

RetificAI: Um verificador de notícias falsas com base em IA e Aprendizado de Máquina

Guilherme Jun Yazaki Grillo¹, Gabriel Tavares Brayn Rosati¹, Jaqueline Matos da Hora¹, Paulo Henrique de Souza Soares¹, Kamila Rios da Hora Rodrigues¹

¹Instituto de Ciências Matemáticas e de Computação - Universidade de São Paulo (USP) São Carlos - SP - Brasil

{guilhermej.un.grillo,gt.rosati,jaqueline.m.hora,paulosoares}@usp.br,
kamila.rios@icmc.usp.br

Resumo. São muitos os caminhos que levam à desinformação. O acesso cada vez mais crescente a dispositivos móveis e aos canais de comunicação potencializou a disseminação de notícias falsas. A sociedade tem se mobilizado em criar soluções que amenizem o fenômeno das Fake News. Este trabalho propõe o RetificAI, um verificador de notícias falsas, composto de um Website e um chatbot, e com base em IA e Aprendizado de Máquina, projetado para disponibilizar de maneira fácil e guiada um feedback sobre notícias veiculadas nos mais diferentes meios. Um estudo prévio foi feito em um protótipo existente e com base nele um novo produto foi projetado. O intuito desse novo produto é ser efetivo, além de assertivo e fácil de usar para diferentes perfis de usuários, sobretudo idosos, que são reconhecidamente aqueles que mais disseminam notícias falsas no Brasil.

1. Introdução

As emoções e as crenças pessoais, para Myers e Dewall (2017), têm grande impacto na formação da percepção que o indivíduo tem do mundo à sua volta. A percepção é caracterizada como a "organização e interpretação de informações sensoriais conscientemente", portanto, não há uma única forma de percepção, já que há vários modos de interpretação e de relação com o repertório individual de informações ou conhecimentos que cada pessoa possui. Essa percepção, no entanto, pode estar ou não de acordo com os fatos. Quando não está de acordo, são criadas realidades alternativas "que não se baseiam em fatos, mas sim em emoções. Realidades alternativas que se baseiam na percepção, não em dados" (MEDRÁN, 2017). É assim que ocorre as distorções dos acontecimentos reais (GOMES, PENNA e ARROIO, 2020). De acordo com Pennycook e Rand (2021), o mau discernimento da verdade está associado à falta de uma análise cuidadosa, sondagem e conhecimento relevante, e o uso de heurísticas como familiaridade. Além disso, há uma desconexão substancial entre o que as pessoas acreditam e o que eles compartilham nas redes sociais, por exemplo. Essa dissociação é em grande parte impulsionada pela desatenção, mais do que pelo compartilhamento proposital de desinformação. Assim, as intervenções podem estimular com sucesso os usuários de mídia social a se concentrarem mais na precisão.

O termo *fake news* ganha a definição de notícias falsas quando traduzido para o Português, mas, para alguns autores, o cenário atual que envolve o termo o coloca em um patamar mais complexo (ZUCKERMAN, 2017). Zuckerman (2017) faz uma categorização das notícias falsas de três maneiras: 1) A notícia falsa para desviar a atenção do problema real; 2) A propaganda, em que é usada a notícia falsa, em um contexto político, para promover um candidato e denegrir a imagem do outro; 3) E um número grande de notícias falsas, confundindo o leitor pelo excesso de informação. A última categoria, segundo o autor, é uma maneira de desacreditar nos meios de comunicação tradicionais e criar o estado de desinformação (ZUCKERMAN, 2017). Allcott e Gentzkow (2017), por sua vez, traçam uma definição mais próxima da tradução, sendo as fake news "[...] artigos ou notícias intencionais para enganar os leitores", tendo assim um propósito de ludibriar o receptor da informação.

Este projeto se insere no contexto de disseminação de *fake news*, uma vez que hoje a tecnologia facilita muito a disseminação de notícias, sendo elas verdadeiras ou não. Um problema claro a se notar é o aumento de fake news, sobretudo em épocas de eleição (como o atual momento do Brasil), assim como uma maior dificuldade em se caracterizar a veracidade de uma notícia. Verificar a veracidade de um fato ou informação recebida por diversos meios midiáticos é uma forma de minimizar os efeitos causados pelas fake news. Os

autores deste trabalho, inspirados em um protótipo de ferramenta desenvolvido no mesmo instituto de pesquisa, realizaram estudos empíricos e analíticos em tal protótipo e, a partir da análise, realizaram o redesign da solução de apoio na identificação de notícias falsas. O intuito é disseminar tal solução, de modo que a sociedade de maneira geral possa pesquisar, de forma rápida e guiada, se uma notícia é verdadeira ou não. A seção a seguir descreve os procedimentos metodológicos adotados neste trabalho.

2. Metodologia/Processo

Para o design da solução aqui proposta adotou-se o método do *Design Thinking* (DT) (BROWN, 2009), usado para estimular ideação ao abordar problemas relacionados a futuras aquisições de informações, análise de conhecimento e propostas de soluções. O DT combina empatia no contexto de um problema, para colocar as pessoas no centro do desenvolvimento de um projeto. Também associa a criatividade para geração de soluções e razão para analisar e adaptar as soluções para o contexto. O DT é uma abordagem centrada no humano, aplicado em 5 etapas, a saber: empatia; definição; ideação; prototipação e teste. A etapa de empatia ou imersão sugere um aprofundamento sobre o problema de design, sobre as partes interessadas na solução, bem como quais são as ameaças, oportunidades, pontos fortes e pontos fracos do produto em questão. Neste trabalho esta etapa foi materializada com o uso de recursos como *benchmarking*, estudo empírico baseado em observação e estudo analítico baseado em inspeção com heurísticas. O *benchmarking* permitiu avaliar outras ferramentas que se propõem a verificar a legitimidade de notícias, tais como: Fake Check¹, Fake News Detection², Fake News BR³. No *benchmarking* foram listadas as principais características dos sites. Entre as características, consideradas aqui como requisitos, elenca-se: oferta de uma descrição sobre o funcionamento da ferramenta com uma breve explicação técnica da lógica ou da técnica utilizada; espaço para colar o texto (caixa de texto), escolher o método (por palavras ou por classes gramaticais, por exemplo) e avaliar (geralmente um botão de enviar). Houve também a realização de pesquisas de outros *chatbots* já existentes no mercado que realizam a verificação de notícias pelo aplicativo de celular Whatsapp, como: Fátima⁴ e Bot da Agência Lupa⁵. O *benchmarking* realizado sobre esses *chatbots* foi feito considerando a análise da escrita (ex.: tempos verbais, formalidade e simplicidade) e outros elementos utilizados no aplicativo Whatsapp, como o uso de emojis e interações exclusivas (ex.: botões e menu de opções).

Para entender o público alvo, um conjunto de personas foi criado (PRUITT e ADLIN, 2006). Um jovem adulto, formado em jornalismo e com interesse em divulgar notícias e fatos relacionados à política e atualidades. Um idoso, aposentado, que costuma ler jornais diariamente, conhece poucas fontes de informação na Internet e não confia em muitas notícias vistas no computador. Gostaria de um meio fácil e seguro de checar suas notícias pelo celular. Além disso, tem dificuldades visuais como daltonismo. Uma mulher, de meia-idade, que estuda para concurso e procura por sites de confiança e com informações sobre concursos que sejam confiáveis e seguros.

Uma avaliação de uso junto a idosos de um curso de letramento digital também foi conduzida na etapa de design. Estudos da literatura apontam que esse é um dos públicos mais suscetíveis a disseminar notícias falsas (ESTABEL, LUCE e SANTINI, 2020; YABRUDE et al., 2020). Após uma aula sobre conscientização e compartilhamento de notícias falsas, doze idosos foram convidados a interagir com a ferramenta *Fake Check* e os pesquisadores observaram tal interação. Os idosos tinham idade entre 65 e 84 anos, de classe média alta e moradores da cidade de São Carlos/SP. Notou-se que os idosos tiveram dificuldades sobretudo no tamanho das fontes usadas no site, no copiar e colar (usando o celular) trechos da notícia para verificação, em identificar o local de clicar para solicitar a verificação, em visualizar a caixa de texto em que o trecho da notícia deveria ser copiado e em visualizar o resultado da checagem, uma vez que tinham que rolar a tela para visualizar, mas o site não apontava que o processo já tinha sido feito. Além disso, o site exigia número alto de palavras para garantir a verificação e todo o trecho copiado era ilustrado na interface, dificultando o acesso às demais

¹ <http://nilc-fakenews.herokuapp.com/>

² <http://fakenews.research.sfu.ca/>

³ <https://sites.google.com/view/detector-de-fake-news/p%C3%A1gina-inicial>

⁴ <https://www.aosfatos.org/fatima/>

⁵ https://api.whatsapp.com/send/?phone=5521991933751&text&app_absent=0

informações e funcionalidades da página. Ressalta-se que essa é uma das atividades previstas para o curso de letramento que tem comitê de ética aprovado pelo CEP-EACH-USP.

Ainda na etapa de empatia, duas avaliações analíticas de inspeção foram realizadas, uma para verificar requisitos de acessibilidade e outra para verificar aspectos de usabilidade. A avaliação heurística de usabilidade (NIELSEN e MOLICH, 1990) foi conduzida na ferramenta Fake Check com 3 avaliadores, com diferentes níveis de experiência na condução dessa atividade. A avaliação apontou problemas como: (**Visibilidade e estado do sistema**) o site não indica previamente se o texto que o usuário colocou para checar tem no mínimo 100 palavras; não tem indicação correta sobre o passo a passo a ser conduzido; não oferece um *loading* eficiente; não tem uma indicação que o resultado da consulta apareceu após clicar em ‘enviar’; não tem uma indicação de onde o usuário está na página; (**Relação entre sistema e mundo real**) - presença de botões falsos, presença de elemento com comportamento de botão, mas com identidade visual não compatível; (**Controle e liberdade do usuário**) - não é possível cancelar o envio, não é possível limpar o texto. (**Consistência e padronização**) - repetição de conteúdos em páginas diferentes, caixa de texto é pequena e não aponta o uso de notícias com mais de 100 palavras. (**Prevenção de erros**) - site não indica previamente se o texto que o usuário colocou tem no mínimo 100 palavras, notícia inserida se mistura com o texto da descrição, caixa de texto não se destaca; (**Design estético e minimalista**) - Site é extremamente minimalista a ponto de faltar informações e passos; (**Ajuda e documentação**) - não tem indicação correta sobre o passo a passo, pouca instrução para o “Modelo de detecção” e não tem FAQ, dúvidas, suporte. A avaliação heurística foi conduzida por 4 especialistas.

Avaliação Simplificada de Acessibilidade (WARAU, 2022) foi conduzida com 4 avaliadores, diferentes daqueles que realizaram a avaliação de usabilidade, com diferentes níveis de experiência na condução dessa avaliação, e identificou os seguintes problemas: não havia textos alternativos para as imagens, o site não era completamente responsivo e ao variar a resolução da tela algum conteúdo se perde, ao redimensionar a janela da aplicação para tamanhos menores que o máximo é possível verificar a rolagem horizontal (violação grave de acessibilidade), ao verificar o contraste alguns links e textos não ficam visíveis, não era possível acessar todo o conteúdo por meio de TAB, Enter e setas direcionais.

Finalizada a etapa de empatia, a equipe conhecia os principais problemas das ferramentas atuais de detecção de notícias falsas, bem como as principais vulnerabilidades do público idoso como exemplo. Passou-se então para a etapa de ideação, em que rodadas de brainstorming foram conduzidas com o intuito de idealizar a nova solução. Diversas tecnologias estão disponíveis no mercado, tais como smartwatches e Internet das Coisas, no entanto, esses recursos ainda não estão disponíveis para a maioria da população. O objetivo da equipe era projetar uma solução que pudesse ser facilmente utilizada, que empregasse dispositivos já de adoção da população e que, de preferência, fosse acessível para a população, incluindo a idosa. Idealizou-se então uma página Web responsiva e um chatbot para o aplicativo de mensagens Whatsapp, que utilizam recursos de Inteligência Artificial e Aprendizado de Máquina para disponibilizar a usuários médios a verificação se as notícias recebidas em suas redes sociais são ou não notícias falsas.

Na etapa de prototipação as soluções foram desenhadas usando a ferramenta Figma, buscando resolver os problemas de usabilidade e acessibilidade identificados na etapa de empatia/imersão. Um estudo de cores e contraste foi realizado na interface. Também foi feita a prototipação do *chatbot* do Whatsapp para simular a interação com uma pessoa, mostrando as principais funcionalidades do bot. Durante a prototipação foi realizado ainda, um estudo de *UX Writing*⁶, para reformular partes do texto da ferramenta, de modo a deixá-la com uma explicação menos técnica e mais objetiva. Além disso, houve mudanças quanto a abordagem das falas, tom e direcionamento das mesmas.

Na etapa de teste, por fim, a ferramenta, com o novo design e funcionalidades, foi avaliada de modo analítico em relação aos aspectos de usabilidade e acessibilidade, utilizando os mesmos instrumentos de inspeção e os mesmos avaliadores. Esses novos estudos não apontaram problemas.

⁶ <https://www.nngroup.com/articles/ux-writing-study-guide/>

3. Justificativa

Conforme descrito acima, a equipe deste trabalho acredita que a solução pensada é uma solução acessível ao público. Além disso, a solução usa tecnologias já adotadas e emprega requisitos de design e interação importantes, tais como boa usabilidade e acessibilidade. O método empregado permitiu ter uma ampla visão do que seria de fato usado pela população, sobretudo pelos idosos. Os artefatos empregados deram retorno dos principais problemas de design existentes em ferramentas já disponíveis, e confirmaram a lacuna existente entre o fenômeno das fake news e os recursos existentes para detectá-las. Estes autores acreditam que a solução pode ser efetiva na detecção de notícias falsas, uma vez que usa um corpus já bastante treinado de IA, mas que oferece uma interface de usuário agradável e mais assertiva que as atualmente disponíveis na Internet.

4. Solução proposta e Cenário de uso

Esta equipe apresenta uma solução que consiste em 2 artefatos computacionais: uma aplicação Web e um *chatbot* para o aplicativo Whatsapp. A aplicação Web consiste em 4 etapas que um usuário precisa passar para verificar se a notícia é verdadeira ou falsa, são elas: 1) selecionar o modelo de detecção, 2) selecionar o modo do conteúdo, 3) adicionar a notícia e, por fim, 4) receber o resultado após a checagem da inteligência artificial. Um exemplo de cenário de uso considera uma pessoa idosa navegando pela Internet e fica na dúvida se a notícia a qual está lendo é ou não uma *fake news*. Assim, acessa o Website do RetificAI e verifica, de maneira rápida, a veracidade da informação.

O *chatbot* no Whatsapp, por sua vez, consiste na replicação da aplicação Web com simplificações, principalmente por conta do modelo da plataforma (aplicativo de mensagens para smartphones). O bot inicia a conversa se apresentando e instruindo sobre a sua utilização, que consistem em: 1) usuário selecionar o modelo de detecção, 2) inserir a notícia (link ou texto) e 3) informar, após a checagem da inteligência artificial, o resultado. Um cenário de uso considera um usuário idoso que recebe, de outro usuário ou grupo do Whatsapp, uma notícia. Para verificar se ela contém informações verdadeiras ou não, o idoso manda mensagem para o bot do RetificAI e encaminha o link da notícia, recebendo o resultado do bot após a verificação. Imagens de ambos os artefatos podem ser vistos no link para o protótipo disponível na Seção 6.

5. Viabilidade

A escolha de uma solução Web para o problema proposto foi devido a ela já ser uma plataforma bastante concretizada no mundo atual, e devido ao fato das pessoas terem facilidade no acesso a computadores e aparelhos celulares pessoais, também devido às redes de Internet estarem mais acessíveis. Além disso, a utilização de serviços de armazenamento e manipulação de dados na nuvem oferecem alta velocidade de processamento por um custo baixo. Já a decisão de desenvolver um chatbot para o Whatsapp se deve ao fato desse ser o principal aplicativo de comunicação via mensagens instantâneas do Brasil, e também por ser um dos principais meios de divulgação e recebimento de informações para muitas pessoas (PANORAMA, 2022), ocasionando no espalhamento de notícias falsas com maior facilidade. Um chatbot de checagem de fatos pode ser criado com facilidade devido à existência de diversas plataformas de bots específicos para o Whatsapp, além de sua API (em Português, Interface de Programação de Aplicações) ser pública e com valores baixos. É intenção do grupo disponibilizar e fazer ampla divulgação de ambos os artefatos durante o período eleitoral no Brasil.

6. Protótipo

Link para protótipo:

<https://www.figma.com/proto/oHrBpM59CM5Ii4xRuyeFcV/retificAI-FINAL?node-id=89%3A8397&scaling=min-zoom&page-id=0%3A1&starting-point-node-id=89%3A8397>

7. Vídeo

Link para o vídeo demonstrativo: <https://youtu.be/szhR7OIDIeA>

Referências

- BROWN, T. "The Making of a Design Thinker." *Metropolis* Oct. 2009: 60-62.
- ESTABEL, L. B., LUCE, B. F., and SANTINI, L. A. "Elderly, fake news and information literacy". *Revista Brasileira de Biblioteconomia e Documentação*; v. 16 (2020); 1-15, 24(2), 15-1.
- GOMES, S. F., PENNA, J. C. B. de O., and ARROIO, A. "Fake news científicas: percepção, persuasão e letramento." *Ciência & Educação (Bauru)* 26 (2020).
- MEDRÁN, A. "No reino da pós-verdade, a irrelevância é o castigo". *UNO*, São Paulo, n. 27, p. 33-35, 2017. Disponível em: <https://tinyurl.com/y9knxgaz> Acesso em: 04 jul. 2022.
- MYERS, D. G.; DEWALL, C. N. "Psicologia 11". ed. Rio de Janeiro: EDEL, 2017.
- NIELSEN, J., and MOLICH, R. "Heuristic evaluation of user interfaces". In *Proceedings of the SIGCHI conference on Human factors in computing systems, 1990* (pp. 249-256).
- PANORAMA. "Pesquisas Independentes sobre o mercado de conteúdos móveis". Disponível em: <https://www.mobilettime.com.br/pesquisas/>. Acessado em: 20 de jul. 2022.
- PENNYCOOK, G., and RAND, D.G. "The psychology of fake news." *Trends in cognitive sciences* 25.5 (2021): 388-402.
- PRUITT, J and ADLIN, T. (2006). *The Persona Lifecycle: Keeping People in Mind Throughout Product Design*. São Francisco: Elsevier.
- YABRUDE, A. T. Z., SOUZA, A. C. M. D., CAMPOS, C. W. D., BOHN, L., AND TIBONI, M. (2020). "Desafios das fake news com idosos durante infodemia sobre covid-19: experiência de estudantes de medicina". *Revista Brasileira de Educação Médica*, 44.
- WARAU, "Websites Atendendo a Requisitos de Acessibilidade e Usabilidade". Disponível em: <https://leanpub.com/warau>. Acessado em: 04 de jun. 2022.
- ZUCKERMAN, E. "Fake news is a red herring". Disponível em: <https://www.dw.com/en/fake-news-is-a-red-herring/a-37269377>. Acesso em: 04 jul. 2022.