ENGENHARIA REVERSA E IA GENERATIVA: RECONSTRUINDO O CÓDIGO PARA ENTENDER O INIMIGO

REVERSE ENGINEERING AND GENERATIVE AI: REBUILDING THE CODE TO UNDERSTAND THE ENEMY

Tadeu Marcos Borges Paes[1]

Francisco Nicolás Isnardi Begot[2]

Resumo: A ascensão da inteligência artificial generativa transformou o modo como sistemas computacionais são concebidos, testados e compreendidos. No campo da cibersegurança, essa revolução reconfigura o papel da engenharia reversa, que deixa de ser apenas um processo de decodificação técnica para se tornar uma prática cognitiva de reconstrução inteligente. Este artigo analisa o uso teórico e investigativo da engenharia reversa associada à IA generativa como instrumento de análise comportamental e de aprendizado adversarial — não apenas para desmontar códigos maliciosos, mas para compreender as lógicas que os produzem. A partir de uma abordagem interdisciplinar, discute-se como modelos generativos (como redes neurais profundas e agentes autônomos) podem reproduzir e simular ameaças, permitindo que pesquisadores e analistas explorem o funcionamento interno de softwares e algoritmos sob uma perspectiva epistemológica. O estudo propõe ainda um modelo conceitual que integra engenharia, cognição e ética digital, alinhado às demandas da era da informação e às diretrizes contemporâneas da ciberdefesa. Os resultados indicam que a integração entre engenharia reversa e inteligência artificial generativa representa um novo paradigma de investigação, capaz de unir reconstrução técnica e reflexão ética na compreensão e enfrentamento de

1 Doutorando em Inteligência Artificial, Senai Centro Desenvolvimento da Amazônia, https://orcid.org/0009-0002-2978-2117

2 Discente de Engenharia Computação, Universidade Federal do Pará, https://orcid.org/0009-0009-0944-6336

ameaças digitais.

Palavras-chave: Engenharia reversa. Inteligência artificial generativa. Cibersegurança. Aprendizado adversarial. Ética computacional.

Abstract: The rise of generative artificial intelligence has transformed how computer systems are designed, tested, and understood. In the field of cybersecurity, this revolution reconfigures the role of reverse engineering, which ceases to be merely a technical decoding process and becomes a cognitive practice of intelligent reconstruction. This article analyzes the theoretical and investigative use of reverse engineering associated with generative AI as an instrument for behavioral analysis and adversarial learning—not only to dismantle malicious code but also to understand the logics that produce it. From an interdisciplinary approach, it discusses how generative models (such as deep neural networks and autonomous agents) can reproduce and simulate threats, allowing researchers and analysts to explore the inner workings of software and algorithms from an epistemological perspective. The study also proposes a conceptual model that integrates engineering, cognition, and digital ethics, aligned with the demands of the information age and contemporary cyber defense guidelines. The results indicate that the integration between reverse engineering and generative artificial intelligence represents a new research paradigm, capable of uniting technical reconstruction and ethical reflection in understanding and addressing digital threats.

Keywords: Reverse engineering. Generative artificial intelligence. Cybersecurity. Adversarial learning. Computational ethics.

INTRODUÇÃO

A inteligência artificial (IA) generativa inaugura uma nova fronteira no campo da engenharia

reversa, redefinindo não apenas as práticas técnicas de análise de software, mas também os paradigmas epistemológicos da cibersegurança. Se antes compreender o inimigo significava descompilar um código, hoje exige reconstruir o raciocínio que o originou. A engenharia reversa, tradicionalmente voltada à desmontagem de binários e sistemas, passa a dialogar com redes neurais, modelos de linguagem e algoritmos capazes de aprender e simular o comportamento de agentes maliciosos (GOODFELLOW; BENGIO; COURVILLE, 2016).

Nesse novo cenário, a IA generativa não é apenas uma ferramenta — ela se torna um espelho cognitivo da criação e da destruição digital. Sistemas como GPT, Gemini e Copilot demonstram que a capacidade de gerar código, texto e imagens é inseparável da capacidade de compreender padrões e intenções, o que transforma a própria noção de autoria, defesa e ataque cibernético. Como observa Harari (2023), "as máquinas deixaram de executar ordens para começar a interpretar contextos", abrindo caminho para um ciclo em que a fronteira entre criador e invasor se torna cada vez mais tênue.

A integração entre engenharia reversa e IA generativa amplia o alcance da análise comportamental de ameaças, permitindo que especialistas explorem não apenas o que o código faz, mas por que ele faz. Por meio de modelos generativos, é possível reconstruir fluxos lógicos, prever mutações de malware e até antecipar vulnerabilidades em tempo real. Essa prática aproxima a engenharia reversa da filosofia da aprendizagem adversarial, conceito segundo o qual compreender o adversário requer imitá-lo, simulá-lo e aprender com seus padrões (GOODFELLOW et al., 2015; MITRE, 2023).

Contudo, essa simbiose entre engenheiros humanos e máquinas inteligentes também introduz dilemas éticos e epistemológicos profundos. Se a IA generativa é capaz de reconstruir o inimigo, ela também pode recriá-lo, intencionalmente ou não. A fronteira entre pesquisa legítima e uso malicioso se torna nebulosa, exigindo que o processo de engenharia reversa seja acompanhado por um marco ético e regulatório robusto, capaz de equilibrar segurança, privacidade e inovação (PAES, 2025).

Este artigo busca analisar o papel investigativo da engenharia reversa associada à IA

generativa como uma nova metodologia de compreensão e enfrentamento da cibercriminalidade. A partir de uma abordagem teórico-investigativa, discute-se como a reconstrução do código, mediada por sistemas generativos, pode revelar dimensões cognitivas e éticas do comportamento digital. O objetivo é propor um modelo conceitual que una engenharia, inteligência e ética, demonstrando que, para entender o inimigo digital, é preciso mais do que decifrar seu código, é necessário aprender a pensar como ele.

A partir dessa problemática e do objetivo delineado, a metodologia adotada neste estudo busca articular fundamentos conceituais e epistemológicos que sustentam o modelo teórico proposto, combinando revisão bibliográfica, análise comparativa e construção reflexiva.

METODOLOGIA

Este estudo adota uma abordagem qualitativa e teórico-investigativa, voltada para compreender as inter-relações entre a engenharia reversa, a inteligência artificial generativa e a análise comportamental de ameaças digitais. O objetivo metodológico não é mensurar fenômenos técnicos, mas interpretar significados, princípios e implicações éticas do uso da IA generativa como instrumento de reconstrução e compreensão de códigos maliciosos (GIL, 2019).

A pesquisa foi conduzida em três eixos complementares:

- Na primeira, realizou-se uma análise bibliográfica e documental, baseada em obras de referência sobre inteligência artificial, cibersegurança e filosofia da técnica, incluindo autores como Goodfellow, Morin, Castells, Lévy e Paes, além de relatórios técnicos do MITRE, ENISA e OWASP.
- Em seguida, procedeu-se a uma análise conceitual comparativa, que buscou examinar como a engenharia reversa e a IA generativa convergem em práticas cognitivas e éticas, observando frameworks de adversarial learning e machine learning explainability.
- Por fim, foi realizada a construção teórica do modelo investigativo, que propõe a

integração entre desmontagem técnica e reconstrução simbólica, originando o conceito de Engenharia Reversa Cognitiva (ERC).

De acordo com Lakatos e Marconi (2020), a pesquisa teórica "busca sistematizar e ampliar conhecimentos já existentes, por meio de uma análise crítica e integradora". Assim, a presente investigação combina revisão sistemática e reflexão epistemológica, procurando revelar as dimensões simbólicas, cognitivas e éticas envolvidas no processo de engenharia reversa mediado por inteligência artificial. Essa articulação teórica sustenta as categorias de análise descritas a seguir, que estruturam a interpretação dos resultados conceituais.

A interpretação dos dados conceituais foi realizada por meio de análise categorial temática e discursiva, segundo os princípios de Bardin (2016), o que permitiu identificar três categorias centrais:

- (a)Engenharia cognitiva – a capacidade da IA generativa de compreender padrões e reconstruir intenções a partir de fragmentos de código;
- (b)Aprendizado adversarial – o uso da simulação generativa como forma de estudar comportamentos maliciosos sem replicá-los de modo destrutivo;
- (c) Ética algorítmica – a reflexão sobre os limites morais da inteligência artificial ao interagir com sistemas vulneráveis.

Como limitação, destaca-se que o estudo é de natureza conceitual e não contempla experimentação empírica direta com ferramentas de engenharia reversa ou IA generativa. Contudo, as inferências teóricas e analíticas aqui desenvolvidas oferecem base epistemológica e metodológica para pesquisas aplicadas futuras, que poderão explorar modelos híbridos de ciberdefesa baseados em IA generativa.

Metodologia de Análise Conceitual

A metodologia analítica foi estruturada em três fases. Na primeira, realizou-se uma revisão narrativa da literatura científica, utilizando bases como Scopus, IEEE Xplore, SpringerLink e SciELO, com recorte temporal de 2015 a 2025. Na segunda fase, procedeu-se à categorização semântica e conceitual, associando os termos "engenharia reversa", "IA generativa", "aprendizado adversarial" e "ética digital" às suas correspondências teóricas. Por fim, na terceira fase, foi elaborado o modelo conceitual de reconstrução cognitiva, que busca representar como a IA generativa pode auxiliar o processo de entendimento do comportamento adversário por meio da reconstrução simbólica de código.

Essa abordagem metodológica visa consolidar um novo paradigma de investigação — a engenharia reversa cognitiva, na qual a análise técnica e a reflexão ética se entrelaçam para compreender a lógica do inimigo sem reproduzir sua destruição.

RESULTADOS

A análise teórica realizada permitiu identificar que a integração entre engenharia reversa e inteligência artificial generativa dá origem a um novo paradigma cognitivo no campo da cibersegurança. Esse paradigma, denominado Engenharia Reversa Cognitiva (ERC), propõe que compreender o inimigo digital não se limita a desmontar o código, mas a reconstruir sua intenção, analisando os padrões simbólicos, comportamentais e éticos embutidos em cada linha de programação.

Os resultados conceituais demonstram que a IA generativa, ao ser aplicada de forma controlada e ética, pode simular ambientes de ataque, prever mutações de malware e criar representações sintéticas de comportamento adversário. Essa capacidade de reconstrução simbólica amplia a noção de engenharia reversa tradicional, que passa de um processo técnico-linear para um ciclo iterativo e reflexivo, em que a máquina aprende com o inimigo e o pesquisador aprende com a máquina.

Segundo Morin (2015), “compreender um sistema complexo é reconstruir o seu modo de organização”, e é justamente nesse sentido que a ERC opera: reconstrói a lógica do código para entender o pensamento do invasor.

Essa visão dialoga com a pesquisa de Mendonça (2023), que ao implementar um dashboard baseado no CVE MITRE para análise e correções de vulnerabilidades, demonstrou como a integração entre sistemas automatizados e análise humana pode revelar padrões ocultos de comportamento digital e fortalecer a capacidade investigativa em segurança da informação. O estudo de Mendonça evidencia que a visualização e a modelagem de dados não apenas aprimoram a eficiência técnica, mas também favorecem a compreensão cognitiva dos processos adversariais — princípio essencial da Engenharia Reversa Cognitiva.

De modo complementar, Paes (2025) propõe que “a verdadeira revolução tecnológica não está nas máquinas que aprendem, mas nas pessoas que aprendem a aprender com as máquinas”, reforçando a ideia de que o processo de reconstrução digital deve ser também um exercício de autoconhecimento humano e ético.

A IA generativa, ao modelar cenários e gerar variantes de ameaças, torna-se não apenas uma ferramenta de defesa, mas um agente cognitivo de análise.

Modelo Conceitual de Engenharia Reversa Cognitiva (ERC)

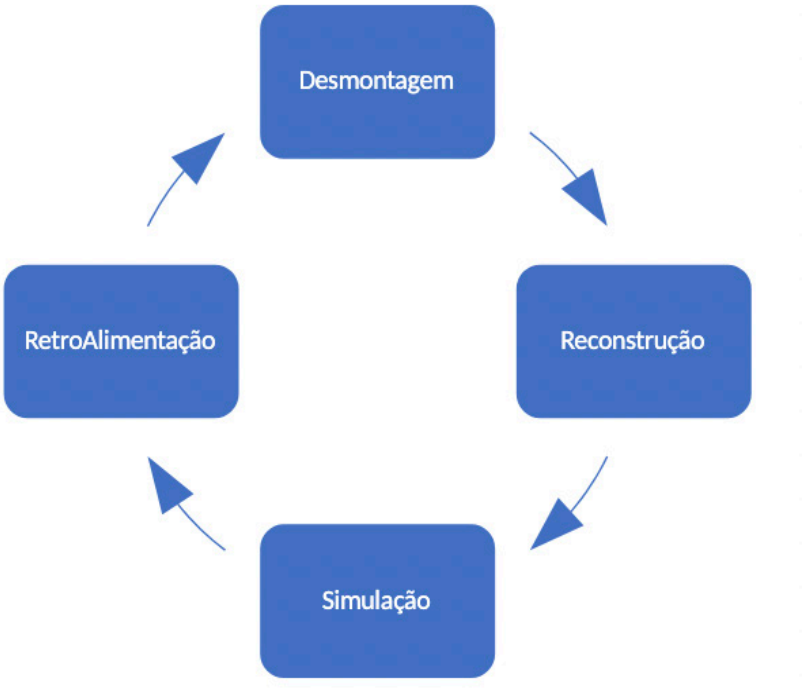
A análise dos resultados teóricos revela que o processo de compreender sistemas digitais maliciosos exige mais do que o simples domínio técnico: requer a capacidade de reconstruir cognitivamente a lógica que sustenta o comportamento do código. Essa reconstrução envolve observar o software como um organismo complexo — com padrões, decisões e intenções embutidas — cuja compreensão depende da combinação entre análise técnica e empatia cognitiva. Nesse sentido, a engenharia reversa tradicional evolui para uma prática de reflexão investigativa, em que cada linha de código é interpretada como expressão de um raciocínio criativo e adaptativo.

A integração da IA generativa nesse contexto transforma a investigação em um processo cooperativo entre mente humana e máquina. Enquanto o analista busca padrões, a IA sugere hipóteses, cria simulações e antecipa comportamentos, formando um ciclo de aprendizado mútuo. Essa dinâmica inaugura uma nova epistemologia da ciberdefesa, em que o objetivo não é apenas descompilar o código, mas compreender as intenções que o geraram. Assim, surge o conceito de Engenharia Reversa Cognitiva (ERC) — uma abordagem que alia raciocínio técnico, cognição simbólica e ética computacional na tentativa de reconstruir não apenas sistemas, mas o próprio pensamento que os concebeu.

A partir dessa integração entre técnica e cognição, propõe-se o Modelo Conceitual de Engenharia Reversa Cognitiva (ERC), estruturado em cinco estágios interdependentes (Figura 1):

- Desmontagem Analítica – extração e descompilação de códigos ou artefatos digitais, com foco na estrutura e no comportamento do sistema;
- Reconstrução Generativa – utilização de modelos de IA (transformers, autoencoders, GANs) para recriar o funcionamento lógico do software ou ameaça analisada;
- Simulação Adversarial – aplicação de algoritmos generativos para reproduzir variações comportamentais e identificar vulnerabilidades emergentes;
- Reflexão Ético-Cognitiva – interpretação das implicações morais e sociais dos resultados gerados pela IA durante o processo de reconstrução;
- Retroalimentação e Aprendizado – integração das descobertas ao ciclo de segurança e pesquisa, formando um sistema dinâmico de aprendizado entre humanos e máquinas.

Figura 1 – Modelo Conceitual de Engenharia Reversa Cognitiva (ERC)



Fonte: Elaboração própria (2025).

Essas etapas compõem um processo cíclico e autorreflexivo. Diferente da engenharia reversa tradicional, que busca restaurar o estado original de um sistema, a ERC busca compreender sua lógica evolutiva — o porquê de sua construção e a intenção de seus algoritmos. Essa visão aproxima-se do que Paes (2024) denomina "inteligência ética aplicada", na qual o ato de investigar um sistema computacional se torna também um exercício de consciência digital.

Os resultados indicam, ainda, que o uso da IA generativa nesse contexto não substitui o analista humano, mas o amplifica cognitivamente. O pesquisador passa a atuar como mediador entre a máquina e a interpretação dos dados, reforçando a noção de simbiose entre inteligência humana e artificial. Esse modelo oferece uma base para futuras aplicações em ciberdefesa, auditoria de algoritmos, ensino de segurança digital e investigação computacional forense.

DISCUSSÃO

Os resultados obtidos indicam a Engenharia Reversa Cognitiva (ERC) redefine o papel da análise de sistemas na era da inteligência artificial generativa. Se antes compreender o inimigo digital significava desmontar linhas de código, hoje significa reconstruir o raciocínio e a intenção que o originaram. Essa transição de um enfoque técnico para um enfoque cognitivo revela a maturidade do campo da ciberdefesa: compreender o código é, em essência, compreender a mente que o concebeu.

De acordo com Morin (2015), compreender um sistema complexo implica reconstruir o seu modo de organização — e essa reconstrução não é apenas técnica, mas epistemológica. A ERC opera nesse sentido: reconstrói o pensamento embutido no código e, ao fazê-lo, espelha o processo de criação humana. Ao empregar a IA generativa nesse contexto, o pesquisador cria um diálogo entre duas inteligências — a humana e a artificial — em que a compreensão emerge da cooperação entre ambas.

Essa lógica de reconstrução ganha respaldo empírico no trabalho de Mendonça (2023), que, ao implementar um dashboard baseado no CVE MITRE, demonstrou que a reconstrução de dados e vulnerabilidades é também uma forma de aprendizado sobre o próprio sistema que os produz. Em sua dissertação, o autor argumenta que "a integração entre análise visual e frameworks MITRE permite prever e compreender o ciclo de vida das ameaças" (MENDONÇA, 2023, p. 57), antecipando, em prática, os princípios teóricos que fundamentam a Engenharia Reversa Cognitiva.

A presença da IA generativa nesse processo amplia a dimensão ética e reflexiva da investigação. Como observa Harari (2023), vivemos a era em que "os algoritmos deixaram de executar ordens para interpretar contextos", o que transforma a relação entre criador e criatura. Nessa perspectiva, o pesquisador que usa IA para compreender o inimigo também precisa compreender a si mesmo — pois a tecnologia reproduz, potencializa e, em certos casos, questiona a própria moral humana. Paes (2025) reforça essa visão ao afirmar que "a singularidade tecnológica educacional inaugura uma era em que a inteligência artificial não é apenas ferramenta, mas consciência colaborativa", destacando a

necessidade de responsabilidade ética e epistemológica no uso da IA generativa.

Assim, a Engenharia Reversa Cognitiva (ERC) assume papel duplo: instrumento técnico e filosofia de investigação. A desmontagem, a reconstrução e a simulação tornam-se atos cognitivos e éticos que reconfiguram a fronteira entre o humano e o algorítmico. Em síntese, compreender o inimigo digital é compreender o reflexo da própria inteligência humana — um exercício de autoconhecimento tecnológico que projeta a IA não como adversária, mas como espelho da consciência criadora.

Em síntese, a discussão evidencia que compreender o inimigo digital é compreender o espelho da própria inteligência humana. A engenharia reversa cognitiva é, antes de tudo, um processo de reconstrução da mente por meio da máquina — um exercício ético, técnico e filosófico que redefine o papel do pesquisador na era da inteligência artificial generativa.

CONCLUSÃO

A convergência entre engenharia reversa e inteligência artificial generativa inaugura uma nova fronteira para a cibersegurança e para a epistemologia da tecnologia. A análise teórica desenvolvida neste estudo demonstrou que compreender o inimigo digital não se resume à desmontagem de códigos, mas envolve reconstruir as intenções e os padrões cognitivos que o originam. Esse deslocamento do foco técnico para o cognitivo e ético dá forma ao conceito de Engenharia Reversa Cognitiva (ERC) — uma abordagem que une raciocínio analítico, aprendizagem de máquina e reflexão humanista.

A aplicação da IA generativa nesse contexto amplia a capacidade de observação e simulação de ameaças, possibilitando a antecipação de comportamentos e vulnerabilidades. Assim, o processo investigativo passa a ser dinâmico, reflexivo e colaborativo, em que a máquina não apenas executa, mas também aprende e propõe, transformando o analista em mediador entre duas inteligências. A ERC surge, portanto, como um modelo de compreensão contínua, em que cada reconstrução de código representa um novo ciclo de aprendizado e autoconhecimento tecnológico.

A contribuição prática da pesquisa de Mendonça (2023) confirma que a reconstrução de

vulnerabilidades e o uso de frameworks MITRE podem ser vistos como etapas concretas dessa filosofia investigativa, enquanto os aportes teóricos de Morin (2015), Harari (2023) e Paes (2025) reforçam que a tecnologia só alcança maturidade quando é acompanhada de consciência ética. Desse modo, compreender o inimigo digital significa, também, compreender o reflexo de nossas próprias inteligências — humanas e artificiais — em constante diálogo.

Conclui-se que a Engenharia Reversa Cognitiva representa mais do que uma metodologia: trata-se de um novo paradigma de pensamento investigativo, em que ética, cognição e técnica se entrelaçam na busca por segurança e verdade digital. Para pesquisas futuras, recomenda-se a expansão empírica do modelo em ambientes de simulação real, explorando o uso de IA explicável, auditoria algorítmica e formação ética em ciberdefesa.

Por fim, reafirma-se que reconstruir o código é também reconstruir a consciência. Entender o inimigo é compreender os limites e as possibilidades da própria mente humana — e, nesse espelho digital, reconhecer que cada linha de código é, antes de tudo, uma linha de pensamento.

REFERÊNCIAS

BARDIN, Laurence. Análise de conteúdo. Lisboa: Edições 70, 2016.

CASTELLS, Manuel. Sociedade em rede. 19. ed. São Paulo: Paz e Terra, 2020.

GIL, Antonio Carlos. Métodos e técnicas de pesquisa social. 7. ed. São Paulo: Atlas, 2019.

GOODFELLOW, Ian; BENGIO, Yoshua; COURVILLE, Aaron. Deep Learning. Cambridge: MIT Press, 2016.

GOODFELLOW, Ian et al. Explaining and Harnessing Adversarial Examples. In: International Conference on Learning Representations (ICLR). San Diego: ICLR, 2015.

HARARI, Yuval Noah. Inteligência Artificial e o Futuro da Humanidade. Londres: Harvill Secker, 2023.

LAKATOS, Eva Maria; MARCONI, Marina de Andrade. Fundamentos de metodologia científica. 9. ed. São Paulo: Atlas, 2020.

LÉVY, Pierre. Cibercultura. 4. ed. São Paulo: Editora 34, 2019.

MENDONÇA, Eudes Danilo da Silva. Implementação de um Dashboard baseado no CVE MITRE para análise e correções de vulnerabilidades. Dissertação (Mestrado em Computação Aplicada) – Universidade Federal do Pará, Tucuruí, 2023.

MITRE. ATT&CK® Framework – Adversarial Tactics, Techniques, and Common Knowledge. McLean: The MITRE Corporation, 2023. Disponível em: https://attack.mitre.org/. Acesso em: 10 nov. 2025.

MORIN, Edgar. O método 6: ética. Porto Alegre: Sulina, 2015.

PAES, Tadeu Marcos Borges. Singularidade Tecnológica Educacional: O Salto Temporal – Uma Revolução Inevitável. Belém: AlphaNexus Academy, 2025.

SIEMENS, George. Conectivismo: uma teoria de aprendizagem para a era digital. Porto Alegre: Penso, 2021.