

Estado da publicação: O preprint não foi publicado em outro meio.

Análise do discurso, linguística forense e iag: autoria e cena de enunciação em ciberataques - phishing no caso eviltokens

Rosangela Carreira, Iwens Sene Junior

<https://doi.org/10.1590/SciELOPreprints.18633>

Submetido em: 2026-10-28

Postado em: 2026-10-30 (versão 1)

(AAAA-MM-DD)

ANÁLISE DO DISCURSO, LINGUÍSTICA FORENSE E IAG: AUTORIA E CENA DE ENUNCIÇÃO EM CIBERATAQUES - PHISHING NO CASO EVILTOKENS

DISCOURSE ANALYSIS, FORENSIC LINGUISTICS AND IAG: AUTHORSHIP AND ENUNCIATION SCENE IN CYBERATTACKS - PHISHING IN THE EVILTOKENS CASE

Tipo de Contribuição – Estudo Piloto

Rosângela Aparecida Ribeiro Carreira¹
Iwens Gervásio Sene Junior²

RESUMO: A utilização da Inteligência Artificial Generativa (IAG) tem provocado mudanças nas formas de produção, apropriação e circulação de enunciados em ambientes digitais, trazendo questões importantes para a compreensão da autoria e da construção de identidades discursivas. No campo dos ciberataques, sistemas de geração textual podem favorecer práticas de engenharia social, por meio da produção de enunciados que simulam sujeitos, instituições e situações comunicativas, criando efeitos de autenticidade e legitimidade. Este artigo objetiva discutir as aproximações possíveis entre Análise do Discurso (AD) e Linguística Forense (LF), considerando a IAG como ferramenta de práticas discursivas relacionadas à simulação identitária e à produção de mensagens potencialmente fraudulentas. Para isso, partimos das contribuições de Maingueneau (2005 a 2024), Foucault (2009), Coulthard e Johnson (2007), Sousa-Silva e Coulthard (2016), Schmitt e Flechais (2024), Sousa-Silva (2026), Carreira e Nascimento(2020), entre outros, para compreender as relações entre autoria, *ethos*, cenas de enunciação e produção de evidências linguísticas. Trata-se de uma pesquisa qualitativa, de cunho interpretativo, cujo *corpus* é constituído por dois recortes enunciados referentes a situações de engenharia social do caso *Eviltoken* analisado pela equipe de segurança da *Microsoft* em seu *blog* em 22 de setembro de 2026. A análise demonstra que a construção de simulacros de identidades institucionais pode produzir efeitos de legitimidade por meio da

¹ Rosângela Aparecida Ribeiro Carreira. Doutora em Letras, professora adjunta da Universidade Federal de Goiás (UFG), vinculada ao Departamento de Linguística e Língua Portuguesa, Goiânia, Goiás, Brasil. Docente permanente do Programa de Pós-Graduação em Letras e Linguística (PPGLL/UFG). Coordenadora do Laboratório de Linguagem Jurídica, Comportamentos e Tecnologias Sociais (LABJUS/UFG). E-mail: rosangela.carreira@ufg.br. ORCID: <https://orcid.org/0000-0001-7167-7666>. Contribuições CRediT: Conceituação da parte de AD e LF; Metodologia da parte de Linguística, escolha e seleção do *corpus*; Investigação de um caso real; Análise formal; Escrita das seções relativas à Linguística – rascunho original; Escrita – revisão e edição; Supervisão.

² Iwens Gervasio Sene Junior é professor titular do Instituto de Informática da Universidade Federal de Goiás (UFG), Brasil, e pesquisador do Centro de Excelência em Inteligência Artificial (CEIA/UFG). É Doutor em Engenharia Elétrica pela Universidade de Brasília (UnB) e realizou estágio de pós-doutorado na Universidade de Lancaster, no Reino Unido. E-mail: iwens@ufg.br. ORCID: <https://orcid.org/0000-0001-7605-0205>. Contribuições CRediT: Conceituação da parte de Ciências da Computação; Escrita da seção 2 relativa à Ciência da Computação – rascunho original e leitura final.

apropriação de formas de dizer socialmente reconhecidas. Concluimos que a interface entre AD e LF possibilita ampliar a compreensão das relações entre enunciação, identidade e autoria, especialmente quando a linguagem integra práticas sociais que podem apresentar consequências jurídicas.

PALAVRAS-CHAVE: Análise do Discurso. Linguística Forense. Inteligência Artificial Generativa. Autoria. Ciberataques.

ABSTRACT: The use of Generative Artificial Intelligence (GAI) has been causing changes in the ways statements are produced, adopted, and circulated in digital environments, raising important questions for understanding authorship and the construction of discursive identities. In the field of cyberattacks, text generation systems can facilitate social engineering practices by producing statements that simulate subjects, institutions, and communicative situations, creating effects of authenticity and legitimacy. This article aims to discuss the possible connections between Discourse Analysis (DA) and Forensic Linguistics (FL), considering GAI as a tool for discursive practices related to identity simulation and the production of potentially fraudulent messages. To do this, we draw on the contributions of Maingueneau (2005 to 2024), Foucault (2009), Coulthard and Johnson (2007), Sousa-Silva and Coulthard (2016), Schmitt and Flechais (2024), Sousa-Silva (2026), Carreira and Nascimento (2020), among others, to understand the relationships between authorship, ethos, enunciation scenes, and the production of linguistic evidence. This is a qualitative research study with an interpretative approach, whose corpus consists of two enunciated excerpts related to social engineering situations from the EvilToken case analyzed by the Microsoft security team on their blog on September 22, 2026. The analysis shows that constructing simulations of institutional identities can produce effects of legitimacy through the appropriation of socially recognized ways of speaking. We conclude that the interface between AD and LF allows for a better understanding of the relationships between enunciation, identity, and authorship, especially when language is part of social practices that can have legal consequences.

KEYWORDS: Discourse Analysis. Forensic Linguistics. Generative Artificial Intelligence. Authorship. Cyberattacks.

RESUMO PARA NÃO ESPECIALISTAS

Este artigo investiga como mensagens fraudulentas podem parecer comunicações verdadeiras de empresas e instituições. O estudo analisa o caso *EvilTokens*, uma plataforma usada em golpes digitais conhecidos como *phishing*, nos quais criminosos tentam convencer pessoas a clicar em links, fornecer informações ou permitir acesso indevido a contas. A pesquisa examina dois exemplos divulgados pela equipe de segurança da Microsoft em setembro de 2026. Para compreender como essas mensagens conquistaram a confiança dos destinatários, os autores combinam conhecimentos da Análise do Discurso, da Linguística Forense e da Ciência da Computação. A análise observa a escolha das palavras, as imagens, a aparência de autoridade e o contexto em que as mensagens circulam. O artigo também discute como ferramentas de Inteligência Artificial podem auxiliar na elaboração de textos enganosos, sem que isso signifique que a tecnologia seja responsável pelas ações criminosas. Os resultados ressaltam a importância de investigar tanto a linguagem quanto as circunstâncias de produção e circulação das mensagens para compreender melhor os golpes digitais e os desafios de identificar seus responsáveis.

Introdução

A apropriação da palavra alheia não é um fenômeno novo na história das sociedades. Desde que a linguagem existe e passou a ser como instrumento de organização das relações humanas, os sujeitos se apropriam, reproduzem, reformulam e deslocam enunciados, atribuindo-lhes sentidos que nem sempre correspondem àqueles produzidos em suas condições primeiras de enunciação. Contudo, com o advento das tecnologias digitais e, mais recentemente, da Inteligência Artificial Generativa (doravante IAG), esse processo assume configurações que merecem atenção, principalmente quando problematizamos as condições de produção de textos que simulam a identidade, o estilo e a autoria de outros sujeitos atribuindo-lhes um simulacro de autoridade.

Se, por um lado, a utilização da IAG facilita o acesso à produção escrita, possibilitando a criação de enunciados em diferentes gêneros discursivos, por outro, também amplia as possibilidades de circulação de textos cuja procedência não pode ser imediatamente reconhecida. Isso ocorre porque os sistemas generativos permitem produzir enunciados com aparência linguística coerente, adequados a diferentes situações comunicativas e, em determinadas circunstâncias, próximos das formas de dizer socialmente atribuídas a sujeitos ou instituições, dependendo, assim, do grau de letramento digital dos sujeitos envolvidos.

Algumas vezes, esses enunciados surgem a partir de uma “parceria” com a máquina em que o nível de letramento acadêmico (Botelho, Carreira e Evangelista, 2024) e digital (Rojo, 2017) entra em jogo e o gênero responde à necessidade imediata do sujeito em seu contexto social. Em outros momentos, seja por falta de letramento (Street, 2014) ou seja por necessidade de dissimulação, a máquina medeia e apaga a autoria propositalmente, como ocorre em ciberataques, isto é, antes mesmo de ocorrer uma possível fraude, há um trabalho discursivo de construção de confiança e de falseamento de identidades em prol de uma possível legitimação dentro do processo de interação.

É importante ter em conta também que vivemos um momento deslegitimação autoral e apagamento de identidades e estilos angariando a autores o fardo de provar-se autor e atestar sua proficiência de escrita socialmente. Por exemplo, ao submetermos ao *ChatGPT* nossos textos antigos escritos sem auxílio de IA foi surpreendente verificar que a máquina os identificou com alto teor de probabilidade de ser escrito por IA, mesmo um deles sendo partes de uma resenha da obra de Antonio Cândido, ou seja, o fato de dominar a escrita do gênero pode alucinar a IA ou confundir algoritmos, o que também pode ocorrer com os identificadores de autoria.

Carreira e Nascimento (2020), ao analisarem a função *pseudônimo de autor*, já preconizavam que a reconfiguração ou desconfiguração autoral sempre esteve em voga porque “a autoria está intimamente relacionada à ideia de legitimação do sujeito do discurso por meio de um dizer que signifique ou que tenha poder de significação social”(Carreira e Nascimento, 2020, p.129).No âmbito digital, os enunciados gerados com uso de IAG’s atingem outras proporções igualmente inquietantes

Nesse trabalho, interessa-nos compreender como essas materialidades linguísticas podem ser apropriadas por sujeitos e utilizadas para construir relações de poder, simular identidades e provocar efeitos de sentido que favoreçam determinadas ações em crimes digitais.

Na instância dos ciberataques, essa discussão torna-se particularmente necessária. Mensagens fraudulentas enviadas por correio eletrônico, aplicativos de comunicação ou redes sociais podem apresentar-se como comunicações de instituições públicas, financeiras ou de sujeitos próximos ao destinatário. Em muitos casos, a estratégia não depende inicialmente da invasão de um sistema computacional, mas da construção de uma situação enunciativa capaz de levar o coenunciador a reconhecer como legítimo um dizer cuja procedência é desconhecida ou simulada.

É nesse espaço que situamos a aproximação entre a Análise do Discurso (doravante AD) de base francesa e a Linguística Forense (LF). Embora essas áreas possam distinguir-se de algum modo em relação aos objetos e aos procedimentos distintos, ambas podem contribuir para a compreensão dos processos de produção e interpretação da linguagem em situações nas quais o dizer ultrapassa a função de comunicar e passa a integrar conflitos, disputas e práticas socialmente relevantes. Entendemos a LF como uma área da Linguística Aplicada, assim, a AD, como disciplina de entremeio, pode fornecer categorias e elementos que possibilitem estudar a situação comunicativa para além dos níveis estruturais, estando *a priori* a serviço da LF em sua aplicação como as demais ferramentas e áreas de saber da Linguística.

A AD, ao considerar as condições sócio-históricas de produção, o interdiscurso e as cenas de enunciação, por exemplo, permite examinar as estratégias por meio das quais determinados enunciados constroem sua legitimidade. A LF, por sua vez, pode mobilizar diferentes abordagens linguísticas para investigar problemas relacionados à autoria, à interpretação de textos e à utilização da linguagem como evidência em situações jurídicas ou investigativas como é possível acompanhar em diversas pesquisas da área. Como observa Sousa-Silva (2026, p. 2):

[a] rápida proliferação dos Grandes Modelos de Linguagem (LLMs) alterou fundamentalmente o panorama das ameaças à cibersegurança. Ataques que antes exigiam muito trabalho e considerável competência linguística formal para parecerem

genuínos — como campanhas de *phishing*, engenharia social e fraudes por personificação — podem agora ser automatizados em larga escala com conhecimento especializado mínimo (Ferrara, 2023; Li e Fung, 2025; Europol, 2026). A ampla disponibilidade de tecnologia capaz de viabilizar usos maliciosos provocou uma necessidade urgente de pesquisa tanto sobre a natureza das ameaças possibilitadas por LLMs quanto sobre os métodos disponíveis para detectá-las e atribuí-las. (Tradução nossa)

Esse caráter de urgência investigativa apontado leva a algumas inquietações iniciais norteadoras desse trabalho, dentro dos limites de um artigo: quem enuncia quando um texto é produzido com a intervenção de um sistema generativo? A autoria permanece vinculada ao sujeito que solicita a produção textual? As marcas estilísticas podem ser utilizadas para atribuir a produção de determinado enunciado a um indivíduo, quando há possibilidade de intervenção tecnológica? E, principalmente, de que maneira a AD pode contribuir para compreender a construção discursiva de identidades simuladas, sem confundir efeitos de autoria com evidências de autoria em um universo tecnodiscursivo (Paveau, 2021)

Partimos da hipótese de que a IAG não elimina a autoria humana, pois como pondera Li (2018) a inteligência artificial nada tem de artificial por ser criada, gerenciada e administrada por humanos, os quais em seus enunciados criam situações nas quais podem deslocar as condições de sua identificação por meio da ajuda das IAG's, ao permitir que sujeitos se apropriem de enunciados produzidos ou reformulados artificialmente para construir simulacros discursivos. Esses simulacros, ao mobilizarem formas de dizer reconhecidas socialmente, podem favorecer a adesão de coenunciadores a práticas comunicativas fraudulentas.

A utilização da noção de simulacro, neste estudo, não pressupõe que todo enunciado produzido por IAG seja falso ou ilegítimo. Trata-se de considerar, em situações específicas de simulação identitária, como um discurso pode construir a imagem de outro discurso ou de outro enunciator, apropriando-se de suas formas de legitimação, porque o objetivo deste artigo é discutir as possibilidades de articulação entre AD e LF para a interpretação de enunciados relacionados à engenharia social, considerando a participação potencial da IAG nos processos de produção, apropriação e circulação textual para refletir sobre essas questões como um primeiro passo para compreensão desses procedimentos na identificação de possíveis atos criminosos.

Para isso, inicialmente, essa escrita nasce de uma parceria interdisciplinar entre autores/investigadores de áreas distintas (AD/LF³ e Ciências da Computação⁴) que juntos buscam pensar essa nova realidade de uma forma salutar para gradativamente apresentar

³ Faculdade de Letras UFG - PPGLL

⁴ Faculdade de Ciências da Computação UFG - CEIA

resultados e aplicações para as diferentes áreas⁵. Assim, trazemos primeiras reflexões sobre aproximações entre AD e LF com foco em conceitos de autoria, cenas de enunciação e *ethos* discursivo para abordar a utilização da IAG em práticas de ciberataques, observando os processos de simulação identitária e construção de legitimidade tendo como corpus *phishings* recentes analisados pela *Microsoft*, divulgados em seu blog em 22 de setembro de 2026. Posteriormente, apresentamos os procedimentos metodológicos e realizamos uma análise discursiva de dois enunciados construídos extraídos do caso *Eviltoken*, que ocorreu de fevereiro a abril de 2026, para representar diferentes situações comunicativas associadas à engenharia social. Por fim, discutimos as implicações dessas práticas para a investigação de autoria e para os estudos da linguagem na relação AD e LF.

1. Análise do Discurso e Linguística Forense: aproximações possíveis

A linguagem não se constitui apenas como um instrumento utilizado pelos sujeitos somente para transmitir informações. Ela participa da constituição das relações sociais, das instituições e dos lugares ocupados pelos indivíduos em diferentes situações históricas. Dito de outro modo, os sujeitos não somente utilizam a linguagem, mas também são constituídos nas e pelas práticas discursivas que organizam sua existência em sociedade.

Essa concepção, bastante presente nos estudos da AD, permite compreender que a interpretação de um enunciado não pode estar dissociada das condições que possibilitaram sua produção, circulação e reconhecimento.

A produção do discurso é, simultaneamente, “controlada, selecionada, organizada e redistribuída” (Foucault, 2012, p. 9) por procedimentos socialmente instituídos e relações de poder que permeiam as camadas sociais em todos os níveis, inclusive, o digital. Se os discursos são atravessados por relações de poder e procedimentos de controle, as possibilidades de enunciação não estão igualmente distribuídas entre os sujeitos, porque há lugares socialmente autorizados/legitimados para dizer, assim como existem formas de dizer que recebem maior legitimidade em determinadas instituições. Um comunicado bancário, por exemplo, adquire estatuto diferente daquele produzido por um indivíduo sem vínculo com a instituição, porque o enunciador institucional ganha na relação interacional um estatuto de confiabilidade, que permite ao coenunciador aceitar a mensagem como verdadeira, ainda que seja falsa, porque há aí uma relação socialmente instituída de controle e poder financeiro que somada a falta de

⁵ Surge da parceria em projetos de pesquisa e extensão como o LABJUS/UFG.

letramento e a cena de enunciação “simulada” pode levar o sujeito a ser vítima de um ciberataque. Não necessariamente porque a estrutura linguística da mensagem/enunciado seja distinta, mas porque há condições institucionais e sociais que autorizam sua produção e circulação, e esse “lugar do dizer” (Carreira, 2026) se consolida nas práticas discursivas.

Entretanto, como reconhecer essas condições quando o enunciado é produzido em ambiente digital e pode ser reproduzido por sujeitos que não ocupam efetivamente o lugar de onde simulam falar? Essa pergunta, entre outras já citadas, nos conduz à discussão sobre as cenas de enunciação.

Maingueneau (2008b) compreende a atividade discursiva considerando diferentes dimensões das cenas enunciativas: *a cena englobante* relaciona-se ao tipo de discurso, enquanto *a cena genérica* remete às condições instituídas pelos gêneros e a cenografia corresponde à cena que o próprio discurso constrói para legitimar sua enunciação constituindo um quadro cênico responsável pela manutenção dos efeitos de sentido.

Essas dimensões são importantes porque demonstram que o enunciado não se apresenta isoladamente ao coenunciador. Ele mobiliza um conjunto de condições que permite reconhecer determinados papéis sociais e interpretar o dizer a partir de expectativas compartilhadas.

Assim, ao receber uma comunicação de uma instituição financeira ou dentro dela, o sujeito reconhece formas de tratamento, terminologias e procedimentos que remetem à *memória discursiva* (Paveau, 2013) daquele campo institucional. O reconhecimento não significa, contudo, que o enunciado seja efetivamente legítimo, mas a cena de enunciação instituída pode levar à adesão e há muitas camadas envolvidas nessa questão, pois não é real atribuir à máquina a excelência nesse tipo de ação, uma vez que são ações humanas e interacionais que possibilitam esse acontecimento discursivo (Foucault, 2001)

É justamente nesse deslocamento entre o reconhecimento da forma discursiva e a confirmação da procedência do enunciado que podemos observar uma questão relevante para a análise de práticas fraudulentas, pois entendemos que a investigação de autoria não se resolve somente pelo reconhecimento de padrões quantitativos, como denota Sousa-Silva (2026), ao discutir as possibilidades de análises forenses mais amplas e qualitativas de modelos generativos:

[d]iferentemente das abordagens estilométricas, que agregam características superficiais, a análise forense de autoria examina dimensões qualitativas, incluindo estratégias pragmáticas idiossincráticas, padrões discursivos, tipologias de erros, mudanças de registro e colocações que tendem a permanecer estáveis ao longo da produção linguística de um autor, formando um conjunto distintivo quando comparadas às características linguísticas exibidas por outros autores (Grant, 2021 apud Sousa-Silva, 2026, p.4) (Tradução nossa).

Essa ideia permite reconhecer que as marcas atribuídas ao estilo não se esgotam nas frequências lexicais, já que a relação entre enunciador, gênero e condições de produção interfere nas escolhas linguísticas. Do mesmo modo, ao privilegiarmos as cenas de enunciação e o *ethos*, não pretendemos substituir os procedimentos de LF, mas ampliar a interpretação das regularidades identificadas, possibilitando uma contribuição entre áreas em prol de resultados que poderão, quem sabe, servir de respaldo para análises factuais de autoria em cibercrimes, uma vez que insere-se no campo dos estudos da linguagem relacionados ao Direito e aos processos investigativos judiciais também, podendo recorrer a diferentes teorias e metodologias linguísticas para compreender situações nas quais a linguagem constitui objeto de interpretação, conflito ou evidência.

Coulthard e Johnson (2007) demonstram a abrangência desse campo ao contemplarem questões relacionadas à linguagem em ambientes jurídicos e à investigação de materialidades linguísticas. Sousa-Silva e Coulthard (2016) também apresentam contribuições para a compreensão das relações entre linguagem e evidência, especialmente quando determinados textos passam a integrar procedimentos investigativos para reconhecimento de autoria.

Uma questão que merece atenção é o fato de que a LF não se confunde com uma teoria de autoria nem se limita à análise de textos produzidos por suspeitos de crimes. Trata-se de um campo mais amplo no qual diferentes fenômenos linguísticos podem apresentar relevância jurídica como análise do discurso jurídico e análise de diferentes *corpora* que podem ir desde os gêneros orais e escritos até áreas específicas como a tradução jurídica.

Nesse sentido, a aproximação com a AD não deve ocorrer apenas pela constatação de que ambas estudam a linguagem, pois essa característica é compartilhada por diferentes áreas da Linguística. O ponto de aproximação que nos interessa está na possibilidade de interpretar os enunciados considerando os sujeitos, as condições de produção e os efeitos de sentido que integram situações potencialmente relevantes para uma investigação constituindo a cena de enunciação (Maingueneau, 2008b) de um crime virtual e instituindo *ethé* (Maingueneau, 2008a) de sujeitos com potencial de autoria a ponto de legitimar um “lugar do dizer” (Carreira, 2026) falso e também compreendendo-o como tecnodiscurso, a partir dos estudos de Paveau (2021) para quem os algoritmos concorrem como elementos participantes dos processos de enunciação.

Tomemos como exemplo uma mensagem na qual o enunciador se apresenta como representante de determinada instituição e solicita informações pessoais ao destinatário. A materialidade linguística poderá ser examinada sob diferentes perspectivas, desde as escolhas lexicais e sintáticas até a organização argumentativa e semiótica. Entretanto, a compreensão de

seu funcionamento discursivo exige observar também a construção da autoridade institucional, as relações entre enunciador e coenunciador e as condições de circulação do texto e da cena de enunciação nas quais estão introjetadas e projetadas relações de poder inerentes ao sistema social capitalista em que os sujeitos, nesse âmbito, deixaram de ser meros utilizadores para serem considerados dados, isto é, produto de valor monetário.

A AD possibilita interpretar essas relações, enquanto a LF poderá utilizar os resultados da análise em conjunto com outros procedimentos adequados à prática investigativa acadêmica e forense, ou seja, a AD entra como uma, entre outras possibilidades disciplinares e teóricas que podem estar a serviço da Linguística Forense.

É importante ressaltar, entretanto, que interpretar uma estratégia de adesão não equivale a comprovar a existência de uma intenção criminosa. Da mesma forma, reconhecer uma identidade institucional simulada não significa identificar o sujeito responsável por sua produção. Essas distinções são necessárias para evitar que a análise discursiva seja confundida com a atribuição pericial de autoria e é nesse aspecto que outros procedimentos metodológicos e a interdisciplinaridade se fazem fundamentais, pois uma coisa é a interpretação desses dados para fins científico-acadêmicos e outra é estar a serviço da justiça elaborando análise pericial.

1.1. O *ethos* e sua participação nos efeitos de sentido de cibercrimes.

O *ethos* discursivo, conforme Maingueneau (2008a, 2015), não corresponde simplesmente à imagem que um sujeito deseja apresentar de si, mas relaciona-se ao modo como determinada imagem do enunciador é construída no próprio funcionamento do discurso, isto é, na enunciação.

Isso significa que a credibilidade de uma mensagem pode ser produzida por escolhas linguísticas que conferem ao enunciador a aparência de competência, autoridade ou proximidade, mesmo quando essas características não correspondem à identidade empírica de quem produziu o texto, mas somente se consolida no processo de enunciação.

A questão torna-se mais complexa quando pensamos nos sistemas de IAG, uma vez que esses recursos podem gerar enunciados linguisticamente adequados a diferentes situações comunicativas, reproduzindo características de gêneros discursivos e estilos institucionais.

Nesse caso, o *ethos* pode ser construído por um sujeito que não redigiu integralmente a mensagem, mas que orientou, selecionou ou utilizou determinada produção textual em um comando/*prompt* para uma IAG.

Dessa maneira, há uma diferença importante entre a construção discursiva de uma identidade e a identificação empírica de seu responsável, que em âmbito digital demanda o reconhecimento do algoritmo e de tecnodiscursos (Paveau, 2021). Se a questão da autoria já vinha sendo colocada em xeque durante todo o século XIX (Carreira e Nascimento, 2020), agora, se tornou fundamental reconsiderá-la a partir de múltiplos olhares.

Foucault (2009), ao problematizar a função-autor, permite compreender que a autoria não se reduz à existência de um indivíduo anterior ao texto, mas participa de práticas de classificação, atribuição e circulação discursiva. Da mesma forma, Maingueneau (2010) discute a autoralidade como uma categoria que ultrapassa a relação imediata entre um nome e uma obra, considerando as dimensões enunciativas e sociais envolvidas.

Se essas discussões já apresentavam dificuldades para a compreensão da autoria em textos tradicionalmente produzidos por sujeitos humanos individual ou coletivamente, a participação de sistemas generativos introduz outras possibilidades de deslocamento e simulacros a serem compreendidos. Para Paveau (2021, p. 31):

no trabalho de análise do discurso digital, para dar conta de uma evolução teórica necessária: falar em tecnodiscurso, tecnopalavra, tecnosigno, tecnogênero do discurso e tecnografismo é inscrever na análise uma opção teórica que modifica a episteme dominante das ciências da linguagem. É também afirmar que os discursos digitais nativos não são de ordem puramente linguageira, que as determinações técnicas coconstroem as formas tecnolinguageiras, e que as perspectivas logo e antropocêntricas devem ser descartadas em prol de uma perspectiva ecológica e integrativa, que reconhece o papel dos agentes não humanos nas produções linguageiras. As ciências da informação e da comunicação falam em tecnossemiótica desde os anos 1990, para representar essa integração dos dispositivostécnicos numa análise da construção do sentido.

Um sujeito pode produzir uma mensagem integralmente, utilizar um sistema para corrigir sua escrita, solicitar a elaboração completa de um texto ou reformular uma materialidade produzida artificialmente. Em todas essas situações, a relação entre sujeito e texto apresenta condições distintas de “produção, circulação e poder” (Foucault, 2009).

Logo, não parece suficiente afirmar que a autoria pertence à IAG porque ela produziu a sequência linguística, tampouco atribuí-la integralmente ao sujeito que solicitou sua produção, sem considerar o processo efetivo de elaboração.

A autoria, nesse contexto, precisa ser investigada considerando diferentes níveis de participação na produção discursiva.

É nesse ponto que a aproximação entre AD e LF se mostra produtiva. A primeira permite interpretar a construção dos lugares enunciativos e os processos de legitimação do dizer; a

segunda oferece procedimentos para investigar questões relacionadas à atribuição de autoria, desde que existam materiais e condições metodológicas adequadas para isso.

Entendemos, assim, que a contribuição dessa interface não consiste em transformar a AD em instrumento automático de identificação de autores, modelo ou protocolo de análise, mas em ampliar a compreensão de como os enunciados constroem identidades, mobilizam discursos anteriores e participam de práticas sociais que podem apresentar consequências jurídicas e também como os âmbitos discursivos intervêm em aspectos semânticos e relações de poder sócio-historicamente construídas que emanam e corroboram para a interpretação de gêneros textuais digitais criminosos.

2. Inteligência Artificial Generativa e ciberataques: produção, apropriação e simulação discursiva

A IAG representa um avanço significativo na Ciência da Computação, especialmente pela capacidade de produzir textos, imagens, áudios e códigos computacionais a partir de padrões aprendidos em grandes conjuntos de dados. Entre suas principais aplicações destacam-se os Grandes Modelos de Linguagem (*Large Language Models* – LLMs), fundamentados predominantemente na arquitetura *Transformer*, introduzida por Vaswani *et al.* (2017).

Esses modelos utilizam mecanismos de atenção para estabelecer relações entre elementos de uma sequência textual, permitindo gerar respostas contextualizadas a partir de instruções formuladas pelos usuários. Embora apresentem resultados linguisticamente coerentes, suas respostas não asseguram necessariamente precisão factual, compreensão semântica ou adequação ao contexto, exigindo mecanismos de validação e supervisão humana. E embora tanto no âmbito educacional quanto em algumas áreas de estudos linguísticos seja muito criticada, seria uma hipocrisia negar seu uso massivo social e não tentar entender o que ela significa para as diferentes áreas de saber significa discutir sobre possibilidades que só existem em sentido restrito sem consideração de um futuro que, provavelmente, será muito diferente do imaginado por todas as áreas não específicas e, ousamos dizer, para as tecnológicas também.

A interação com esses sistemas ocorre principalmente por meio de instruções em linguagem natural, denominadas *prompts*, que orientam a geração das respostas. Conforme demonstrado por Brown *et al.* (2020), modelos de linguagem podem adaptar seu comportamento a diferentes tarefas a partir de instruções e exemplos apresentados no contexto.

Em pesquisa sobre a aplicação de IAG à identificação de similaridades entre precedentes judiciais, Sene Junior *et al.* (s.d.) exploram o uso do ChatGPT e do Gemini na comparação de documentos jurídicos, destacando a importância da contextualização das consultas, da engenharia de *prompts* e da verificação das respostas produzidas. Embora desenvolvida em um domínio distinto da segurança cibernética, essa investigação evidencia como a formulação das instruções influencia os conteúdos gerados e reforça a necessidade de avaliar criticamente os resultados apresentados pelos modelos.

No campo da segurança da informação, a IAG apresenta uma característica de uso dual: as mesmas funcionalidades empregadas em atividades legítimas podem ser exploradas em ações maliciosas. Weidinger *et al.* (2022) identificam diferentes categorias de riscos associados aos modelos de linguagem, incluindo a produção de informações enganosas e sua utilização para finalidades prejudiciais. Nos ciberataques, essas tecnologias podem auxiliar na elaboração de mensagens fraudulentas, na simulação de comunicações institucionais e na adaptação de conteúdos direcionados a determinados públicos. Entretanto, é importante distinguir a capacidade de gerar conteúdos da efetiva realização de um ataque, pois o comprometimento de sistemas ou usuários depende de condições técnicas, comportamentais e contextuais específicas.

Uma das aplicações mais relevantes nesse cenário é o *phishing*, modalidade de engenharia social baseada em comunicações fraudulentas destinadas a induzir usuários a fornecer informações confidenciais, acessar páginas maliciosas ou executar operações indevidas. Com a utilização da IAG, torna-se possível produzir mensagens gramaticalmente elaboradas e adaptadas a diferentes situações comunicativas, reduzindo o esforço necessário para sua elaboração e personalização. Entretanto, a qualidade linguística não garante o sucesso da fraude, uma vez que sua eficácia depende também da confiança atribuída ao remetente (enunciador), do contexto de circulação e do comportamento do destinatário (coenunciador/enunciatário). Nesse sentido, os riscos discutidos por Weidinger *et al.* (2022) permitem compreender como sistemas generativos podem ampliar as possibilidades de produção e disseminação de conteúdos potencialmente enganosos.

Considere-se, por exemplo, uma mensagem eletrônica que simula uma comunicação institucional solicitando a atualização urgente de credenciais de acesso que se assemelha as estratégias do *Eviltoken*. O texto pode reproduzir expressões administrativas, referências organizacionais e estruturas típicas de comunicados oficiais, procurando estabelecer uma aparência de legitimidade, criando uma cena de enunciação que se apropria de elementos reais para criar uma cena falsa, como Carreira (2026) demonstra em *fakenews* e também elementos

não-usuais podem aderir as necessidades dos agentes criminosos. Sob a perspectiva da AD, a linguagem não atua apenas como meio de transmissão de informações, mas participa da construção de sentidos e de relações de autoridade. Expressões como "regularização cadastral", "suspensão de acesso" e "prazo obrigatório" podem mobilizar sentidos associados à urgência e à obediência institucional, influenciando a interpretação do destinatário/coenunciador.

A incorporação da IAG a essas práticas introduz novas questões sobre as condições de produção discursiva. O conteúdo fraudulento pode resultar da interação entre um agente humano e um modelo computacional, mediante sucessivas instruções, reformulações e ajustes. Nesse processo, o sistema não deve ser compreendido como sujeito dotado de intenção própria, mas como um recurso tecnológico utilizado na produção textual, cuja intencionalidade do ataque, obviamente, permanece associada aos agentes responsáveis por sua elaboração e utilização.

No âmbito da LF, esse cenário amplia os desafios relacionados à análise de autoria e à interpretação de evidências textuais. A utilização de modelos generativos permite modificar estilos de escrita, reorganizar estruturas sintáticas e produzir diferentes versões de uma mesma mensagem, dificultando a identificação de padrões individuais. A presença de fluência linguística, correção gramatical ou determinadas escolhas lexicais não constitui, isoladamente, evidência suficiente para determinar se um texto foi produzido por uma pessoa ou por um sistema de IAG. Assim, a análise forense deveria considerar conjuntamente características linguísticas, elementos pragmáticos, condições de circulação e, quando disponíveis, evidências técnicas, como metadados, cabeçalhos de mensagens e registros de acesso.

Outra dimensão importante envolve os ataques de *prompt injection*, nos quais instruções maliciosas são inseridas em conteúdos processados por modelos de linguagem com o objetivo de alterar seu comportamento. Perez e Ribeiro (2022) investigam técnicas de manipulação de instruções que evidenciam vulnerabilidades relacionadas ao processamento de comandos em linguagem natural. Diferentemente do *phishing*, direcionado à manipulação de pessoas, a injeção de *prompts* procura interferir no comportamento de sistemas computacionais. Essa distinção demonstra que a linguagem pode desempenhar diferentes funções nos ciberataques mediados por IAG, atuando tanto na construção de mensagens persuasivas maliciosas quanto na tentativa de redirecionamento de sistemas automatizados.

Dessa forma, a aproximação entre Ciência da Computação, AD e LF oferece uma perspectiva interdisciplinar para investigar os ciberataques mediados por IAG. Enquanto a dimensão computacional permite compreender os mecanismos de geração textual e suas limitações e o percurso tecnológico da engenharia social, as áreas da Linguística contribuem

para examinar componentes linguageiros específicos. Compreender essas práticas exige, portanto, reconhecer que as vulnerabilidades exploradas não se restringem aos sistemas tecnológicos, mas envolvem também os processos sociais de interpretação, confiança e interação mediados pela linguagem.

Há, nesse processo, uma questão discursiva importante: a possibilidade de apropriação de formas de dizer que, ao serem reconhecidas pelo coenunciador, podem conferir legitimidade a uma identidade simulada.

A memória discursiva (Paveau, 2013) permite ao coenunciador reconhecer determinadas práticas comunicativas e isso não é suficiente para o reconhecimento da veracidade ou falsidade da mensagem, pois a cena de enunciação pode legitimar para o usuário uma pseudo-verdade ou uma verossimilhança. Assim, podemos observar uma distinção entre o lugar institucional representado no discurso e as condições efetivas de sua produção.

A partir da perspectiva de Maingueneau (2008a, 2018), essa distinção permite examinar as cenas de enunciação e os processos de construção do *ethos* discursivo, no que concerne aos aspectos de *categoria (papéis/estatuto)*, *experiência (modos de ser)* e *ideologia (posicionamento)*. O enunciador que simula uma instituição procura construir uma imagem de autoridade, competência e confiabilidade, mobilizando características discursivas que podem favorecer a adesão do coenunciador. Não se trata apenas de reproduzir palavras utilizadas por uma instituição, mas de construir uma cenografia na qual o discurso aparece como legítimo.

Desse modo, não basta classificar uma mensagem como artificial ou humana. É necessário investigar quais características decorrem do gênero, quais podem ser atribuídas às condições interacionais e quais seriam, de fato, pertinentes à hipótese de autoria. A distinção torna-se particularmente relevante quando os textos examinados são curtos, reformulados ou produzidos em condições discursivas distintas.

Se a identificação de autoria já constitui um problema relevante para a LF, a possibilidade de intervenção da IAG na elaboração textual introduz novas dificuldades. Um mesmo enunciado pode resultar da participação de diferentes sujeitos, de processos de reformulação e de intervenções computacionais, sem que essas operações sejam imediatamente reconhecíveis no texto final. Assim, torna-se necessário distinguir a autoria discursiva, a produção material da sequência linguística e a responsabilidade pela circulação do enunciado, ou seja, a autoria tecnológica da arquitetura que baseia essa circulação.

Jawahar, Abdul-Mageed e Lakshmanan (2020), ao discutirem a detecção automática de textos produzidos por máquinas, situam um problema que se tornou ainda mais relevante com

o desenvolvimento dos modelos generativos contemporâneos: a identificação da origem computacional de uma materialidade linguística.

Entretanto, é preciso considerar que os procedimentos de detecção apresentam limitações, principalmente quando os textos são curtos, passam por intervenções humanas ou são produzidos mediante instruções específicas. Isso significa que a identificação de autoria não pode estar condicionada apenas ao reconhecimento de determinadas escolhas lexicais, sintáticas ou estilísticas.

A pesquisa de Sousa-Silva (2026) também explicita por que a mudança de gênero exige prudência na comparação de textos. Como observa:

[L]inguistas forenses comparam rotineiramente textos produzidos sob diferentes restrições situacionais — como escrita acadêmica ou formal *versus* textos ofensivos, tóxicos ou supostamente difamatórios. Como resultado, é incomum que os textos conhecidos disponíveis coincidam em gênero com os textos questionados. (...) exigindo controle cuidadoso de características dependentes do gênero e da situação ao avaliar traços relacionados à autoria. (Sousa-Silva, 2026, p.12) (Tradução nossa)

Do ponto de vista discursivo, essas possibilidades ampliam as questões relacionadas à construção de *ethé* e aos processos de legitimação do “lugar do dizer” (Carreira, 2020).

Uma identidade pode ser simulada por meio de formas de tratamento, escolhas lexicais, características de voz ou representações visuais. Entretanto, em qualquer dessas situações, é necessário distinguir a identidade construída na enunciação daquela efetivamente responsável pela produção e circulação do material.

A discussão também permite retomar a noção de simulacro, considerando as contribuições de Maingueneau (2008a) para a compreensão das relações entre posicionamentos discursivos. Quando determinado sujeito se apropria de formas de dizer associadas a uma instituição, não está necessariamente reproduzindo seu discurso original, mas construindo uma representação desse discurso a partir de determinadas características consideradas reconhecíveis.

A utilização da IAG em práticas de ciberataques evidencia que a vulnerabilidade explorada nem sempre se encontra no funcionamento dos dispositivos tecnológicos. Em determinadas situações, está relacionada às condições de reconhecimento, confiança e interpretação construídas nas interações discursivas. Assim, compreender a relação entre IAG e ciberataques talvez exija ultrapassar uma perspectiva exclusivamente tecnológica, considerando os sujeitos e as práticas sociais nas quais os enunciados adquirem sentido.

Dessa maneira, entendemos que a IAG não inaugura a simulação discursiva nem elimina a autoria humana. Contudo, pode ampliar as possibilidades de produção e circulação de simulacros, introduzindo novas condições para os estudos de autoria e para a interpretação de evidências linguísticas no campo forense.

3. Metodologia

A presente pesquisa é de natureza qualitativa, de cunho interpretativo, orientada pelos pressupostos da AD, em interface com a LF, tendo como base os estudos Sousa-Silva (2026)

Nosso objeto de investigação é a construção discursiva de identidades simuladas em mensagens relacionadas à engenharia social, considerando as possibilidades de intervenção da IAG na produção de enunciados em *phishings*.

O *corpus* constitui-se por um caso recentes analisado pela de segurança da *Microsoft* para chegar a sua composição, fizemos uma busca simples utilizando *ChatGPT*:

Etapas I - prompt [Pesquise pishings de ciberataques reais comprovados e divulgados pela mídia]

Figura 1 – Levantamento de Casos Recentes

1. Pedido inicial — casos reais de phishing		
USUÁRIA — Pesquise pishings de ciberataques reais que tenham sido divulgados pela mídia		
ASSISTENTE — Foram indicados casos de phishing, spear phishing e engenharia social cobertos por imprensa ou investigadores, com atenção à origem documental e à possibilidade de análise do enunciado fraudulento.		
Caso	Ano	Modalidade descrita
John Podesta / <u>campanha de Hillary Clinton</u>	2016	Spear phishing: alerta falso de segurança Google
RSA Security	2011	Spear phishing com arquivo Excel "2011 Recruitment plan"
Twitter (<u>atual X</u>)	2020	Spear phishing por telefone
Uber	2022	Engenharia social e solicitações MFA
MGM Resorts	2023	Engenharia social por telefone
Microsoft / <u>EvilTokens</u>	2026	Phishing por código de dispositivo; campanhas com IA
<u>Energia na Ucrânia</u>	2015-2016	Spear phishing com anexos maliciosos

Etapas II – prompt – Quais foram feitos por IA?

Figura 2 – Delimitação e composição do corpus

3. Pergunta sobre uso de IA

USUÁRIA — Quais foram feitos por IA?

ASSISTENTE — Entre os casos inicialmente listados, o relatório referente à campanha Microsoft/*EvilTokens* foi apresentado como o caso com comprovação explícita de uso de IA na campanha. A documentação dos demais casos não demonstrava uso de IA na redação dos e-mails.

Caso	Ano	Situação do uso de IA indicada na conversa
Podesta	2016	Não documentado
RSA	2011	Não documentado
Twitter	2020	Não documentado
Uber	2022	Não documentado
MGM	2023	Não documentado
Microsoft/ <i>EvilTokens</i>	2026	IA utilizada na campanha; não atribuível automaticamente a cada e-mail
Ucrânia	2015–2016	Não documentado
Silent Lynx	2026	Não confirmado
Microsoft Power BI	2026	Não confirmado

Etapa III – Prompt – Quero algumas mensagens/phishings produzidos por IA

4. Solicitação das mensagens da Microsoft

USUÁRIA — Quero algumas mensagens enviadas no caso Microsoft para analisar

ASSISTENTE — Foi apresentada a existência de uma figura de e-mail de phishing no estudo Microsoft “Unmasking EvilTokens”, publicado em 22/09/2026. A conversa transcreveu um fragmento legível da mensagem comercial, com trechos ocultados pelos investigadores.

Trecho em inglês conforme apresentado na conversa:

Good morning,
I would like to invite you to submit a proposal for an upcoming project with [trecho ocultado]. The full Request for Proposal (RFP), along with all relevant project details, can be accessed via adobe.
[Link de referência]
If you have any questions or require any clarification, please do not hesitate to reach out to me directly.
I look forward to reviewing your proposal.

Tradução apresentada: Bom dia. Gostaria de convidá-lo a apresentar uma proposta para um próximo projeto com [trecho ocultado]. A solicitação completa de proposta (RFP), juntamente com todos os detalhes relevantes do projeto, pode ser acessada pelo Adobe. [Link de referência]. Caso tenha alguma dúvida ou necessite de esclarecimentos, não hesite em entrar em contato diretamente comigo. Aguardo sua proposta.

Etapa IV – Notadamente, houve perda de características originais da mensagem. Fomos ao link fornecido: <https://www.microsoft.com/en-us/security/blog/2026/09/22/unmasking-eviltokens-getting-to-the-root-of-device-code-phishing/>.

Na conferência, nos detivemos na compreensão do ataque, recortamos os elementos da cena de enunciação, aparentemente, mais significativos, considerando os limites da análise de um artigo. Os enunciados apresentados foram retirados de parte do relatório investigativo intitulado *Unmasking EvilTokens: Getting to the root of device code phishing (Desmascarando EvilTokens: Chegando à raiz do phishing de código de dispositivo)* de autoria de *Microsoft Threat Intelligence, Especialistas do Microsoft Defender e Security Reserch* no blog da *Microsoft*, publicado em 22 de setembro de 2026, assim, não correspondem a mensagens coletadas em investigações jurídicas diretamente, mas constituem evidência de utilização de IAG por autores de crimes no caso e contemplam nosso objetivo nesta pesquisa, pois, conforme os especialistas, “uma das principais funcionalidades do *EvilTokens* é o seu fluxo de *phishing* por código de dispositivo” (Equipe Microsoft Defender, 2026, s/n)

Ao adotar esse procedimento, buscamos evitar a redução da análise à identificação de palavras ou estruturas gramaticais, uma vez que o funcionamento discursivo depende das relações entre os diferentes planos da enunciação que se estabelece em camadas enunciativas enredadas como demonstraremos nos resultados.

A análise não procura estabelecer uma classificação automática de mensagens fraudulentas, mas observar os modos pelos quais o discurso pode favorecer a construção de simulacros discursivos em determinadas situações comunicativas estabelecendo imagens discursivas e cenas de enunciação que dissimulam e concorrem em espaços de poder. Assim os recortes foram escolhidos pelas etapas de enredamento das cenas de enunciação: Recorte I – Figuras 1 e 2 – A ciberdistribuição como rompimento da bolha econômica e apropriação de poder: necessários para a compreensão do surgimento dos *phishings*; Recorte II – O *phishing* por *e-mail*.

4. *Corpus* e análise: a construção discursiva da legitimidade

A *EvilTokens*, plataforma de *phishing* como serviço – cuja sigla é apresentada como PhaaS - surgiu em fevereiro de 2026 como um serviço disponibilizador de ferramentas para criar mensagens fraudulentas personalizadas por IAG, comprometer contas de e-mail e identificar vítimas de maior valor financeiro ou valor estratégico. Segundo o relatório, as campanhas afetaram mais de 12 mil caixas de entrada em mais de 10 mil organizações até abril do mesmo ano.

A característica central do ataque é a exploração de um mecanismo legítimo de autenticação por código de dispositivo da *Microsoft*, que pode permitir ao criminoso obter *tokens* de acesso sem conhecer diretamente a senha da vítima. O crime acontecia nas seguintes etapas:

Tabela I – Etapas do *Eviltokens*

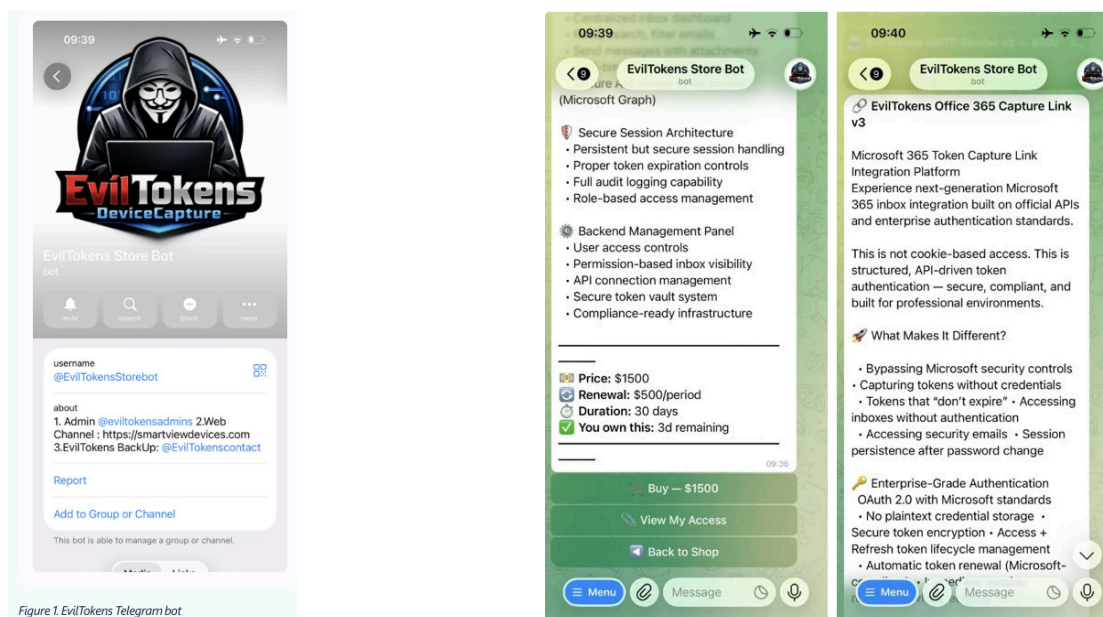
Etapa	Funcionamento
Preparação do ataque	Uso da plataforma e de modelos; a IA auxilia na personalização das mensagens conforme o destinatário.
Construção da mensagem	Simulação de avisos de senha, faturas, documentos compartilhados ou propostas comerciais.
Envio e interação	A vítima recebe o <i>e-mail/pishing</i> e é estimulada a clicar no <i>link</i> ou abrir o anexo.
Autenticação fraudulenta	Uma página fornece um código e encaminha a vítima ao site oficial de autenticação da <i>Microsoft</i> .
Roubo de <i>tokens</i>	A inserção do código autoriza inadvertidamente a sessão do criminoso.
Investigação da conta	Recursos de IA auxiliam na análise de mensagens e na identificação de alvos relevantes.
Manutenção do acesso	O operador pode registrar dispositivos ou criar regras de <i>e-mail</i> para ocultar atividades.

Adaptado de Microsoft. Disponível em: <https://www.microsoft.com/en-us/security/blog/2026/09/22/unmasking-eviltokens-getting-to-the-root-of-device-code-phishing/>. Acesso em 15 de out. de 2026.

As funções da IAG, entre outras, são: *produção enunciativa* (auxílio à redação de mensagens ajustadas ao interlocutor e ao contexto profissional); *análise de informações* (exame de correspondências após o comprometimento e identificação de informações financeiras e novas vítimas) e escala dos ataques (apoio, junto à automação, à realização de campanhas personalizadas para muitos destinatários). Percebemos, assim, que existe um trajeto ramificado de ações humanas que sobrepõem discursos, interdiscursos e, conseqüentemente, cenas de enunciação, que se dão em rede, pela rede e para a rede interna a empresa e externa, em forma de engenharia social, enredando sentidos alcançáveis e inalcançáveis para criação de simulacros por meio de enunciados que diremos que são *enredados* por todo esse ciclo tecnológico discursivo.

4.1. Recorte I — A autoridade como condição de legitimação da autoralidade

Enunciado 1 — Eviltokens - Telegram bots



Adaptado de Microsoft. Disponível em: <https://www.microsoft.com/en-us/security/blog/2026/09/22/unmasking-eviltokens-getting-to-the-root-of-device-code-phishing/>. Acesso em 15 de out. de 2026.

As figuras 1 e 2 nos interessam, conforme citado na seção anterior por mostrar ciberdistribuição como cibercultura de determinado grupo social e enunciativo digital que surge como rompimento da bolha econômica, cuja análise apropriação da forma como se apropriam desse lugar de poder é necessária para a compreensão do surgimento dos *phishings* dentro do enredamento cenográfico como um todo. O primeiro recorte apresenta uma situação comunicativa que se aproxima de mensagens comerciais em espaço de vendas, o enunciador

vende a ferramenta para que outros criminosos possam “aplicá-la”. As imagens estão apresentadas no *blog* exatamente como apareciam nas telas dos celulares, pois a venda ocorria por meio do *telegram*. A máscara da imagem que representa o enunciador principal, ao centro, estabelece interdiscursividade clara com a máscara utilizada em filmes como “V de Vendetta” e outros. No filme, ela é símbolo de resistência e poder da coletividade sobre um sujeito, inspirada em Guy Fawkes⁶, segundo seu criador gráfico David Lloyd.

Com o movimento *Anonymous*⁷, ela ganha novos significados. A máscara cria uma espécie de unidade linguística que apela à memória discursiva (Paveau, 2013) remetendo ao anonimato em prol de uma causa coletiva e mobiliza associações culturais com resistência, desafio à autoridade e confronto com instituições, torna-se, também o centro da cena de enunciação, cuja cena englobante, nessa etapa, é o discurso midiático/digital/comercial, a cena genérica é a propaganda para a venda do *Eviltokens* e a cenografia é de ataque ao sistema digital e capitalista das grandes empresas, contudo, cria um simulacro para a ação criminosa de forjar senhas, *tokens*, *e-mails* e identidades, ou seja, transferir dinheiro e informação. Essa imagem é a semiose central de um *ethos* de confiabilidade que atribui a si o *status* de poder financeiro e político social. Trata-se de um *ethos* também de especialista, cujo coenunciador também é especialista, revelada pela própria forma como o discurso se apresenta, a saber, os enunciados curtos, organizados em tópicos com ícones, o que pode sugerir uso de IAG, o que reforça essa imagem de especialista e deixa marcas discursivas para atrair a adesão desses coenunciadores e vender, mas os enunciados aparentam ter sido copiados e colados ali, pois não há formatação de espaços ou preocupação com a pontuação e não são feitos em formato empresarial para divulgação de um produto.

Na seleção lexical, seu produto se diferencia pelo que promete em relação à segurança da *Microsoft*. Todo esse quadro cênico, somado a essas características convergem com os estudos de LF, pois, com um número maior de mensagens reais, seria possível verificar se o enunciador mantém o idioleto tecnicista e essa imagem de si, deixando somente marcas de formatação e pontuação como a ausência de ponto-e-vírgula ao final das enumerações e os

⁶ Guy Fawkes (1570–1606) foi um conspirador inglês que participou da chamada Conspiração da Pólvora (Gunpowder Plot), em 1605, cujo objetivo era explodir o Parlamento britânico e assassinar o rei Jaime I. Sua imagem foi popularizada pela história em quadrinhos V de Vingança, de Alan Moore e David Lloyd, posteriormente adaptada para o cinema em 2005. A máscara estilizada de Guy Fawkes, utilizada pelo protagonista da narrativa, foi apropriada pelo *Anonymous*.

⁷ *Anonymous* - coletivo descentralizado de ativistas e hacktivistas surgido no ambiente digital dos anos 2000. Associado a ações de protesto, mobilizações políticas e operações cibernéticas, o *Anonymous* adotou essa representação como símbolo de anonimato, resistência à vigilância e contestação das estruturas de poder. Nesse processo de apropriação e ressignificação, a máscara passou a constituir um signo de identificação coletiva, articulando memória histórica, cultura digital e práticas contemporâneas de contestação política

espaçamentos, sem deixar de considerar a intenção comunicativa criada para vender o produto criminoso.

Logo, percebemos que existe uma intenção de autoria, uma autoralidade dissimulada pela imagem. Desse modo, para sua identificação seria necessária uma busca algorítmica para alcançar os protocolos, percursos e fundamentos da arquitetura *Transformer*, introduzida por Vaswani et al. (2017), por trás desse enredamento discursivo.

4.2. Recorte II — A apropriação do discurso administrativo

Enunciado 2



Adaptado de Microsoft. Disponível em: <https://www.microsoft.com/en-us/security/blog/2026/09/22/unmasking-eviltokens-getting-to-the-root-of-device-code-phishing/>. Acesso em 15 de out. de 2026.

Transcrição do recorte

Good morning,
I would like to invite you to submit a proposal for an upcoming project with [trecho ocultado]. The full Request for Proposal (RFP), along with all relevant project details, can be accessed via adobe. [Link de referência]
If you have any questions or require any clarification, please do not hesitate to reach out to me directly.
I look forward to reviewing your proposal.

Esse *pushing* foi enviado a CEO de empresas e a departamentos financeiros, no entanto, a forma como o *Eviltoken* é implementado nas empresas cria uma situação comunicativa de confiabilidade por mimetização e personalização institucional, o que faz com que os sujeitos não desconfiem e sigam os comandos, cedendo sem saber *tokens* de acesso à transferências monetárias ou informações relevantes. Embora a mensagem exposta no *blog* da *Microsoft* esteja sem os elementos de identificação empresarial como *slogans*, assinaturas e endereço digital, trata-se de um *phishing* de circulação interna o que administra a confiabilidade da mensagem que se constitui em cena de enunciação, assim, embora haja elementos linguísticos que notadamente marcam possibilidade de desvio de autoria para esse tipo de comunicação como iniciar com “good morning” e usar a primeira pessoa do singular. A equipe da *Microsoft* identificou o uso de IA nessa etapa, logo, notamos um autor que não soube elaborar ou admitiu/aderiu ao enunciado criado pela IAG como legítima para um ambiente formal, ainda

assim, apesar de tudo, muitos coenunciadores foram levados a seguir os passos e clicar nos *links*.

Entretanto, a questão que nos interessa não é apenas identificar essas expressões, mas compreender como elas participam da construção de uma cena de enunciação (Maingueneau, 2008b) que enreda e leva sujeitos a aderirem ao proposto. Nela, o sujeito se apresenta como representante de uma instituição. Nessa etapa, a cena englobante é a profissional, a cena genérica é o *email profissional* e a cenografia é a de proposta de negócios, justamente, o que supõe a credibilidade. No entanto, o gênero que se sobrepõe à cena genérica é o *phishing* ressignificando todo o quadro cênico transformando-o em ciberataque.

Essa escolha constrói um enunciador que faz um convite, que não se apresenta como sujeito individual, mas como instância dotada de autoridade para acompanhar operações financeiras da empresa, projeta um *ethos* de autoridade institucional que pode levar o coenunciador a aderir à cena de enunciação que é de risco à integridade financeira do enunciador potencial e institucional. O enunciado “If you have any questions or require any clarification, please do not hesitate to reach out to me directly...”. Gera inquirição e atribui a esse *ethos* (Maingueneau, 2008) o poder de dominar o conhecimento necessário para a execução da proposta do enunciador/cliente, replicando relações de poder geradas pelo sistema capitalista que transforma seus usuários em ferramentas de lucro e em produtos de dados para terceiros, o sujeito/enunciador, cujos letramentos digitais não estabeleçam conexões entre as orientações e a realidade interna da empresa, certificando-se da veracidade, podem perfeitamente aderir a esse discurso, sendo meros expectadores subservientes e atendendo aos comandos, assim, de um jeito criminoso e enviesado os cibercriminosos rompem a bolha do capital e transferem esse poder para si.

Essa posição é reforçada pela expressão “I look forward to reviewing your proposal”, que apresenta a solicitação como procedimento de expansão de lucros, mas leva ao *phishing* de códigos que gerou o vazamento de capitais.

Temos, portanto, um deslocamento importante: o sujeito que solicita informações coloca-se discursivamente no lugar daquele que protege informações e amplia negócios, enquanto o coenunciador passa a ocupar a posição de alguém que necessita de sua intervenção.

A cenografia construída aproxima-se daquela de uma comunicação institucional. O discurso institui uma relação na qual o enunciador apresenta competência técnica e o coenunciador deve reconhecer essa competência para agir adequadamente. Importante salientar que a identificação de IA nessa fase não é suficiente para atribuir a ela a autoria do texto, há uma participação na construção dessa autorialidade criminosa.

Do ponto de vista do *ethos*, o enunciador apresenta-se como autoridade competente e responsável, ainda que essa condição não esteja comprovada, assim, a legitimidade é construída no próprio discurso.

O problema que surge para a LF diz respeito à distinção entre a identidade representada pela mensagem e a identidade efetiva de seu produtor.

Esse enunciado integra uma investigação técnico-judicial, por isso, é necessário compreender sua origem, as condições de envio e os eventuais vínculos com outros materiais. A análise das escolhas linguísticas poderia contribuir para descrever o funcionamento da mensagem, mas não seria suficiente para identificar quem a produziu apenas a partir de uma amostragem sem uso de outras ferramentas.

A possibilidade de utilização de IAG introduziria ainda outra dimensão: a materialidade textual poderia ter sido elaborada, reformulada ou adaptada por um sistema, sem que essa intervenção estivesse explicitada no enunciado.

Nesse sentido, o discurso constrói a autoridade, mas ainda não comprova sua procedência, por isso, a conjugação de análises interdisciplinares pode levar a hipóteses que levem a sua origem não somente pela identificação do IP ou registro da máquina, ou pelo caminho de rede de envio, mas por possíveis características idioletais. O coenunciador, na relação de poder instituição x máquina x humano pode se sentir coagido a cumprir essa “exigência”, ainda que tenha letramento financeiro e digital, pode sentir-se co-participante da cena de enunciação que “informa”.

O enunciador apresenta sua solicitação como possibilidade ou convite e atribui-lhe caráter compulsório, ainda que não informe quais normas justificariam esse envio. A enunciação aqui procura legitimar a solicitação a partir de uma autoridade institucional presumida.

A relação entre obrigação e manutenção de acesso à partilha de um negócio também contribui para a construção do efeito de sentido pretendido. O coenunciador é colocado diante de uma alternativa aparentemente simples: realizar o procedimento ou correr o risco de perder a oportunidade de apresentar uma proposta. No entanto, como a anterior, a mensagem não apresenta elementos suficientes para identificar a instituição responsável, a natureza da atualização ou a autenticidade do procedimento, ambas dependem ainda do meio de divulgação: *whatsapp*, *e-mail*, *sms* etc. É justamente isso que torna relevante a construção da cenografia (Maingueneau, 2008b).

O discurso procura produzir uma cena administrativa reconhecível, mobilizando expressões frequentemente associadas a comunicações institucionais. A familiaridade dessas

formas de dizer pode favorecer o reconhecimento da mensagem como legítima, mesmo quando as condições efetivas de sua produção permanecem desconhecidas.

No âmbito da AD, essa apropriação (Maingueneau, 2025) pode ser examinada a partir das relações interdiscursivas, uma vez que a mensagem mobiliza formas de dizer que circulam em diferentes instituições.

O enunciador não precisa reproduzir integralmente um documento oficial para construir sua legitimidade, seu lugar do dizer, conforme demonstra Carreira (2026). Basta que mobilize características linguísticas e enunciativas capazes de remeter a práticas socialmente reconhecidas.

Se considerarmos a participação potencial da IAG, podemos observar que sistemas generativos são capazes de produzir enunciados adaptados ao gênero administrativo, mobilizando formas de tratamento, vocabulário e construções sintáticas adequadas ao contexto solicitado.

Uma investigação forense precisaria considerar outras evidências de produção, circulação e eventual intervenção tecnológica.

A contribuição da AD encontra-se na compreensão do processo de legitimação, enquanto a questão da atribuição de autoria permanece dependente de procedimentos específicos.

Os dois recortes analisados demonstram que a legitimidade discursiva pode ser construída a partir de diferentes posições enunciativas. No primeiro, a autoridade digital e política, por meio da interdiscursividade da máscara e da disposição dos elementos, funciona como elemento de reconhecimento; no segundo, a obrigação administrativa em prol das relações de poder que envolvem dinheiro em uma empresa e a familiaridade interpessoal constitui a principal condição de credibilidade.

Apesar das diferenças, encontramos um funcionamento comum: a construção de uma identidade cuja legitimidade depende da relação entre o discurso e as expectativas sociais do coenunciador. Essa observação nos permite ampliar a discussão sobre a autoria.

Quando determinado sujeito se apropria de formas de dizer socialmente reconhecidas para produzir uma identidade simulada, não está apenas reproduzindo expressões linguísticas, mas está mobilizando condições discursivas que permitem ao enunciado ocupar um lugar de legitimidade, nesse sentido, podemos pensar a relação entre apropriação e aderência.

A apropriação discursiva (Maingueneau, 2024 e 2025) corresponde, na perspectiva aqui proposta, ao movimento pelo qual um sujeito mobiliza formas de dizer associadas a outras práticas enunciativas, reinscrevendo-as em novas condições de produção e circulação.

A aderência (Maingueneau, 2025), por sua vez, pode ser compreendida provisoriamente como o processo pelo qual determinados elementos discursivos favorecem o reconhecimento e a aceitação de uma posição enunciativa pelo coenunciador, assumindo novos sentidos, no caso dos *phishing* há a apropriação do discurso de autoridade institucional e consideramos que o slogan, o formato aderem ao enunciado *fake* e o meio cria um simulacro que permite uma cena de enunciação legítima.

Não se trata, neste momento, de apresentar a aderência como categoria teórica definitivamente estabelecida nos mesmos moldes apresentados por Maingueneau (2025), mas de considerar sua produtividade analítica para compreender práticas nas quais a legitimidade do enunciado depende da correspondência entre a cenografia construída e as expectativas socialmente compartilhadas.

No contexto da IAG, essa questão adquire importância porque a tecnologia pode facilitar a produção de enunciados adequados a diferentes gêneros, mas não controla integralmente as interpretações realizadas pelos sujeitos e o sentido permanece relacionado às “condições de produção, circulação e recepção dos discursos”.

Do ponto de vista da LF, esse deslocamento exige considerar que a autoria de uma mensagem não se confunde necessariamente com a identidade construída em sua enunciação, mas um somatório de características idioletais e tecnodiscursivas.

Um discurso pode apresentar uma instituição como enunciativa, ser produzido materialmente com auxílio de IAG e circular por decisão de um sujeito que não pertence à instituição representada.

Assim, entendemos que a produção textual mediada por IAG torna mais complexa a relação entre o sujeito que aparece no discurso, aquele que participa de sua elaboração e aquele que assume sua circulação, sendo intermediados por algoritmos responsáveis pela engenharia social.

Não estamos, portanto, diante do desaparecimento da autoria, mas de uma redistribuição possível das operações que participam da construção e da utilização dos enunciados.

Essa redistribuição exige novas reflexões teórico-metodológicas, especialmente quando a linguagem passa a integrar situações de investigação criminal.

5. Conclusões

A utilização da Inteligência Artificial Generativa em práticas sociais contemporâneas tem ampliado as possibilidades de produção e circulação de enunciados, introduzindo questões

importantes para os estudos da linguagem, principalmente no que concerne à autoria, à construção de identidades e à responsabilização por práticas discursivas.

Ao longo deste estudo, procuramos demonstrar que a aproximação entre AD LF e Ciência da Computação pode contribuir para compreender determinados mecanismos de simulação identitária presentes em mensagens relacionadas à engenharia social como a imagem dos sujeitos (*ethé*) e a autoria ou autoralidade.

A análise permitiu observar que a construção da legitimidade discursiva não depende exclusivamente da identificação explícita do enunciador, mas também das condições pelas quais determinadas formas de dizer são reconhecidas socialmente como pertencentes a instituições ou sujeitos e enredadas no e pelo ambiente digital.

Nesse sentido, as cenas de enunciação e o *ethos* discursivo apresentam-se como categorias relevantes para a interpretação de enunciados que simulam relações de autoridade, obrigação ou proximidade no processo *transformer* de arquitetura da IAG.

Entretanto, a análise também demonstra que a interpretação dessas estratégias não corresponde à identificação de autoria nem permite comprovar a participação da IAG na produção de determinado texto, mas permite observar um simulacro de autoralidade.

Esse aspecto merece destaque porque a presença de sistemas generativos pode provocar a tendência de atribuir automaticamente a essas tecnologias a responsabilidade por materialidades linguísticas que apresentam características semelhantes às suas produções.

Tal procedimento desconsidera que os sujeitos humanos também mobilizam padrões linguísticos, reproduzem formas de dizer e utilizam recursos tecnológicos diversos na elaboração de textos.

Assim, entendemos que a discussão sobre autoria precisa considerar não apenas a materialidade verbal, mas os processos de produção, apropriação e circulação discursiva.

A proposta de compreender as relações entre apropriação e aderência constitui uma possibilidade de desenvolvimento teórico para investigações posteriores, especialmente quando houver materiais empíricos que permitam examinar como determinadas formas de dizer são reconhecidas e legitimadas pelos coenunciadores.

A relação entre discurso e interpretação permanece atravessada pelas condições sociais, históricas e interacionais que constituem os sujeitos.

Por isso, não consideramos que a IAG elimine a autoria humana. Ao contrário, sua utilização evidencia a necessidade de distinguir diferentes níveis de participação na produção dos enunciados, especialmente quando esses textos integram práticas potencialmente ilícitas.

A contribuição da AD para a LF encontra-se, portanto, na possibilidade de compreender como os discursos constroem lugares de enunciação, simulam identidades e mobilizam relações de poder que ultrapassam a materialidade linguística.

A LF, por sua vez, pode oferecer procedimentos de investigação que permitam examinar questões relacionadas à origem, à comparação e à atribuição de autoria textual.

Ressaltamos, finalmente, que o caráter experimental do *corpus* não permite generalizar resultados para situações reais de cibercrimes. Uma etapa posterior deverá incluir materiais documentados, informações sobre as condições de circulação e, sempre que possível, registros que permitam verificar a intervenção de sistemas de IAG.

Dessa maneira, mais do que identificar se um texto foi produzido por uma máquina, interessa compreender como a apropriação de enunciados pode construir simulacros de autoria, produzir efeitos de legitimidade e participar de práticas discursivas com possíveis consequências jurídicas.

É nessa articulação entre linguagem, tecnologia, autoria e responsabilidade que situamos a contribuição possível da Análise do Discurso para os estudos contemporâneos em Linguística Forense.

Link para Preprint

[INSERIR LINK OU DOI APÓS DEPÓSITO EM REPOSITÓRIO DE PREPRINTS QUE ACEITE COMENTÁRIOS PÚBLICOS.]

Conflito de Interesse

Declaramos a quem possa interessar que não há conflitos de interesses.

Protocolo e Pré-Registro de Pesquisa

Não se aplica.

Declaração de Disponibilidade de Dados

O *corpus* analisado consiste em materiais divulgados publicamente pela *Microsoft* no relatório “Unmasking EvilTokens: Getting to the root of device code phishing”, disponível em <https://www.microsoft.com/en-us/security/blog/2026/09/22/unmasking-eviltokens-getting-to-the-root-of-device-code-phishing/>.

Declaração de Uso de Inteligência Artificial

Os autores utilizaram o *ChatGPT* para uma busca exploratória de casos de *phishing* e para auxiliar na delimitação inicial do *corpus*, conforme os *prompts* e registros apresentados na seção Metodologia. As informações e os recortes selecionados foram confrontados com o relatório público da *Microsoft*.

REFERÊNCIAS

- BOMMASANI, Rishi *et al.* **On the opportunities and risks of foundation models**. ARXIV, 2021. Disponível em: <https://arxiv.org/abs/2108.07258>. Acesso em: 20 out. 2026.
- BOTELHO, Laura; CARREIRA, Rosângela Aparecida Ribeiro; SILVA, Thiago Evangelista. **Percepções de alunos ingressantes no curso de Letras sobre práticas de leitura: reflexões sobre letramentos acadêmicos e formação docente**. In: BOTELHO, Laura; VIANINI, Carolina (org.). *Letramentos e ensino: reflexões a partir da linguística aplicada*. Campinas: Mercado de Letras, 2024. p. 21-46
- BROWN, Tom B. *et al.* **Language models are few-shot learners**. *Advances in Neural Information Processing Systems*, v. 33, p. 1877-1901, 2020.
- CARREIRA, Rosângela Aparecida Ribeiro. **É preciso falar sobre o lugar do dizer: a múltipla perspectiva (para)tópica**. In: NASCIMENTO, Jarbas Vargas; CANO, Márcio Rogério de Oliveira; ELIAKIM, Jonatas (org.). *Paratopia*. São Paulo: Blucher, 2020. v. 3, p. 170-202. Disponível em: <https://openaccess.blucher.com.br/article-list/9786555500325-482/list/>. Acesso em: 20 out. 2026.
- CARREIRA, Rosângela Aparecida Ribeiro e NASCIMENTO, Jarbas Vargas. **Autoria e autoralidade: o movimento paratópico de um pseudônimo**. *Caderno de Letras, Pelotas*, n. 36, p. 127-139, jan./abr. 2020. DOI: <https://doi.org/10.15210/cdl.v0i36.17197>.
- COLEMAN, Gabriella. **Hacker, hoaxer, whistleblower, spy: the many faces of Anonymous**. London: Verso, 2014.
- COULTHARD, Malcolm; JOHNSON, Alison. **An introduction to forensic linguistics: language in evidence**. London: Routledge, 2007.
- EUROPOL. **ChatGPT: the impact of large language models on law enforcement**. Luxembourg: Publications Office of the European Union, 2023. DOI: <https://doi.org/10.2813/255453>.
- FOUCAULT, Michel. **A ordem do discurso**. São Paulo: Loyola, 2001.
- FOUCAULT, Michel. **O que é um autor?**. In: FOUCAULT, Michel. *Ditos e escritos III: estética — literatura e pintura, música e cinema*. Rio de Janeiro: Forense Universitária, 2009.
- FOUCAULT, Michel. **A arqueologia do saber**. Tradução de Luiz Felipe Baeta Neves. 8. ed. Rio de Janeiro: Forense Universitária, 2012.
- FRASER, Antonia. **Faith and treason: the story of the Gunpowder Plot**. New York: Doubleday, 1996.
- JAWAHAR, Ganesh; ABDUL-MAGEED, Muhammad; LAKSHMANAN, Laks V. S. **Automatic detection of machine generated text: a critical survey**. In: INTERNATIONAL CONFERENCE ON COMPUTATIONAL LINGUISTICS, 28., 2020, Barcelona. *Proceedings [...]*. Barcelona: International Committee on Computational Linguistics, 2020. p. 2296-2309.
- LLOYD, David. **I do not have a name: David Lloyd on the V for Vendetta mask**. The Quietus, 2011. Disponível em: <https://thequietus.com/interviews/david-lloyd-v-for-vendetta-mask/>. Acesso em: 20 out. 2026.
- MAINGUENEAU, Dominique. **Ethos, cenografia, incorporação**. In: AMOSSY, Ruth (org.). *Imagens de si no discurso: a construção do ethos*. São Paulo: Contexto, 2005. p. 69-92.
- MAINGUENEAU, Dominique. **A propósito do ethos**. In: MOTTA, Ana Raquel; SALGADO, Luciana (org.). *Ethos discursivo*. São Paulo: Contexto, 2008a. p. 11-29.
- MAINGUENEAU, Dominique. **Cenas da enunciação**. Organização de Sírio Possenti e Maria Cecília Pérez de Souza-e-Silva. São Paulo: Parábola Editorial, 2008b.
- MAINGUENEAU, Dominique. **Gênese dos discursos**. Tradução de Sírio Possenti. São Paulo: Parábola Editorial, 2008c.
- MAINGUENEAU, Dominique. **Doze conceitos em análise do discurso**. Organização de Sírio Possenti e Maria Cecília Pérez de Souza-e-Silva. São Paulo: Parábola Editorial, 2010.
- MAINGUENEAU, Dominique. **Discurso e análise do discurso**. Tradução de Sírio Possenti. São Paulo: Parábola Editorial, 2015.
- MAINGUENEAU, Dominique. **Análise de textos de comunicação**. Tradução de Maria Cecília Pérez de Souza-e-Silva e Décio Rocha. São Paulo: Cortez, 2024.
- MAINGUENEAU, Dominique. **As margens do discurso**. Organização de Nelson Barros da Costa, Maria das Dores Mendes e José Wesley Matos. São Paulo: Contexto, 2025.
- MICROSOFT. **Disrupting EvilTokens: the AI chatbot built for cybercrime**. Microsoft On the Issues, 22 set. 2026. Disponível em: <https://blogs.microsoft.com/on-the-issues/2026/09/22/disrupting-eviltokens-the-ai-chatbot-built-for-cybercrime/>. Acesso em: 20 out. 2026.
- MICROSOFT THREAT INTELLIGENCE; MICROSOFT DEFENDER EXPERTS; MICROSOFT SECURITY RESEARCH. **Unmasking EvilTokens: getting to the root of device code phishing**. Microsoft Security Blog, 22 set. 2026. Disponível em: <https://www.microsoft.com/en-us/security/blog/2026/09/22/unmasking-eviltokens-getting-to-the-root-of-device-code-phishing/>. Acesso em: 15 out. 2026.
- MIRSKY, Yisroel; LEE, Wenke. **The creation and detection of deepfakes: a survey**. *ACM Computing Surveys*, v. 54, n. 1, 2021.
- MOORE, Alan; LLOYD, David. **V for Vendetta**. New York: DC Comics, 2005.
- MOORE, Alan; LLOYD, David. **V for Vendetta: 30th Anniversary Deluxe Edition**. Burbank: DC Comics, 2018.
- PAVEAU, Marie-Anne. **Os pré-discursos: sentido, memória, cognição**. Campinas, SP: Pontes Editores, 2013.

- PAVEAU, Marie-Anne. **Análise do discurso digital: dicionário das formas e das práticas**. Organização da tradução de Julia Lourenço Costa e Roberto Leiser Baronas. Campinas, SP: Pontes Editores, 2021.
- PEREZ, Fábio; RIBEIRO, Ian. **Ignore previous prompt: attack techniques for language models**. arXiv, 2022. Disponível em: <https://arxiv.org/abs/2211.09527>. Acesso em: 20 out. 2026.
- SCHMITT, Marc; FLECHAIS, Ivan. **Digital deception: generative artificial intelligence in social engineering and phishing**. Artificial Intelligence Review, v. 57, artigo 324, 2024. DOI: <https://doi.org/10.1007/s10462-024-10973-2>.
- SENE JUNIOR, I. G. et al. **Aplicação de Inteligência Artificial Generativa em Precedentes: identificação de similaridades entre casos**. [S. l.: s. n.], [s. d.]. Manuscrito acadêmico. Referência provisória.
- STREET, Brian V. **Letramentos sociais: abordagens críticas do letramento no desenvolvimento, na etnografia e na educação**. Tradução de Marcos Bagno. São Paulo: Parábola Editorial, 2014.
- SOUSA-SILVA, Rui. **From detection to attribution: forensic linguistics and adversarial red teaming as complementary responses to LLM misuse**. In: INTERNATIONAL CONFERENCE ON NATURAL LANGUAGE PROCESSING AND ARTIFICIAL INTELLIGENCE FOR CYBER SECURITY, 2., 2026, Alicante. Proceedings [...]. Alicante: Department of Languages and Information Systems, University of Alicante, 2026. p. 122-133. Disponível em: <https://aclanthology.org/2026.nlpaics-1.13/>. Acesso em: 12 out. 2026.
- ROJO, Roxane. **Entre plataformas, ODAs e protótipos: novos multiletramentos em tempos de Web2**. The ESpecialist: Descrição, Ensino e Aprendizagem, São Paulo, v. 38, n. 1, 2017.
- SOUSA-SILVA, Rui; COULTHARD, Malcolm. **Linguística Forense**. In: DINIS-OLIVEIRA, Ricardo Jorge; MAGALHÃES, Teresa (org.). *O que são as Ciências Forenses? Conceitos, abrangência e perspectivas futuras*. Lisboa: Pactor, 2016. p. 137-144.
- VASWANI, Ashish et al. **Attention is all you need**. Advances in Neural Information Processing Systems, v. 30, 2017.
- WEIDINGER, Laura et al. **Taxonomy of risks posed by language models**. In: ACM CONFERENCE ON FAIRNESS, ACCOUNTABILITY, AND TRANSPARENCY, 2022. Proceedings [...]. New York: ACM, 2022. p. 214-229.

Este preprint foi submetido sob as seguintes condições:

- Os autores declaram que os necessários Termos de Consentimento Livre e Esclarecido de participantes ou pacientes na pesquisa foram obtidos e estão descritos no manuscrito, quando aplicável.
- Os autores declaram que a elaboração do manuscrito seguiu as normas éticas de comunicação científica.
- Os autores declaram que estão cientes que são os únicos responsáveis pelo conteúdo do preprint e que o depósito no SciELO Preprints não significa nenhum compromisso de parte do SciELO, exceto sua preservação e disseminação.
- Os autores declaram que os dados, aplicativos e outros conteúdos subjacentes ao manuscrito estão referenciados.
- O manuscrito depositado está no formato PDF.
- Os autores declaram que a pesquisa que deu origem ao manuscrito seguiu as boas práticas éticas e que as necessárias aprovações de comitês de ética de pesquisa, quando aplicável, estão descritas no manuscrito.
- Os autores declaram que uma vez que um manuscrito é postado no servidor SciELO Preprints, o mesmo só poderá ser retirado mediante pedido à Secretaria Editorial do SciELO Preprints, que afixará um aviso de retratação no seu lugar.
- Os autores concordam que o manuscrito aprovado será disponibilizado sob licença [Creative Commons CC-BY](#).
- O autor submissor declara que as contribuições de todos os autores e declaração de conflito de interesses estão incluídas de maneira explícita e em seções específicas do manuscrito.
- Os autores declaram que o manuscrito não foi depositado e/ou disponibilizado previamente em outro servidor de preprints ou publicado em um periódico.
- Caso o manuscrito esteja em processo de avaliação ou sendo preparado para publicação mas ainda não publicado por um periódico, os autores declaram que receberam autorização do periódico para realizar este depósito.
- O autor submissor declara que todos os autores do manuscrito concordam com a submissão ao SciELO Preprints.