



CIÊNCIAS SOCIAIS E APLICADAS
FEMIC JOVEM

Aquilis Alves de Melo Oliveira

Isabella dos Santos Caruso

Vítor Emanuel da Silva Rodrigues

Abraão Lima Sousa

Victor Eduardo Alves da Silva Carvalho

IFPI

Teresina, Piauí e Brasil



abraaolmsousa@gmail.com

Verifica AI: bot que usa inteligência artificial para detecção de fake news nas redes sociais



Apresentação



- O surgimento da internet e a evolução dos smartphones geraram grandes impactos no mundo moderno. Como resultado, vivemos na era digital, onde a maioria das pessoas está conectada por meio de celulares e tem amplo acesso às redes sociais.
- A desinformação é disseminada principalmente em ambientes virtuais, sobretudo nas redes sociais, onde conteúdos polêmicos tendem a viralizar rapidamente (DELMAZO, 2018). Situações como a pandemia e os períodos eleitorais demonstram que a desinformação distorce a realidade, intensifica a polarização, provoca pânico e leva a sociedade a decisões equivocadas que afetam diretamente a saúde, a segurança, as finanças e até mesmo a democracia.

Objetivos

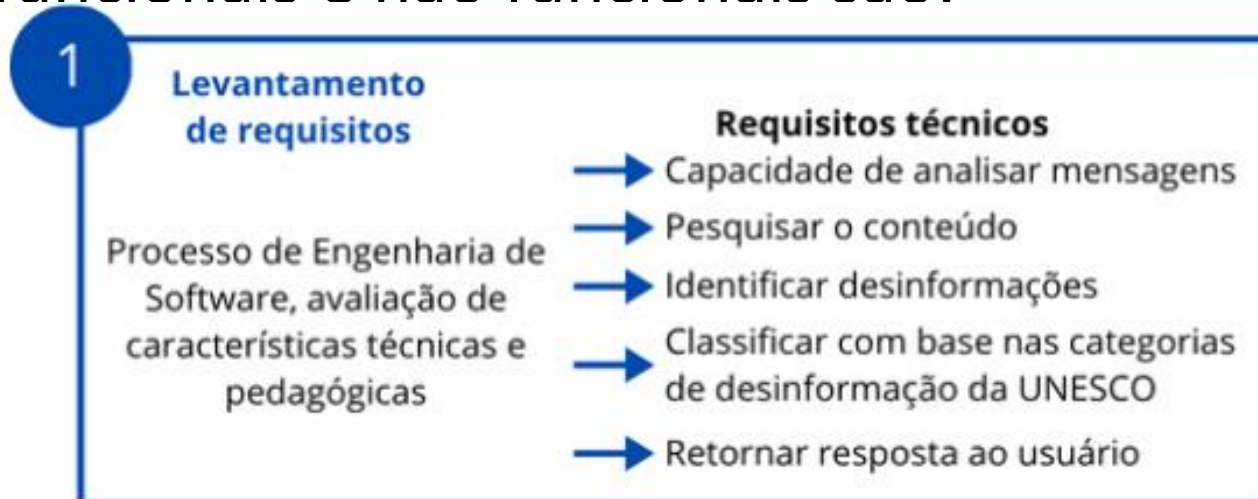


- Desenvolver uma ferramenta que combata ativamente a desinformação ao verificar, dentro da própria rede social, a veracidade de conteúdos publicados (posts e mensagens).
- **Objetivos Específicos**
 - Selecionar a rede social alvo e configurar as permissões oficiais necessárias para envio e recebimento de mensagens diretas (DM);
 - Desenvolver o *backend* para processar conteúdos multimodais, convertendo áudio, vídeo e imagem em texto para análise;
 - Integrar LLM do *Gemini* ao algoritmo, habilitando a análise com inteligência artificial;
 - Projetar e ajustar *prompts* com finalidade de identificar e classificar desinformação;
 - Adotar as diretrizes da UNESCO para classificar dentre as sete categorias;
 - Avaliar a confiabilidade e desempenho do protótipo através da matriz de confusão;

Metodologia



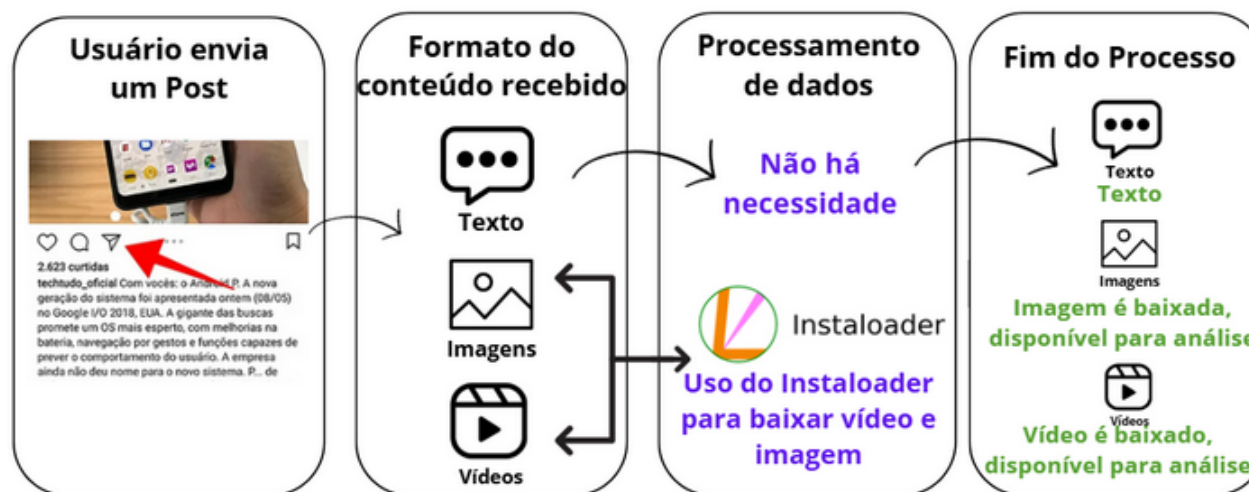
- Com base na análise de soluções já existentes, buscou-se desenvolver uma ferramenta que adota uma abordagem semelhante à do Grok, mas adaptada para o Instagram. A solução proposta consiste em um agente de IA integrado a um perfil da rede social.
- Os requisitos funcionais e não funcionais são:



Metodologia



- A integração com o Instagram foi estabelecida por meio das ferramentas da *Meta for Developers*, possibilitando a recepção e o envio de mensagens diretas (DM) entre o perfil do projeto e os usuários. Permitindo o sistema para receber, em tempo real, o conteúdo compartilhado pelos usuários.

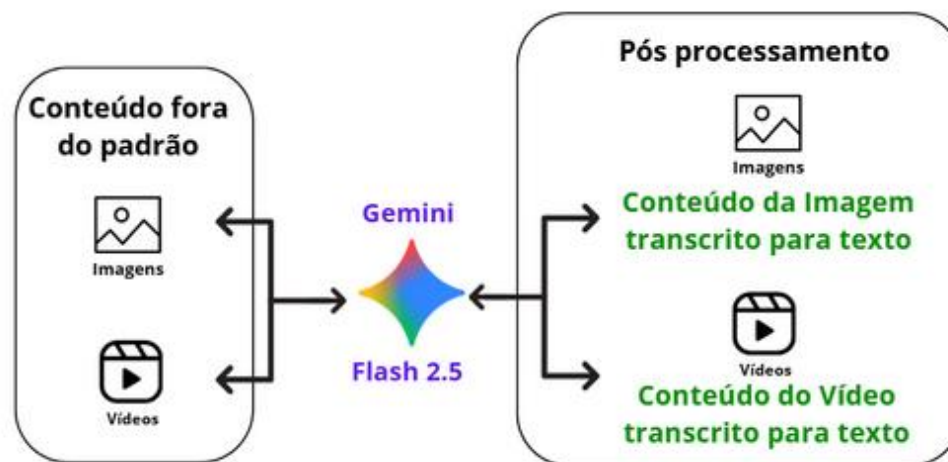


Fonte: Próprio Autor (2025)

Metodologia



- Nesta etapa, mesmo após o processamento inicial, três tipos de conteúdo (texto, imagem e vídeo) foram identificados. O passo seguinte consistiu em padronizar todo o material para o formato de texto utilizando o *Gemini Flash 2.5*. Reconhecimento óptico de caracteres para imagens e transcrição de áudio e análise de vídeos para vídeos.

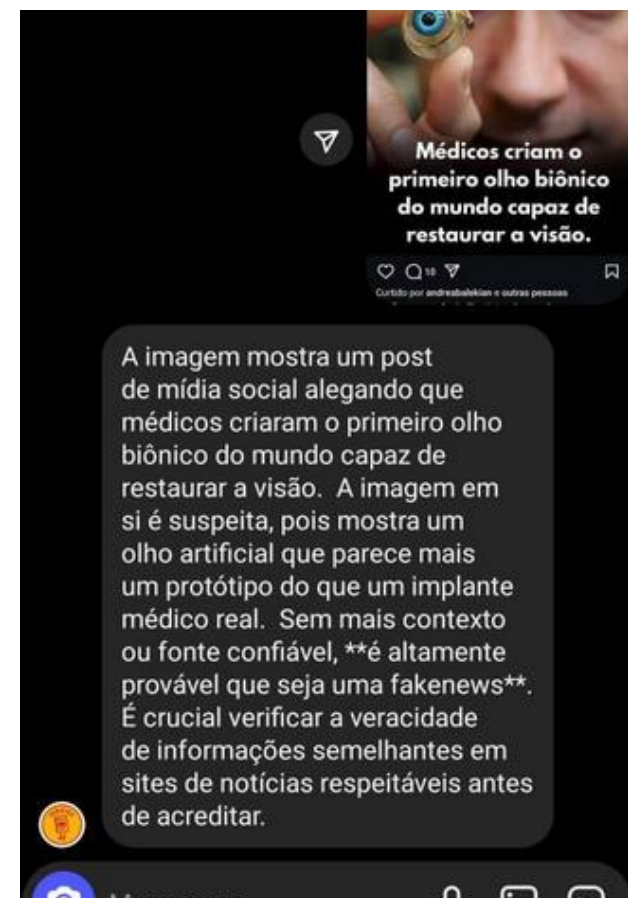


Fonte: Próprio Autor (2025)

Metodologia



- Testes preliminares foram conduzidos para verificar a integração do *Gemini* com o perfil do Instagram.
- A etapa de integração do *Gemini Flash 2.5* foi confirmada como bem-sucedida por meio de uma das respostas obtidas.



Fonte: Próprio Autor (2025)

Metodologia



- O *Gemini* pode analisar o conteúdo, mas o modelo gratuito *Flash 2.5* não possui conectividade direta com a internet.
- Para possibilitar a verificação do conteúdo com fontes recentes, o algoritmo submete o conteúdo analisado a uma busca em múltiplas fontes relacionadas ao tema na internet, reunindo evidências para validação.



Pesquisa Web
sobre o conteúdo

Fonte: Próprio Autor (2025)

Metodologia



- Em seguida, por meio de *prompts* previamente definidos, os resultados da pesquisa são comparados com o conteúdo original, permitindo concluir se a notícia constitui ou não uma desinformação.



Comparar o conteúdo
original e pesquisa

Fonte: Próprio Autor (2025)

Metodologia



- Se a probabilidade de o conteúdo ser classificado como desinformação for no mínimo de 95%, o sistema o categoriza conforme uma das sete categorias recomendadas pelo Manual da UNESCO:
- Sátira e paródia
- Conexão falsa
- Conteúdo enganoso
- Contexto falso
- Conteúdo impostor
- Conteúdo manipulado
- Conteúdo fabricado



Classificar a
desinformação de
acordo com UNESCO

Fonte: Próprio Autor (2025)

Resultados alcançados



- O protótipo desenvolvido está disponível aos usuários do Instagram por meio do perfil @verifica_ai_. Sua arquitetura foi pensada para que o usuário viva toda a experiência dentro da própria da própria plataforma da rede social.
- Ele opera de forma contínua, gratuita e sem a necessidade de instalação.



Fonte: Próprio Autor (2025)

Resultados alcançados



- O protótipo atendeu aos seguintes requisitos funcionais:
 - A comunicação é nativa e realizada via mensagens diretas (DM);
 - O sistema processa diferentes tipos de mídia, incluindo textos, imagens e vídeos.
 - O conteúdo recebido é comparado com fontes recentes, e a veracidade da informação é verificada.
 - Quando a desinformação é identificada, o caso é classificado em uma das sete categorias definidas pelo Manual da UNESCO.
 - A resposta ao usuário é fornecida em um tempo ágil, com média de 30 segundos.
 - As respostas são claras e objetivas, e incluem as fontes utilizadas na pesquisa para validação.

Resultados alcançados



- ENVIANDO NOTÍCIAS VERÍDICAS AO VERIFICA AI

< verifica_ai_



Verifica AI

0
posts

44
seguidores

15
seguindo

Usando Inteligência Artificial para conferir fake news

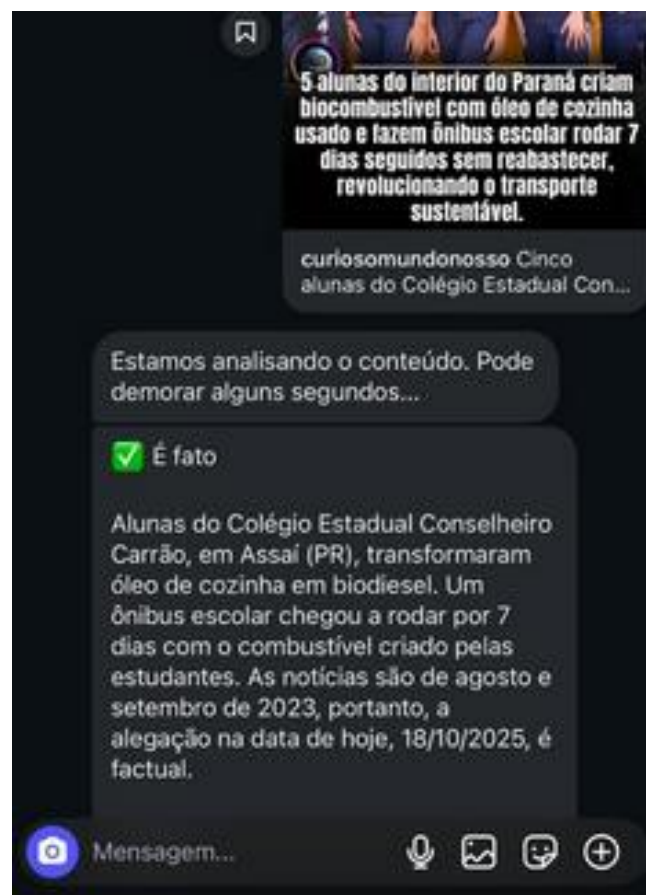


Seguido(a) por abraaollima07

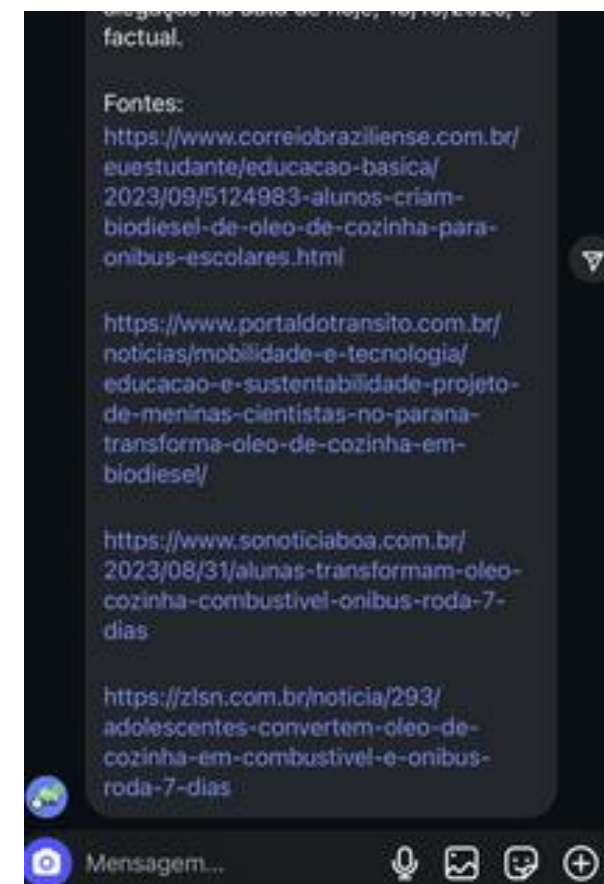
Seguir

Mensagem

Fonte: Próprio Autor (2025)



Fonte: Próprio Autor (2025)



Fonte: Próprio Autor (2025)

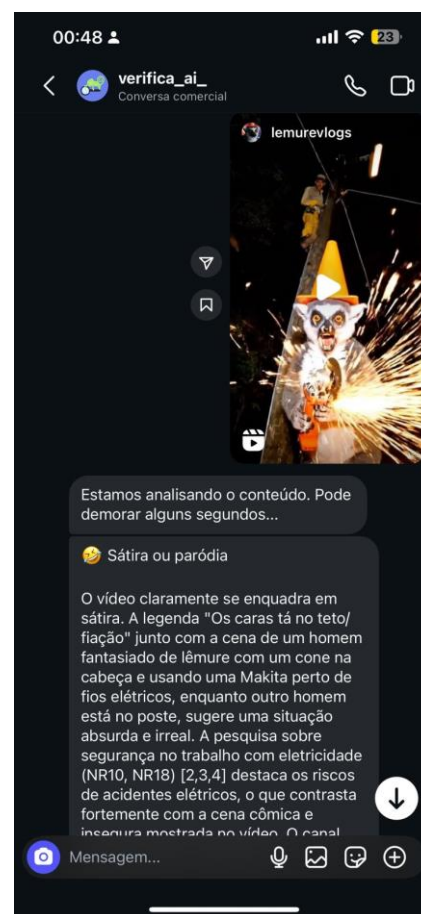
Resultados alcançados



- Outros testes do verifica AI



Fonte: Próprio Autor (2025)



Fonte: Próprio Autor (2025)



Fonte: Próprio Autor (2025)

Resultados alcançados



- Para avaliar o desempenho do modelo do sistema, utilizou-se a matriz de confusão. Para cálculo foi empregado um conjunto de 200 postagens, com temas variados, distribuídos de forma igualitária. Desse conjunto, 150 postagens eram verdadeiras e 50 falsas, contemplando os três formatos (texto, imagem e vídeo). O modelo apresentou acurácia de 97%, precisão de 98% e sensibilidade (recall) de 98%, podendo-se concluir que classifica corretamente a grande maioria dos casos.

MATRIZ DE CONFUSÃO

200 POSTS VERIFICADOS

	ACERTOS	ERROS
VP	FP	
147	3	
FN	VN	
3	47	

VP - Verdadeiro Positivo

VN - Verdadeiro Negativo

FP - Falso Positivo

FN - Falso Negativo

Fonte: Próprio Autor (2025).

TABELA 1 - RESULTADOS DAS MÉTRICAS DA MATRIZ DE CONFUSÃO		
MÉTRICA	Cálculo	%
ACURÁCIA	$\text{Precision} = (147 + 47) / (147 + 47 + 3 + 3)$	97%
PRECISÃO	$\text{Precision} = 147 / (147 + 3)$	98%
SENSIBILIDADE (RECALL)	$\text{Recall} = 147 / (147 + 3)$	98%

Fonte: Próprio Autor (2025).

Aplicabilidade dos resultados no cotidiano da sociedade



- O trabalho desenvolvido com o Verifica AI apresenta aplicabilidades práticas na detecção e classificação de desinformação multimídia (texto, imagem e vídeo) dentro da própria rede social, permitindo verificações rápidas via DM com acurácia de 97% e alinhamento às sete categorias da UNESCO, o que diretamente atende à utilidade pública de combater fake news de maneira ativa no Brasil, onde 57% da população falha em identificar conteúdos falsos.
- O trabalho surgiu a partir de observações da pandemia e pesquisas que apontavam a urgência de ferramentas que combatam as “fake News”. Ao verificar que os modelos da @Meta AI, e IA Studio da Meta apresentam apresentam alucinações e fornecer respostas incorretas.

Criatividade e inovação



- O projeto Verifica AI destaca-se por sua robustez em criar uma ferramenta que utiliza o Graph API, Gemini Flash 2.5, OCR, análise comparativa com buscas web, e prompts otimizados que classificam desinformação nas sete categorias da UNESCO. Como resultado temos uma solução que verifica conteúdos multimodais (texto, imagem e vídeo) que supera as limitações de ferramentas existentes, como Meta AI e AI Studio, ao analisar posts de forma nativa e contribuir para a redução da viralização de desinformação.

Considerações finais



- Este projeto alcançou todos os objetivos previamente estabelecidos. Os resultados obtidos na avaliação do protótipo Verifica AI revelam um desempenho robusto em termos de funcionalidade, mas também expõe limitações ao uso de tecnologias gratuitas. Restrições operacionais como limites de uso diário, e apesar raramente, latência nas respostas e erros críticos, inviabilizam o uso em larga escala.
- Temos limitações também quanto a ocorrência de falsos positivos e falsos negativos encontrados, causados pela extração de dados em sites não confiáveis em casos de conteúdos totalmente fabricados, apontam a necessidade de melhorias nos prompts usados no modelo do Gemini.



9ª Feira Mineira de Iniciação Científica



De 08 a 28 de novembro de 2025

Realização



Associação Mineira de
Pesquisa e Iniciação Científica

UNIVERSIDADE
DO ESTADO DE MINAS GERAIS



MINISTÉRIO DA
CIÊNCIA, TECNOLOGIA
E INOVAÇÃO



Apoiadores e parceiros