188.3445922. Acesso em: 8 maio 2026.

BROWN, T. et al. **Language models are few-shot learners. Advances in Neural Information Processing Systems**, v. 33, p. 1877-1901, 2020. Disponível em: https://proceedings.neurips.cc/paper/2020/hash/1457c0d6bfcb4967418bfb8ac142f64a-Abstract.html. Acesso em: 19 maio 2026.

CAMACHO, A. **Sapphos: 'Tinder para mulheres' tem falha grave denunciada por usuário e sai do ar.** TecMundo, 2025. Disponível em: https://bit.ly/3KJi6X8. Acesso em: 28 abril 2026.

FAWZY, A.; TAHIR, A.; BLINCOE, K. **Vibe coding in practice: motivations, challenges, and a future outlook.** 2025. Disponível em: https://doi.org/10.48550/arXiv.2510.00328. Acesso em: 19 maio 2026.

GEORGOPOULOS, M. et al. **Mitigating demographic bias in facial datasets with style-based multi-attribute transfer.** International Journal of Computer Vision, v. 129, n. 7, p. 2288-2307, 2021. Disponível em: https://doi.org/10.1007/s11263-021-01448-w. Acesso em: 16 maio 2026.

HEATHERS, J. **Researchers have been sneaking secret messages into their papers in an effort to trick artificial intelligence (AI) tools into giving them a positive peer-review report.** Nature, 2025. Disponível em: https://doi.org/10.1038/d41586-025-02172-y. Acesso em: 14 maio 2026.

IMPERVA. **Bad Bot Report 2025: the rapid rise of bots and the unseen risk for business.** 2025. Disponível em: https://www.imperva.com/resources/resource-library/reports/2025-bad-bot-report/. Acesso em: 19 maio 2026.

KALAI, A. T. et al. **Why language models hallucinate.** 2025. Disponível em: https://arxiv.org/abs/2509.04359. Acesso em: 26 abril 2026.

KOBAK, D. et al. **Delving into LLM-assisted writing in biomedical publications through excess vocabulary.** Science Advances, 2025. Disponível em: https://arxiv.org/abs/2406.07016. Acesso em: 18 maio 2026.

KNIVETS, D. **Insecure direct object reference (IDOR)**. OWASP FOUNDATION. 2023.

Disponível em: https://owasp.org/www-community/attacks/insecure_direct_object_reference. Acesso em: 4 maio 2026.

LIU, B. **Mais que um 'Grindr lésbico': criadoras de novo app de relacionamento para mulheres sáficas no Brasil prometem revolucionar encontros.** Marie Claire, 2025. Disponível em: http://bit.ly/43cvl8Y. Acesso em: 5 maio 2026.

MAES, S. **The gotchas of AI coding and vibe coding: it's all about support and maintenance.** Center for Open Science, 2025. Disponível em: https://doi.org/10.31219/osf.io/kjz9t_v1. Acesso em: 13 maio 2026.

MAIBERG, E. **The software engineers paid to fix vibe coded messes**. 404 Media, 2025. Disponível em: https://www.404media.co/the-software-engineers-paid-to-fix-vibe-coded--messes/. Acesso em: 12 maio 2026.

MESKE, C. et al. **Vibe coding as a reconfiguration of intent mediation in software development.** 2025. Disponível em: https://doi.org/10.48550/arXiv.2507.21928. Acesso em: 19 maio 2026.

MITCHELL, R. **Web scraping with Python: collecting more data from the modern web.** 2. ed. Sebastopol: O'Reilly Media, 2018.

MUNZERT, S. et al. **Automated data collection with R: a practical guide to web scraping and text mining.** Chichester: Wiley, 2014.

OPENAI. **What are tokens and how to count them?** OpenAI Help Center, 2026. Disponível em: https://help.openai.com/en/articles/4936856-what-are-tokens-. Acesso em: 5 maio 2026.

PIMENOVA, V. et al. **Good vibrations? a qualitative study of co-creation, communication, flow, and trust in vibe coding.** 2025. Disponível em: https://doi.org/10.48550/arXiv.2509.12491. Acesso em: 12 maio 2026.

RADFORD, A. et al. **Language models are unsupervised multitask learners.** OpenAI blog, v. 1, n. 8, p. 9, 2019. Disponível em: https://cdn.openai.com/better-language-models/language_models_are_unsupervised_multitask_learners.pdf. Acesso em: 5 maio 2026.

ROPEK, L. **After AI led to layoffs, coders are being hired to fix 'vibe-coded' screwups.** Gizmodo, 2025. Disponível em: https://bit.ly/4h6eVVn. Acesso em: 19 maio 2026.

SENNRICH, R.; HADDOW, B.; BIRCH, A. Neural machine translation of rare words with subword units. In: **Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (ACL)**, Berlin, Germany, 2016. Stroudsburg: ACL, 2016. p. 1715-1725. DOI: 10.18653/v1/P16-1162. Disponível em: https://aclanthology.org/P16-1162/. Acesso em: 30 abril 2026.

SHUMAILOV, I. et al. **AI models collapse when trained on recursively generated data.** Nature, v. 631, n. 8022, p. 755-759, 2024. Disponível em: https://doi.org/10.1038/s41586-024-07566-y. Acesso em: 15 maio 2026.

SPITZER, P. et al. **Misinformation effect of explanations: incorrect explanations in human-AI collaboration impair procedural knowledge and reasoning.** 2024. Disponível em: https://arxiv.org/abs/2409.12809v2. Acesso em: 19 maio 2026.

SUN, Z. et al. **Are we in the AI-generated text world already? quantifying and monitoring AIGT on social media.** 2024. Disponível em: https://arxiv.org/abs/2412.18148v3. Acesso em: 23 abril 2026.

TAEB, M.; BERNADIN, S.; CHI, H. **Assessing the effectiveness and security implications of AI code generators.** Journal of the Colloquium for Information Systems Security Education, v. 11, n. 1, p. 6, 2024. Disponível em: https://doi.org/10.53735/cisse.v11i1.180. Acesso em: 15 maio 2026.

THOMPSON, B. et al. A shocking amount of the web is machine translated: insights from multi-way parallelism. In: **Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics**, 2024. Disponível em: https://arxiv.org/abs/2401.05749. Acesso em: 25 abril 2026.

ULAM LABS. **We clean up after vibe coding. Literally.** 2023. Disponível em: https://www.ulam.io/software-services/we-clean-up-after-vibe-coding. Acesso em: 10 maio 2026.

VASWANI, A. et al. Attention is all you need. In: **Advances in Neural Information Processing Systems (NeurIPS)**, Long Beach, CA, USA, 2017. Proceedings [...]. Red Hook: Curran Associates, 2017. p. 5998-6008. Disponível em: https://proceedings.neurips.cc/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf. Acesso em: 8 maio 2026.

XIAOMENG, W. et al. **CPGVA: code property graph based vulnerability analysis by deep learning.** IEEE, 2018. p. 184-188. Disponível em: https://doi.org/10.1109/icait.2018.8686548. Acesso em: 10 maio 2026.

YIN, R. K. **Estudo de caso: planejamento e métodos**. 4. ed. Porto Alegre: Bookman, 2010.