



25º Encontro Nacional de Ensino de Jornalismo

ENEJOR

O ENSINO DE JORNALISMO E A CRISE CLIMÁTICA

Brasília-DF

22,23 E 24 DE ABRIL

2026

PROMOÇÃO

REALIZAÇÃO

FOMENTO

APOIO



RELATO DE EXPERIÊNCIA

MINERAÇÃO E CURADORIA DE DADOS SOBRE QUEIMADAS NO CERRADO: UM RELATO DE EXPERIÊNCIA NO PROJETO FOGO NA FAKE

Santos, Arthur; arthur-ferreira.af@aluno.unb.br

Santos, Marcio; marcio.carneiro@ufma.br (orientador)

Moraes, Lauro; lauro.moraes@unb.br (coorientador)

Moura, Dione; moura@unb.br (Coordenadora do Projeto Associado 5 - Engajamento Público com a Ciência)

Colli, Guarino; grcolli@unb.br (Coordenador Geral da Rede Biota Cerrado)

RESUMO

Este relato de experiência detalha o fluxo de coleta, limpeza e curadoria técnica de dados midiáticos sobre o fogo no bioma Cerrado, oriundo dos resultados do projeto de pesquisa e extensão “Fogo na Fake: Lidando com a Desinformação sobre Queimadas no Cerrado”, da Rede Biota Cerrado (RBC), desenvolvido com apoio da Universidade de Brasília (Edital DEX/DEG/DGP/DPI nº 01/2025). A metodologia integrou a SerpApi à plataforma n8n para a extração automatizada de registros entre 2004 e 2025. O processo técnico envolveu a transição de formatos JSON para CSV e o desenvolvimento de uma pipeline em Python para extração de conteúdo textual e verificação de acessibilidade. Os resultados indicam que, de um universo de 27.190 registros brutos, apenas 5.097 foram validados, resultando em um descarte deliberado de 81,25% dos dados. Tal volume de ruído evidencia a complexidade da filtragem semântica necessária para o monitoramento da desinformação ambiental em larga escala.

PALAVRAS-CHAVE

Mineração de Dados. Cerrado. Queimadas. Desinformação. Curadoria de Dados.



25º Encontro Nacional de Ensino de Jornalismo

ENEJOR

O ENSINO DE JORNALISMO E A CRISE CLIMÁTICA

Brasília-DF

22,23 E 24 DE ABRIL

2026

PROMOÇÃO

REALIZAÇÃO

PÓS-GRADUAÇÃO

APOIO



1. INTRODUÇÃO

O projeto "Fogo na Fake", filiado à Rede Biota Cerrado¹ (UnB), surge como uma resposta interdisciplinar à necessidade de monitorar e combater a desinformação ambiental. Sob a coordenação do Prof. Dr. Márcio Carneiro dos Santos, o Grupo de Mineração atuou na arquitetura de fluxos informacionais capazes de lidar com a velocidade, a variedade e o volume das narrativas digitais. O foco geográfico é o Cerrado brasileiro, bioma que ocupa aproximadamente 23% do território brasileiro e que, historicamente, sofre pressão antrópica decorrente do uso irregular do fogo para a expansão de fronteiras agrícolas. Neste contexto, sendo o bioma estrategicamente vital para a segurança hídrica e a biodiversidade, a circulação de informações sobre queimadas é frequentemente obstruída por processos de desinformação.

[...] o Cerrado está se tornando progressivamente mais vulnerável a cenários de aridificação e desequilíbrio hídrico. Essa vulnerabilidade não é apenas climática, mas também socioambiental, pois repercute sobre setores essenciais como agricultura, abastecimento urbano, energia e saúde pública. (Ávila; Bolzam, 2025)

A desinformação, termo utilizado para designar informações falsas, enganosas ou distorcidas, divulgadas com a intenção deliberada de manipular percepções, influenciar comportamentos ou comprometer decisões (Merriam-Webster, 2025), possui impacto que vai além da simples circulação de inverdades: ele afeta diretamente a formação da opinião pública, fragiliza políticas públicas e retarda respostas a problemas urgentes (Dryzek *et al.*, 2011; Lewandowsky *et al.*, 2013). Utilizando como exemplo os contextos ambientais, a desinformação cria obstáculos à conservação e à gestão sustentável dos recursos naturais, ao mesmo tempo em que favorece interesses econômicos específicos e discursos negacionistas (Dryzek *et al.*, 2011; Supran, Oreskes, 2017). Lidar com esse tipo de conteúdo é, portanto, crucial



25º Encontro Nacional de Ensino de Jornalismo

ENEJOR

O ENSINO DE JORNALISMO E A CRISE CLIMÁTICA

Brasília-DF

22,23 E 24 DE ABRIL

2026



para viabilizar ações eficazes de mitigação das mudanças climáticas e preservação da biodiversidade.

Sua versão moderna caracteriza-se por narrativas emocionalmente apelativas, uso de espantalhos argumentativos, como “investir em energia limpa vai quebrar o país” ou “como existe aquecimento global se ontem estava frio?”, linguagem alarmista, imagens descontextualizadas, testemunhos anônimos e citações de supostos especialistas sem credibilidade. Esses elementos reforçam sua capacidade de persuasão e viralização nas redes sociais (Diethelm, McKee, 2009; Cook, 2020).

Cook (2017) destaca que a desinformação não é apenas um erro teórico, mas um fenômeno com consequências reais e letais. Um exemplo histórico é a negação do consenso científico de que o HIV causa a AIDS, responsável por políticas públicas equivocadas na África do Sul que contribuíram para aproximadamente 330 mil mortes entre 2000 e 2005. A hesitação vacinal, impulsionada por alegações infundadas, também colocou populações inteiras em risco de doenças evitáveis. Cook (2017) enfatiza ainda o efeito backfire da cosmovisão, no qual evidências corretivas que contradizem crenças prévias reforçam, paradoxalmente, essas mesmas crenças, tornando a correção da desinformação uma tarefa complexa.

As pesquisas de Lewandowsky *et al.* (2012, 2013) e de Supran e Oreskes (2017) aprofundam o entendimento sobre a relação entre ciência, percepção pública e comunicação. Lewandowsky *et al.* (2012) demonstram que o reconhecimento de um consenso científico é decisivo para aumentar a aceitação das evidências científicas. Quando o público percebe que os especialistas convergem em suas conclusões, tende a confiar mais na ciência e a reduzir resistências ideológicas.

Essa estrutura é mantida por indústrias poluidoras, *think tanks* conservadores, políticos e setores da mídia que utilizam narrativas ideológicas para moldar a opinião pública. Estudos apontam que homens brancos conservadores tendem a apresentar maior propensão a negar as mudanças climáticas, movidos pela



25º Encontro Nacional de Ensino de Jornalismo

ENEJOR

O ENSINO DE JORNALISMO E A CRISE CLIMÁTICA

Brasília-DF

22,23 E 24 DE ABRIL

2026



defesa de sua identidade social e pela manutenção de valores políticos e econômicos (McCright, Dunlap, 2011). Compreender quem produz, recebe e reproduz a desinformação é, portanto, essencial para formular estratégias de comunicação mais eficazes, capazes de enfrentar discursos negacionistas.

Hansson (2017) complementa ao classificar o negacionismo científico, que inclui o negacionismo climático e o evolucionário, como uma das principais formas de pseudociência. Segundo o autor, negacionistas utilizam estratégias como a seleção seletiva de evidências (*cherry-picking*) para sustentar conclusões desejadas, ignorando deliberadamente o conjunto total de provas disponíveis. As raízes históricas da desinformação moderna remontam à década de 1950, quando a indústria do tabaco desenvolveu estratégias de manipulação da opinião pública para desacreditar descobertas científicas sobre os malefícios do cigarro. Essa tática, baseada em conflitos de interesse e engenharia da controvérsia, foi posteriormente replicada por grandes corporações, como a ExxonMobil, em campanhas que buscavam desacreditar as evidências do aquecimento global antropogênico (Supran, Oreskes, 2017). Tais práticas demonstram como a desinformação evoluiu de uma ferramenta de relações públicas para um instrumento sofisticado de controle narrativo, capaz de distorcer a percepção coletiva sobre a própria ciência.

O problema central reside na dificuldade de identificar discursos negacionistas e estratégias de manipulação sem o apoio de recursos computacionais intensivos. O monitoramento manual é insuficiente diante da escala da infosfera contemporânea. Assim, o objetivo deste relato é detalhar o fluxo de trabalho técnico para a coleta, limpeza e análise de dados midiáticos sobre o fogo no Cerrado ao longo de 20 anos (2004-2025), consolidando uma metodologia de ciência de dados aplicada à comunicação científica, que se estendeu até a conclusão técnica em fevereiro de 2026.



25º Encontro Nacional de Ensino de Jornalismo

ENEJOR

O ENSINO DE JORNALISMO E A CRISE CLIMÁTICA

Brasília-DF

22,23 E 24 DE ABRIL

2026



2. REFERENCIAL TEÓRICO (A Desinformação como Obstáculo à Governança Climática)

A base teórica deste trabalho fundamenta-se na compreensão de que a desinformação ambiental é uma estratégia deliberada de manipulação das percepções (Merriam-Webster, 2025). Conforme apontam Cook (2017) e Lamb *et al.* (2020), não lidamos apenas com notícias falsas pontuais, mas também com "discursos de atraso" que visam fragilizar políticas públicas e retardar as respostas à emergência climática.

No ensino de jornalismo, é imperativo incorporar conceitos como a "Cascata de Rumores". Como observado, a viralidade estrutural de uma informação falsa no ambiente digital é potencializada pela "Novidade" (*Novelty*). Notícias enganosas tendem a ser mais inusitadas, o que aumenta a atenção e o valor social do compartilhamento. Portanto, o jornalismo contemporâneo precisa de ferramentas que identifiquem esses padrões semânticos e temporais antes que a cascata de rumores se torne incontrolável.

3. METODOLOGIA E PROCEDIMENTOS TÉCNICOS

A metodologia aplicada foi estruturada para ser auditável e escalável, dividindo-se em quatro fases críticas:

3.1 Fluxo de Coleta e Tratamento de Dados

A coleta de dados brutos foi operacionalizada por meio da SerpApi, integrada ao software de automação n8n. Para garantir a captura de resultados pertinentes, foram estruturadas queries booleanas, como (fogo OR incêndio OR queimada) AND (cerrado OR savana), e realizadas buscas quinzenais automatizadas no Google Search. A recepção inicial dos dados ocorreu em formato JSON, contendo metadados brutos e trechos de pré-visualização (snippets).

O tratamento técnico envolveu a conversão de JSON para CSV por meio de um node de código dedicado no n8n. Posteriormente, os dados foram exportados para o



25º Encontro Nacional de Ensino de Jornalismo

ENEJOR

O ENSINO DE JORNALISMO E A CRISE CLIMÁTICA

Brasília-DF

22,23 E 24 DE ABRIL

2026



Google Sheets para organização inicial em lotes anuais, o que facilitou a gestão dos registros pela equipe de pesquisadores.

3.2 Pipeline de Extração em Python

Para o refinamento do dataset, desenvolveu-se uma pipeline em Python utilizando as bibliotecas Requests, BeautifulSoup e Streamlit. O script operou em três fases críticas:

- Limpeza Inicial: Remoção de links que apontavam para PDFs, imagens e domínios não textuais (YouTube, Instagram, etc.);
- Verificação de Acessibilidade: Execução de requisições de status (HTTP 200) para confirmar a atividade das URLs;
- Processamento em Lotes (Batch Processing): A extração do conteúdo textual das páginas foi realizada em lotes de 1.000 URLs, estratégia adotada para garantir a estabilidade do sistema, evitar a sobrecarga dos servidores e permitir o salvamento automático parcial dos resultados.

3.3 Taxonomia de Referência

A análise semântica subsequente foi orientada pelos modelos teóricos FLICC (Cook, 2020), que categoriza as técnicas de negação da ciência, e Discursos de Atraso (Lamb *et al.*, 2020), focados na inação climática. Essas taxonomias serviram de base de conhecimento para que modelos de Inteligência Artificial Generativa realizassem a filtragem e a categorização dos discursos.

4. DESENVOLVIMENTO: LIMPEZA, ANÁLISE E CURADORIA DE DADOS

4.1 Análise Quantitativa da Limpeza

O processo de curadoria técnica resultou em uma redução drástica do volume inicial de dados, conforme as métricas consolidadas abaixo:

Tabela 1: Tabela comparando quantidade de linhas por período

Período	Marcadas	Limpas	Diferença	Total de Linhas
2004 a 2006	2137	94	2043	2231



25º Encontro Nacional de Ensino de Jornalismo

ENEJOR

O ENSINO DE JORNALISMO E A CRISE CLIMÁTICA

Brasília-DF

22,23 E 24 DE ABRIL

2026



Período	Marcadas	Limpas	Diferença	Total de Linhas
2007 a 2010	3928	266	3662	4194
2011 a 2014	4593	487	4106	5080
2015 a 2017	3414	627	2787	4041
2018 a 2020	2384	962	1422	3346
2021 a 2023	3265	1304	1961	4569
2024 a 2025	2372	1357	1015	3729
TOTAL	22093	5097	16996	27190

Fonte: do autor (2026)

4.2 Eficiência da Curadoria e Estratégia de Captura

A curadoria resultou na seleção de apenas **5.097 linhas**, o que representa um descarte de **81,25%** do material bruto. Apenas **18,75%** do conteúdo foi considerado relevante. É imperativo destacar que esse alto índice de descarte decorreu de uma estratégia deliberada de "**broad-capture**" (captura ampla). A equipe identificou que o aumento da especificidade das palavras-chave reduzia a coleta de resultados pertinentes; optou-se, então, por um volume inicial massivo, seguido de um processamento técnico refinado em pós-coleta.



25º Encontro Nacional de Ensino de Jornalismo

ENEJOR

O ENSINO DE JORNALISMO E A CRISE CLIMÁTICA

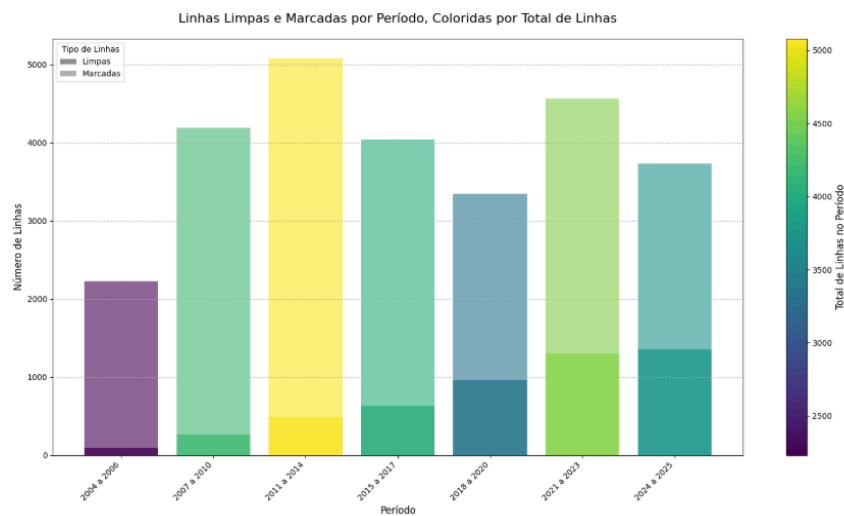
Brasília-DF

22,23 E 24 DE ABRIL

2026



Gráfico 1: Gráfico que mostra a quantidade total de linhas antes e após o processamento.



Fonte: do autor (2026)

4.3 Qualidade dos Dados e Ruído Semântico

O ruído detectado incluiu padrões de erro sistemáticos, tais como:

- Receitas culinárias: instruções de preparo que mencionavam o termo "fogo";
- Comércio especializado: lojas de extintores e produtos de proteção contra incêndio;
- Ambiguidade Linguística: Ocorrências do termo "cerrado" em espanhol, que significa "fechado" ou "cercado", sem relação com o bioma brasileiro.

A análise temporal demonstrou um crescimento do interesse midiático qualificado a partir de 2017, embora o pico de coleta bruta tenha ocorrido entre 2011 e 2014, período de maior diversidade (e, conseqüentemente, menor qualidade) dos registros.



25º Encontro Nacional de Ensino de Jornalismo

ENEJOR

O ENSINO DE JORNALISMO E A CRISE CLIMÁTICA

Brasília-DF

22,23 E 24 DE ABRIL

2026



5. DISCUSSÃO DOS RESULTADOS E USO DE IAG

O uso de Inteligência Artificial Generativa (IAG), especificamente o modelo GPT-nano-5, atuou como um sistema de filtragem de alta precisão. A integração via RAG (Retrieval Augmented Generation) permitiu que o modelo operasse sobre a taxonomia curada, mitigando "alucinações" e garantindo rigor analítico. A IAG identificou que o *Cherry-picking* (~54%) foi a tática de desinformação predominante, seguida de Contexto Falso e de Alarmismo. Essas estratégias alinham-se à tendência do "Novo Negacionismo", que desloca o ataque à existência do aquecimento global para a inviabilização das soluções climáticas e ambientais. A limpeza rigorosa das planilhas foi o fator determinante para que a IAG traçasse o caminho temporal da informação, revelando como as narrativas são distorcidas até atingirem o usuário final.

Os dados mostram que o interesse qualificado da mídia pelo Cerrado cresceu efetivamente a partir de **2016**, com forte alta após **2022**, acompanhando a transformação da pauta científica em uma discussão sobre eventos tangíveis e trágicos.

A escolha por focar em texto em detrimento de áudio e vídeo (IAG multimodal) deve-se a limitações metodológicas e técnicas claras:

- **Escopo e Prazos:** O processamento de imagens e vídeos demandaria uma linha de processamento distinta, que não estava prevista no escopo e nos prazos originais do projeto;
- **Potencial de Experimentação:** O grupo optou por explorar os Modelos de Linguagem de Larga Escala (LLMs) por serem um campo mais recente e promissor para a detecção de **padrões semânticos e interpretação de nuances**, o que seria mais complexo de validar em larga escala com áudios e vídeos no tempo disponível;
- **Limpeza e Tratamento:** A etapa de limpeza de dados já se mostrou altamente consumidora de tempo; lidar com arquivos não textuais sobrecarregaria a infraestrutura de coleta e análise manual necessária para validar os resultados da IA;



25º Encontro Nacional de Ensino de Jornalismo

ENEJOR

O ENSINO DE JORNALISMO E A CRISE CLIMÁTICA

Brasília-DF

22,23 E 24 DE ABRIL

2026

PROMOÇÃO

REALIZAÇÃO

PÓS-GRADUAÇÃO

APOIO



- Gastos com ferramentas: A viabilidade técnica desses experimentos é diretamente impactada pela questão dos custos, fator que, somado às dificuldades de acesso, impediu a conclusão da coleta de dados em plataformas privadas como X e Instagram.

Com o objetivo de mitigar o impacto financeiro, o projeto implementou uma etapa de limpeza e curadoria manual dedicada, que reduziu significativamente o volume de dados brutos e, conseqüentemente, os custos de utilização da API para o processamento dos modelos.

Os resultados indicam que o ecossistema digital favorece a monetização da desinformação via infraestruturas de publicidade (AdTech), onde narrativas negacionistas geram lucro para as plataformas e seus proponentes, superando muitas vezes o alcance da ciência sóbria.

Alguns outros impedimentos relacionados ao ambiente externo foram:

- Barreiras em Redes Sociais: A coleta em plataformas como X (antigo Twitter) e Instagram não foi concluída devido às "grandes dificuldades atuais de acesso e custo" impostas por essas empresas privadas.
- Algoritmos de Recomendação: É crucial mencionar que a desinformação no ambiente digital é potencializada por algoritmos que operam sob a lógica de **velocidade, variedade e volume**, o que torna o monitoramento humano insuficiente e justifica o uso de recursos computacionais intensivos.

6. CONSIDERAÇÕES FINAIS

A experiência no projeto "Fogo na Fake" reforça que automações brutas não substituem a curadoria técnica e o refinamento humano. A metodologia de "broad-capture" demonstrou que a mineração de dados em biomas complexos, como o Cerrado, exige flexibilidade na coleta e rigor na limpeza. O desenvolvimento da pipeline de extração e do assistente Fake_Alert consolida uma metodologia promissora para o enfrentamento da desinformação em escala. O aprendizado técnico evidencia que a integridade da biosfera está intrinsecamente ligada à da infosfera; somente por meio de dados curados e auditáveis é possível promover uma conscientização ambiental eficaz e fundamentada em evidências científicas.

