

Colégio Técnico de Campinas
Departamento de Processamento de Dados

ID:FAKE: Ferramenta para auxílio no reconhecimento de Fake News

André Yuhji Terada
Pedro Henrique Perez Dias
Rafael Lacerda Silva

Campinas - SP
2022

André Yuhji Terada
Pedro Henrique Perez Dias
Rafael Lacerda Silva

ID:FAKE: Ferramenta para auxílio no reconhecimento de Fake News

Trabalho de Conclusão de Curso apresentado ao Departamento de Processamento de Dados
do Colégio Técnico de Campinas, localizado em R. Culto à Ciência, 177 - Centro, Campinas
- SP, 13020-060

Orientadora: Simone Pierini Facini Rocha
Coorientador: Sergio Luiz Moral Marques

Campinas - SP
2022

Sumário

1. Introdução	4
2. Problema	5
3. Objetivos	5
3.1 Objetivo Geral	5
3.2 Objetivos Específicos	5
4. Desenvolvimento	5
4.1 Definição Inicial	5
4.2 Pesquisa	6
4.3 Identidade	6
4.4 Prototipagem	8
4.4.1 Procedimentos	8
4.4.2 Testes	11
4.5 Custos	11
5. Resultados	12
6. Conclusões	14
7. Referências	15

Resumo

O ID:FAKE busca idealizar e desenvolver uma ferramenta que possa ser utilizada como auxílio no combate à disseminação de notícias falsas na internet, que vem afetando desde o debate político nacional e internacional, até a saúde dos brasileiros. O projeto tem como base o uso de uma inteligência artificial capaz de avaliar a probabilidade de uma notícia se enquadrar como *fake news*. Em cima dessa tecnologia, desenvolvemos uma aplicação para WhatsApp que recebe mensagens em diversas formas de mídia, as converte para texto, realiza a análise da mensagem e retorna se a notícia parece verdadeira ou falsa. As funcionalidades já desenvolvidas apresentaram uma boa acurácia na transcrição para texto, e a inteligência artificial apresentou uma precisão de 86% de acertos. Junto com a acessibilidade e praticidade fornecida pelo WhatsApp, é possível afirmar que o projeto desenvolvido possui grande potencial no combate à desinformação na internet.

Palavras-chave: Desinformação, Inteligência Artificial, Social

1. Introdução

Atualmente, vivemos na era da informação, em que recebemos inúmeras notícias de diversas mídias e as consumimos desenfreadamente e de forma irresponsável. Grande parte dessas notícias circula pelas redes sociais, meio em que informações como autor e fontes se tornam irrelevantes para a maioria das pessoas. Isso as levam a acreditar nas chamadas *fake news*, notícias falsas, que fazem com que o leitor tome conclusões equivocadas em relação aos fatos.

Para impedir que pessoas caiam em *fake news*, buscamos identificar o que as levam a acreditar nestas notícias. Durante nossas pesquisas, dois tópicos se destacaram bastante, as “bolhas sociais” e a “pós-verdade”. O primeiro, responsável por isolar o usuário em um único ambiente da internet, fazendo com que seja mais fácil manipulá-lo com notícias falsas ou tendenciosas. O segundo, sendo referente à ideia de usar o apelo emocional e as crenças pessoais para convencer o usuário, ao invés de fatos e argumentos sólidos. Ambos tópicos combinados criam um ambiente ideal para a propagação de *fake news*, onde dificilmente as informações verdadeiras chegam.

Este projeto tem como foco agir no ambiente de bolha social. Buscando proporcionar auxílio rápido e acessível às pessoas na identificação de informações de qualidade.

2. Problema

Como facilitar o reconhecimento de *fake news* e levar essa técnica de maneira acessível às pessoas?

3. Objetivos

3.1 Objetivo Geral

Desenvolver uma ferramenta prática, acessível e confiável que ajude as pessoas a validarem a veracidade das informações *online*.

3.2 Objetivos Específicos

- Estudar e compreender as ferramentas existentes de combate a *fake news*
- Estudar o comportamento e o impacto da difusão das *fake news* nos tempos modernos
- Ajudar na democratização do acesso à informação de qualidade

4. Desenvolvimento

4.1 Definição Inicial

Antes de começar efetivamente a desenvolver o projeto, foi dedicado tempo a pensar nas funcionalidades que o mesmo teria, levando em consideração as principais ferramentas de combate às notícias falsas existentes.

As ferramentas existentes analisadas foram: o “Fato ou Fake” do grupo Globo, a “Agência Lupa” e a ferramenta desenvolvida pelo projeto “Contributions to the Study of Fake News in Portuguese: New Corpus and Automatic Detection Results”. Os dois primeiros são ferramentas profissionais que contam com a checagem manual de informações, sendo mais confiáveis em suas conclusões, porém mais demoradas em divulgá-las. No projeto citado por último foi desenvolvida uma ferramenta que consiste em uma inteligência artificial que analisa textos e gera uma resposta automática, indicando a provável classificação deste texto em verdadeiro ou falso, com uma taxa de 89% de acerto. Este, porém, não é muito acessível por estar presente em um site e verificar apenas um tipo de notícia, as enviadas em

textos. No entanto, a inteligência artificial desenvolvida por este projeto foi disponibilizada para treinamento de maneira local através de tutoriais no Github, possibilitando assim um funcionamento melhor.

Depois dessas análises, a versão pensada para desenvolvimento teria que contar com ferramentas para verificação de notícias falsas contidas em vários tipos de mídia (como áudios, imagens e textos) e consistiria em um sistema acessível de respostas automáticas, conhecido como *bot*. A plataforma escolhida para conter nosso *bot* foi o WhatsApp devido ao gigantesco alcance desse aplicativo em território nacional, segundo inúmeras pesquisas. Como observado na pesquisa realizada pelo instituto Statista em 2021, que indica que o aplicativo possui cerca de 120 milhões de usuários no Brasil e na realizada pelo instituto Opinion Box que o coloca como o software mais presente nos celulares brasileiros.

4.2 Pesquisa

A primeira parte do desenvolvimento do trabalho foi inteiramente dedicada à pesquisa de ferramentas existentes na internet que pudessem ser integradas para gerar uma ferramenta completa que alcançasse o objetivo proposto e obedecesse à definição inicial.

4.3 Identidade visual

Durante o desenvolvimento desse projeto, foi necessária a criação da identidade do mesmo, contando com um nome e um logotipo. O nome, “ID:FAKE”, foi pensado para transmitir a ideia por trás do projeto (“ID”: identificação e “FAKE”: falso). A primeira versão do logotipo contém as iniciais do nome do projeto, além de uma lupa que traz a ideia de investigação e descoberta. O logotipo pode ser visto abaixo:



Figura 1. Logo do projeto. Fonte: Autoral

Após pesquisa de opinião com alguns dos colegas de classe, sobre a logo, foi visto uma dificuldade de diferenciar o 'I' do 'D', o que os confundiu muito. Desse jeito, foi realizada a confecção de uma nova logo, levando em consideração todas as informações coletadas e o desenvolvimento do projeto.



Figura 2. Logo atualizada do projeto. Fonte: Autoral

4.4 Prototipagem

4.4.1 Procedimentos

Foram identificadas duas maneiras para a criação do *bot*: a primeira, utilizando o sistema Venom que foi desenvolvido em Javascript e permitiria o desenvolvimento de maneira simples e livre. A segunda, utilizando as API's da Twilio, programáveis através da linguagem Python e Node, que permitiriam desenvolver o projeto de maneira mais profissional. Inicialmente, a maneira escolhida foi a primeira. Optou-se pelo desenvolvimento da segunda opção. Também, optou-se pela criação de uma API que fizesse a comunicação com os serviços requisitados, tendo o *bot* apenas como um intermediário entre o usuário e as funcionalidades. O desenvolvimento do *bot* com a Twilio foi através de um tutorial disponibilizado pela própria empresa.

Para fazer a análise das notícias em texto, foi decidido utilizar a inteligência artificial desenvolvida pelo projeto acima citado, disponibilizada abertamente no GitHub² do projeto. A análise consistiria no recebimento de um texto pelo *bot* e envio de uma resposta para o usuário, classificando o texto como verdadeiro ou falso, através da classificação gerada pela inteligência artificial. A inteligência artificial foi treinada localmente através do modelo MultinomialNB e com as pré-definições especificadas no Github do projeto inicial.

Para realizar a análise de textos contidos em imagens, modo comum de se divulgar *fake news*, foi utilizada a API Vision AI da Google para realizar a extração destes textos e, posteriormente, analisá-los com o processo acima citado.

Para análise de imagens, foi utilizada a Custom Search API do Google para realizar a busca reversa da imagem na internet, e retornar links de ocorrências desta imagem para o usuário.

Para tratar de notícias enviadas por áudios, trabalhou-se com a API Speech-To-Text da Google para transcrever o áudio para texto e posteriormente analisá-lo com a inteligência artificial especificada anteriormente.

Por fim, além das funcionalidades previamente estabelecidas no plano de pesquisa, conseguiu-se realizar a verificação da veracidade de notícias em vídeos, o trabalho se deu

¹ Em ciência da computação, a Interface de Programação de Aplicações (em inglês, API) é um conjunto de serviços que foram programadas em um software, que são disponibilizados para que aplicativos possam utilizar dessas funcionalidades diretamente, sem envolver-se em detalhes da implementação do software.

² GitHub é uma plataforma de hospedagem de código-fonte e arquivos com controle de versão

com a utilização da ferramenta ‘ffmpeg’ que consiste em um conversor de vídeo que roda localmente. Esse, faz o recorte do vídeo em *keyframes* e extrai seu áudio, fazendo com que eles possam ser analisados através das ferramentas já descritas para imagens e áudios. Essa última parte, por ser mais complexa, foi desenvolvida por último.

O esquema de funcionamento da ferramenta pode ser visto abaixo:

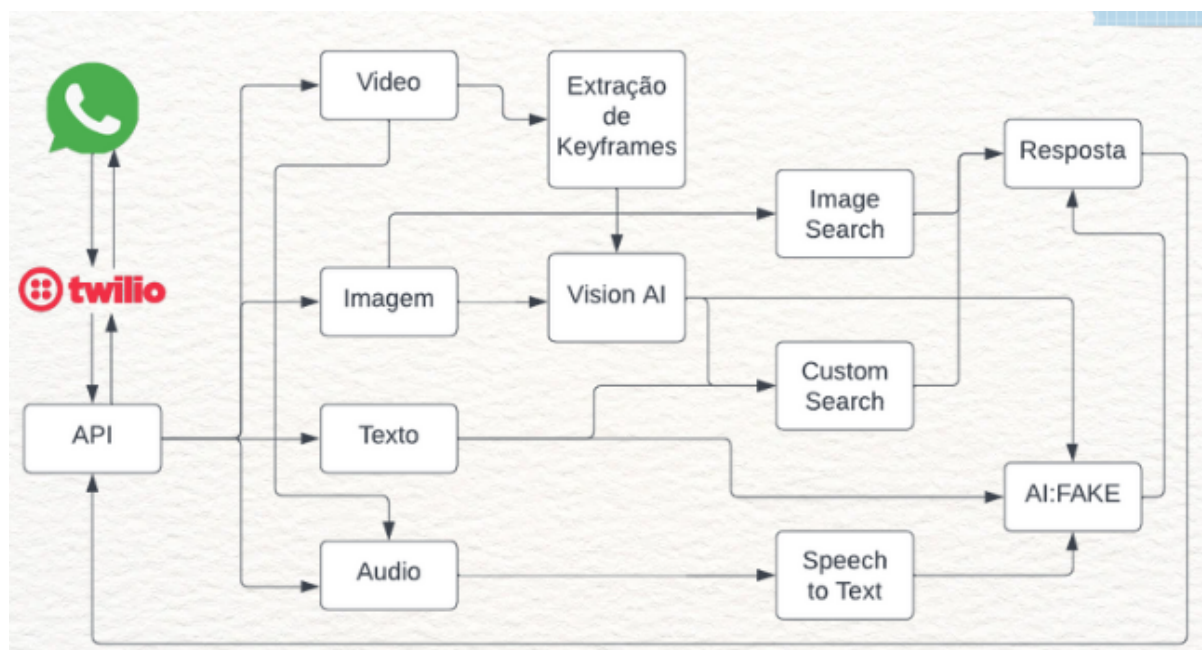


Figura 3. Esquema de funcionamento do projeto. Fonte: Autoral

Para organização do desenvolvimento do projeto, foi elaborado o cronograma abaixo:

Tarefas	Meses									
	Mar	Abr	Mai	Jun	Jul	Ago	Set	Out	Nov	Dec
Escolha do tema	X									
Planejamento e Pesquisa	X	X	X							
Revisão Bibliográfica	X	X	X	X	X	X	X	X		

Desenvolvimento do Plano de pesquisa			X	X						
Diário de Bordo	X	X	X	X	X	X	X	X	X	
Criação do <i>bot</i>				X	X					
Implementar I.A projeto				X	X					
Implementar Vision AI						X	X			
Implementar uso do TinEye							X	X		
Implementar Speech-to-Text								X	X	
Implementar Análise de Vídeo									X	X
Testes para protótipos				X	X	X	X	X	X	X
Relatório do projeto								X	X	X

Devido ao fato de o projeto trabalhar com muitas APIs que utilizam credenciais próprias para funcionamento, foram observados riscos de segurança durante o projeto com a possível exposição dessas credenciais. Para evitar esse problema, foi necessário utilizar o gerenciamento de credenciais da GCloud.

4.4.2 Testes

Os testes realizados se deram paralelamente ao desenvolvimento do projeto para contribuir com o desenvolvimento. Eles foram feitos em pequena escala, através de amigos e familiares.

4.5 Custos

Para ser possível realizar o projeto em larga escala, teriam que ser custeados os usos das APIs Speech-to-Text, da Vision AI e da Custom Search API da Google, atualmente desconsiderados por contar com período de testes e limite de utilização gratuito. Os custos estão apresentados abaixo:

Custos de uso da API Vision AI:

Recurso	Preço por 1.000 unidades		
	Primeiras 1.000 unidades/mês	1.001 - 5.000.000 unidades/mês	5.000.001 unidades/mês ou um número maior
Detecção de rótulos	Grátis	US\$ 1,50	US\$ 1,00
Detecção de texto	Gratuito	US\$ 1,50	US\$ 0,60
Detecção de texto de documentos	Grátis	US\$ 1,50	US\$ 0,60

Figura 3. Tabela de custos de uso da API Vision AI. Fonte nas referências

Custos de uso da API Speech-to-Text:

Seleção de	Modelos padrão (todos os modelos, exceto chamadas aprimoradas de telefone e vídeo)		Modelos aprimorados (chamada de telefone, vídeo)	
	De 0 a 60 minutos	Mais de 60 minutos até 1 milhão de minutos	De 0 a 60 minutos	Mais de 60 minutos até 1 milhão de minutos
Reconhecimento de fala (sem geração de registros de dados – padrão)	Grátis	US\$ 0,006/15 segundos **	Gratuito	US\$ 0,009/15 segundos **
Reconhecimento de fala (com possibilidade da ativação da geração de registros de dados)	Grátis	US\$ 0,004/15 segundos **	Grátis	US\$ 0,006/15 segundos **

Figura 4. Tabela de custos de uso da API Speech-to-Text. Fonte nas referências

A API JSON da Pesquisa personalizada do Google fornece 100 consultas de pesquisa por dia sem custo financeiro. Se você precisar de mais, inscreva-se para [faturamento](#) no Console de APIs. Solicitações adicionais custam US \$5 por 1.000 consultas, até 10 mil consultas por dia.

Se você precisar de mais de 10 mil consultas por dia e o mecanismo de pesquisa programável pesquisar até 10 sites, talvez seja interessante usar a [API JSON de Site restrito restrito a pesquisas personalizadas](#), que não tem um limite de consulta diária.

Figura 5. Custos de uso da Custom Search API. Fonte nas referências

5. Resultados

Os resultados alcançados até o momento são bastante promissores. Contando já com as análises de texto, imagem, vídeo e áudio funcionais, o protótipo atual do projeto possui uma boa taxa de acurácia e cumpre muito bem com o objetivo inicial planejado com ele.

Os resultados preliminares podem ser observados abaixo:

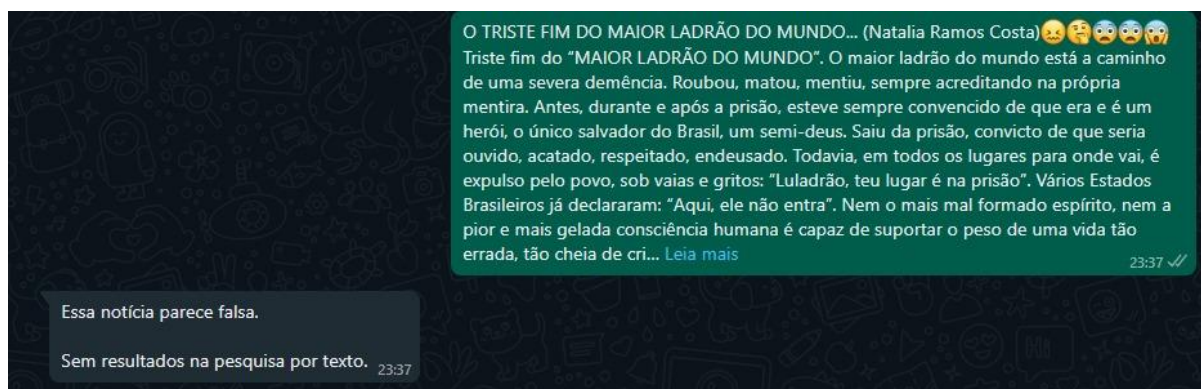


Figura 6. Resposta indicada pelo bot perante uma notícia em texto comprovadamente falsa. Fonte: Autoria própria



Figura 7. Imagem que contém *fake news* em texto e em imagem. Fonte: Autoria própria

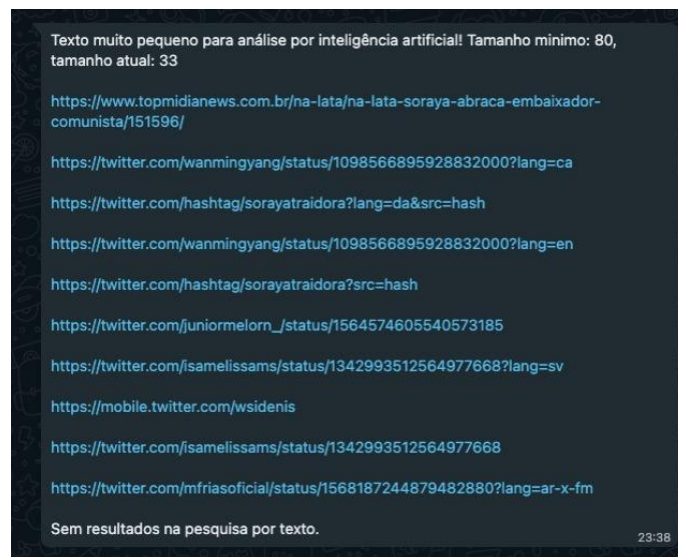


Figura 8. Resposta do *bot* à notícia mostrada acima. Autoria própria



Figura 9. Imagem que contém *fake news* em texto. Fonte: Autoria própria

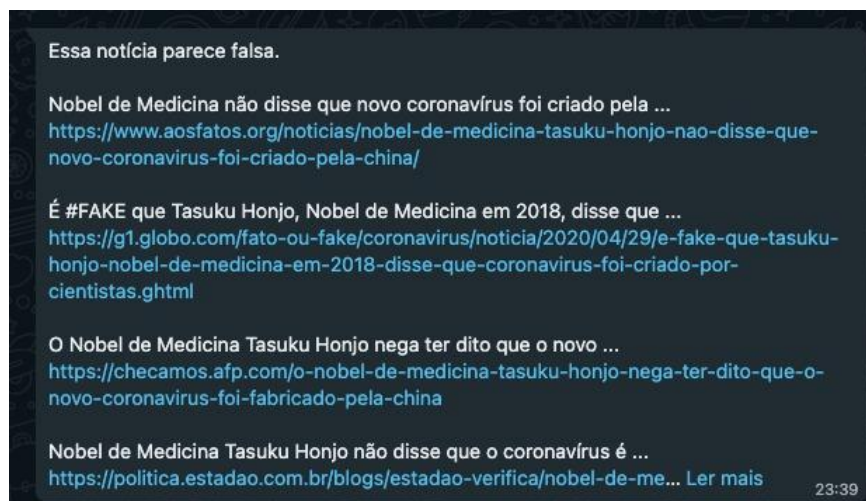


Foto 10. Resposta do *bot* à fake news citada acima. Fonte: Autoria própria

6. Conclusões

Com base nos dados apresentados em relação à presença e impacto das notícias falsas na nossa sociedade atual, pode-se concluir a necessidade de uma ferramenta que possa

auxiliar na identificação dessas em meio à grande quantidade de informações que circulam *online*.

A pesquisa encontra-se em sua fase final, com previsão de término até o final do ano. No entanto, os resultados parciais já indicam o potencial da mesma em ser de grande ajuda no combate à desinformação.

Para o futuro, pretende-se divulgar o projeto a fim de ampliar o alcance do mesmo. Para isso, seria necessária uma rede de apoio e investimento, para lidar com os custos indicados. Além disso, podem ser pensadas mais formas de incrementar a ferramenta, como a verificação de notícias contidas em *links* recebidos pelo usuário.

7. Referências

Empresa ‘Statista’. **Messenger: Whatsapp in Brasilien 2022 Brand Report** (nome original da pesquisa). Dados referenciados disponíveis em:

<https://www.messengerpeople.com/pt-br/whatsapp-no-brasil/#:~:text=Com%20quase%20120%20mil%C3%B5es%20de,meio%20predileto%20para%20andar%20mensagens.>

Monteiro, R. A. **Contributions to the Study of Fake News in Portuguese: New Corpus and Automatic Detection Results**. Disponível em:

https://link.springer.com/chapter/10.1007/978-3-319-99722-3_33

OpinionBox. **Pesquisa sobre aplicativos no Brasil: apps mais populares, hábitos e preferências dos brasileiros**. Dados disponíveis em:

<https://blog.opinionbox.com/pesquisa-sobre-aplicativos-no-brasil/#:~:text=Uso%20de%20smartphones%20e%20aplicativos%20no%20Brasil&text=Al%C3%A9m%20disso%2C%2098%25%20dos%20usu%C3%A1rios,manteve%20a%20propor%C3%A7%C3%A3o%20de%2015%25>

Vision AI API. Dados de custo disponíveis em: <https://cloud.google.com/vision/pricing/>

Speech-to-Text API. Dados de custo disponíveis em:

<https://cloud.google.com/speech-to-text/pricing>

Custom Search API. Dados de uso disponíveis em:

<https://developers.google.com/custom-search/v1/introduction> e em:

<https://programmablesearchengine.google.com/controlpanel/create>. Dados de custo disponíveis em: <https://developers.google.com/custom-search/v1/overview>

Ferramenta Venom. Código e informações disponíveis em: <https://github.com/orkestral/venom>

Tutorial para criação do *bot* com a Twilio. Dados disponíveis em:

<https://www.twilio.com/blog/criar-um-chatbot-para-whatsapp-com-python-flask-e-twilio>

GCloud. Dados Disponíveis em:

<https://googleapis.dev/python/google-api-core/latest/auth.html>

MultinomialNB. Dados disponíveis em:

https://scikit-learn.org/stable/modules/generated/sklearn.naive_bayes.MultinomialNB.html