# Uma abordagem híbrida para detecção de notícias falsas: Aprendizado de Máquina e Verificação Baseada em Conhecimento

**Caroline Barbosa de Oliveira Lira**[1] **, Kamila Rios da Hora Rodrigues**[1]

[1]Instituto de Ciências Matemáticas e de Computação - Universidade de São Paulo (ICMC-USP), Avenida Trabalhador São-carlense, 400 - Centro - São Carlos/SP

clira23@usp.br, kamila.rios@icmc.usp.br

***Resumo. Introdução****: A disseminação de notícias falsas (fake news) se tornou uma problemática de escala global. Eventos como pandemias, conflitos armados, guerras comerciais e processos eleitorais marcados por intensa polarização transformaram as redes sociais em canais propícios à manipulação, à difusão de teorias da conspiração e à veiculação de discursos de protesto e de ódio. As vulnerabilidades humanas — notadamente as de ordem cognitiva — são frequentemente exploradas, especialmente no caso de pessoas idosas, para fomentar a circulação de desinformação.* ***Objetivo****: O presente trabalho descreve a proposta de uma abordagem híbrida com o objetivo de reduzir a disseminação e o compartilhamento de fake news em redes sociais, utilizando recursos acessíveis de verificação por meio de uma estratégia inicialmente fundamentada na disseminação do conhecimento, seguida pela incorporação de ferramentas específicas de detecção automática de notícias falsas, baseadas em técnicas de Aprendizado de Máquina.* ***Metodologia****: Um corpus de notícias em língua portuguesa foi previamente refinado, e testes foram realizados para avaliar sua acurácia.* ***Resultados****: Os resultados demonstraram desempenho promissor em relação à precisão, levando à implementação desse corpus na ferramenta ADA. Tal ferramenta, destinada à verificação da veracidade de notícias, foi desenvolvida por meio de técnicas de Design Participativo (DP) com o público 60+, visando promover o engajamento e incentivar a adesão ao uso da tecnologia como suporte na identificação de fake news. Os resultados apontam para uma ferramenta com boa adesão do público alvo testado e com recursos de acessibilidade.*

***Palavras-Chave*** *Notícias falsas, Abordagem Híbrida, Verificação baseada em conhecimento, Aprendizado de Máquina, Ferramenta ADA.*

## 1. Introdução

Em 2020, a revista do MIT (*Massachusetts Institute of Technology) - MIT Technology Review*, apontou que os usuários mais velhos compartilham 7 vezes mais *fake news*. Parte do conteúdo de desinformação é geralmente alimentado e compartilhado por adultos mais velhos nas redes sociais. Nadia Brashier, psicóloga do Departamento de Psicologia da Universidade de Harvard, expõe a falta de evidência de muitas suposições sobre a razão dos idosos compartilharem mais desinformação e também argumenta que não existe uma resposta única[1].

[1]Disponível em: https://mittechreview.com.br/usuarios-mais-velhos--/mais-informacoes-falsas-seu-palpite-do-porque-pode-estar-errado

Nadia Brashier ainda afirma que algumas estratégias, como a checagem de fatos, fracassaram, inclusive quando as plataformas de redes sociais usam o sistema de verificação de dados ou rotulam mensagens como enganosas ou falsas [Brashier e Schacter 2020]. Para o público idoso, quando elas são repetidamente rotuladas como "falsas", causam o efeito contrário, aumentando a crença no conteúdo. Para a pesquisadora, isso não significa que os idosos sejam piores em discernir o quanto uma notícia é real. Para ela, em uma avaliação sobre a veracidade das manchetes, eles se saíram melhor. Ou seja, o problema está nas abordagens atuais de checagem de fatos, pois, para a profissional, não é o melhor caminho, mas ela sugere que, para lidar com soluções mais eficazes, é preciso trabalhar com a alfabetização digital, uma vez que esse público tem menos afinidade com as plataformas sociais e precisa confiar nas que estão mais próximas.

Um dos recursos tecnológicos para lidar com as *fake news* nas mídias sociais é o uso de PLN (Processamento de Linguagem Natural), o qual, na prática, reduz o tempo humano e amplia os esforços para detectar e prevenir a disseminação de notícias falsas. Isso é, o uso automático de detecção de notícias falsas. Um exemplo desse uso eficiente é uma pesquisa pioneira no Brasil desenvolvida pela Universidade de São Paulo (USP) e Universidade Federal de São Carlos (UFSCar) [Monteiro et al. 2018], na qual os pesquisadores criaram um detector de *fake news*, a ferramenta *FakeCheck*. O trabalho teve como objetivo explorar os métodos existentes para detecção de conteúdo enganoso, utilizando a técnica de PLN e um *corpus* em Português [Santos 2022].

Inspirados na ferramenta *FakeCheck*, neste trabalho foi desenvolvida a ferramenta Ada, que passou por adaptações após avaliação com o público idoso e contém um *corpus* também em Português, com mais de 3000 notícias, abrangendo temas como saúde, política e política internacional.m

O presente estudo propõe a interação do usuário com um método híbrido e multidisciplinar. Tais abordagens enfatizam a condução de pesquisas em Interação Humano-Computador (IHC), buscando alinhamento com o documento do GranDIHC-BR 2025–2035, sobretudo ao desafio GC1: Novas Abordagens Teóricas e Metodológicas em IHC [da Silva Junior et al. 2024] e ao GC4: Aspectos socioculturais na interação humano-computador [Neris et al. 2024].

Este artigo está dividido da seguinte maneira: Seção 2 descreve os trabalhos correlatos, Seção 3 descreve o design de soluções computacionais, como fundamentação e conceitos para o desenvolvimento da Ada, na Seção 4 são descritas as etapas da construção da ferramenta,a Seção 5 traz uma explicação detalhada sobre o processo de construção da Ada, na Seção 6 são apontadas as limitações encontradas para o desenvolvimento da ferramenta e as avaliações feitas, na Seção 7 são descritos os cuidados éticos e, na Seção 8, são descritas as considerações finais.

## 2. Trabalhos Correlatos

Diante das várias tarefas possíveis aplicadas pelo processamento de linguagem natural, a seguir são descritos trabalhos que descrevem o processamento de textos, recursos para classificação de notícias falsas e verdadeiras, que serviram de base para este trabalho. Nesses são apresentados diferentes classificadores: Correção de erro de concordância; Melhoria na fluidez e clareza do texto.

[Monteiro et al. 2018] apontam a escassez de conjuntos de dados rotulados em Português, o que dificulta a classificação automática de documentos. O trabalho foi inspirado em outras iniciativas, principalmente na língua inglesa. No entanto, foi introduzido um *corpus* em Português contendo notícias falsas e verdadeiras. Os autores utilizaram a técnica de aprendizado de máquina. Entre as melhorias do estudo estão o ajuste na estrutura da frase para maior clareza, e a substituição de termos.

[Della Vedova et al. 2018] realizaram um estudo na Finlândia e propuseram um novo modelo para detecção de notícias falsas com *Machine Learning* (ML). Além disso, desenvolveram um chatbot do Facebook Messenger com aplicação no mundo real, alcançando uma precisão de detecção de notícias falsas de 81,7%. Foram feitas melhorias no modelo quanto à precisão e clareza das informações no mesmo.

[Kong et al. 2020] apresentaram um estudo realizado na Malásia, utilizando notícias de mídias sociais para combater a disseminação de falsas histórias. Foram empregadas técnicas de PNL, incluindo pré-processamento de textos, tokenização, lematização e aplicação de TF-IDF *(Term Frequency-Inverse Document Frequency)* para analisar a frequência de termos. Os resultados dos modelos apresentados demonstram que *"[...] os modelos treinados com conteúdo de notícias podem alcançar melhor desempenho com o tempo de computação [...]*". O estudo resultou em acréscimo de informações às técnicas utilizadas no mesmo e na melhoria dos resultados.

[Zervopoulos et al. 2020] conduziram um estudo com notícias de Hong Kong, utilizando PLN para detecção de notícias falsas. Ao final, os autores apresentaram resultados que evidenciam diferenças morfológicas e lexicais entre os textos com notícias falsas do Twitter, relacionados aos protestos de Hong Kong ocorridos em 2019.

Todavia, embora os classificadores e técnicas para detecção de *fake news* sejam recorrentemente explorados, a ferramenta Ada se destaca por ser a primeira a ser desenvolvida com o apoio de idosos e para eles, bem como levando em consideração questões como baixo conhecimento tecnológico e dificuldades do público 60+ devido à idade.

## 3. Design de Soluções Computacionais

A revolução industrial causou transformações e a ruptura dos modos convencionais de uma sociedade tecnológica, levando a modelos centrados em humanos que passaram a ter significado, especialmente para uma sociedade inclusiva. As mudanças também abriram novas possibilidades para a acessibilidade e para a interação entre humanos e ferramentas computacionais. Assim, o conceito de design por [Rogers et al. 2013], considera "essencialmente um conjunto de ideias para o design. É composto por cenários, imagens, *mood boards* ou documentos textuais".

A Interação Humano-Computador (IHC) foca na qualidade do uso de sistemas e como isso impacta a vida dos usuários. O estudo da interação está relacionado a como as pessoas utilizam e interagem com os sistemas computacionais. As características humanas também influenciam na própria arquitetura, visto que as interfaces buscam construir sistemas que possam favorecer a experiência de uso [Barbosa e Silva 2010].

IHC é uma área multidisciplinar, pois o desenvolvimento de um sistema requer, às vezes, conhecimento de Psicologia, Sociologia, Design, Linguística e Semiótica.

O que melhora a qualidade, eficácia e satisfação do usuário com sua experiência. [Barbosa e Silva 2010] definem que critérios de usabilidade são um conjunto de fatores que irão qualificar o software conforme uma pessoa pode interagir com o mesmo.

A acessibilidade, por sua vez, faz parte do processo de interação, pois é um critério para acessar o sistema e interagir com o mesmo, sem obstáculos. Pode ser definida como essa flexibilidade que o usuário tem em ter acesso a uma informação, respeitadas as diferentes necessidades. É importante que o usuário não encontre barreiras, seja um usuário com deficiência auditiva, motora ou visual [Barbosa e Silva 2010]. Isso inclui considerar usuários com daltonismo, dislexia e até limitações físicas que impeçam algum acesso [Rogers et al. 2013].

[Rogers et al. 2013] definem Design de Interação (DI) como "projetar produtos interativos para apoiar o modo como as pessoas se comunicam e interagem em seus cotidianos, seja em casa ou no trabalho". Enquanto no DI, existe a visão mais ampla, incluindo a teoria e prática de design, bem como a própria prática de design de experiência, a IHC tem seu foco no design, avaliação e implementação, tratando os fenômenos que rodeiam a interação.

A questão cognitiva é crucial para DI e IHC. Neste trabalho, os aspectos cognitivos estão relacionados à forma como uma pessoa acredita e compartilha notícias falsas. A cognição abrange processos que vão desde a atenção, memória, percepção e raciocínio até a tomada de decisões [Rogers et al. 2013].

A participação do usuário no desenvolvimento visa atender às necessidades de interação, orientando para o sucesso do sistema [Lowdermilk 2019]. Neste trabalho o Design Centrado no Usuário (DCU) [Abras et al. 2004] foi adotado como metodologia para considerar as características dos usuários-alvo no processo de desenvolvimento, em conjunto com o *Design Thinking* [Melo e Abelheira 2015] como metodologia para o processo de design .

A Figura 1 é uma adaptação das etapas de *Design Thinking* [Gaspar et al. 2021]. Ela contém a etapa de **imersão** - com pesquisa e levantamento das dificuldades previamente relatadas em estudos e artigos científicos; **ideação** - com a análise e síntese desses resultados; **desenvolvimento** do protótipo e implementação provisória, com bases em personas e os critérios que nortearam o protótipo, para facilitar a participação do usuário. Após a implementação provisória, os usuários podem visualizar um protótipo e, por meio de **testes** e formulários, opinarem sobre mudanças e melhorias. Esse *feedback* é fundamental para orientar a implementação e o projeto final centrado no usuário. Para a ferramenta de detecção automática de notícias aqui descrita, a participação de pessoas interessadas é crucial para garantir a satisfação desses usuários finais.

A seção a seguir descreve o processo de construção do *corpus* da ferramenta Ada.

## 4. Construção do *corpus* da ferramenta Ada

[Moreira Filho 2021] define que linguística de *corpus* é a área que estuda uma grande quantidade de dados linguísticos reais, com ajuda de recursos computacionais. Pode ter várias aplicações, entre elas para processamento de línguas naturais, tradução e análise forenses. A definição de *corpus*, ainda pelos autores, *"[...]é uma coleção de textos ou transcrições de gravações de áudios, produzidos naturalmente na comunicação humana,*

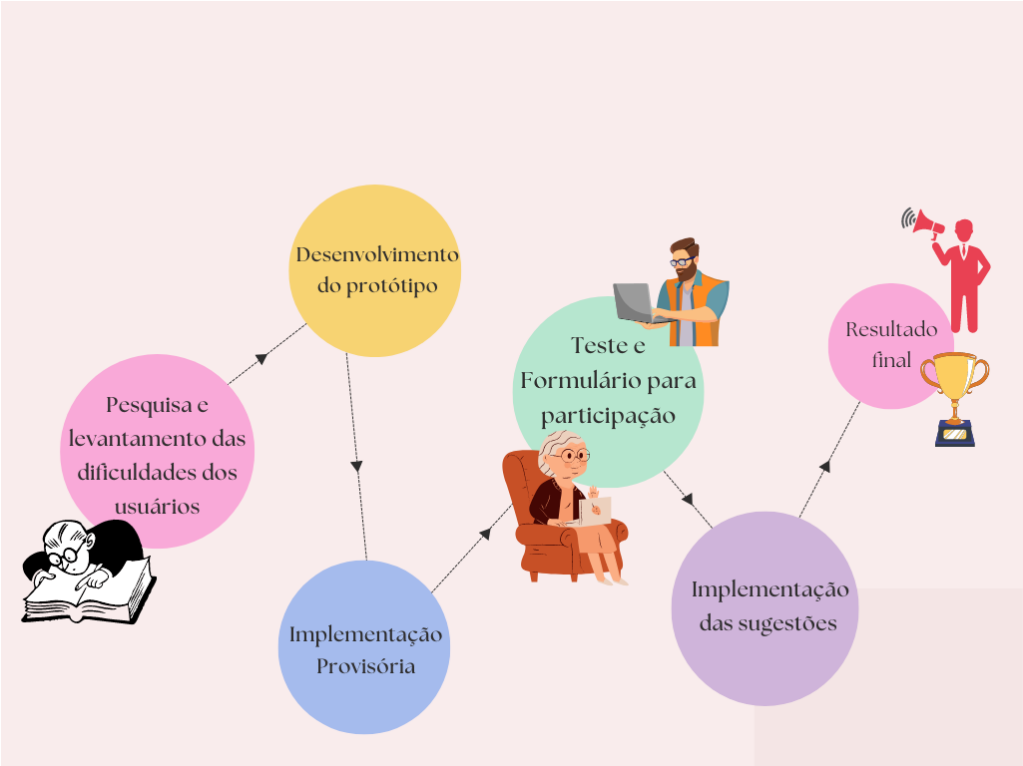



**Figura 1. Etapas do *Design Thinking* na construção da ferramenta.**

*legíveis por computador [...]*", que tem como plural, *corpora*. Assim, a sua construção foi desenvolvida da seguinte forma neste trabalho:

- **Etapa 01**: *Open Source Intellingece* - OSINT (Inteligência de Código Aberto): uma técnica usada por serviços de inteligência, segurança, investigação e negócios [Galvão 2023]. Nessa etapa foram feitos levantamento de dados e informações encontradas nas mídias sociais, metadados, inclusive de vídeos do *Youtube*, Twitter (atual X), *TikTok* e arquivos digitais disponibilizados na w Web, sem que violasse a privacidade ou questões éticas.
- **Etapa 02**: Coleta de Dados: A etapa da coleta foi realizada de forma manual, após a seleção ocorrida na anterior, assim extraindo dados inclusive de documentos oficiais em formato de pdf usando *python*. As notícias foram coletadas, na maioria, dos anos de 2022 e 2023. Nesse trabalho, o *corpus* foi desenvolvido com notícias falsas e verdadeiras. As notícias que compõem foram retiradas de mídias sociais e de notícias da Web, como *The Economist, The Guardian, The Times of Israel*, BBC, CNN, *Foreign Affairs, Financial Times*, Estadão, Folha, R7, Gazeta, Metrópoles e páginas oficiais do governo, como a do Ministério Exterior com o Itamaraty. O *corpus* tem ainda notícias rotuladas como "Científico", nos quais são privilegiados também artigos de revistas científicas na área médica como: *Science* e *The New England Journal of Medicine*. O critério utilizado foi com base no levantamento de temas relevantes. A diversidade visa ampliar a credibilidade das notícias. Foram coletados dados referentes às notícias de 2022 e 2023, cobrindo eleições, copa e guerras. A Figura 2 ilustra os rótulos.
  Foram adicionadas notícias do *corpus* de outra ferramenta, a*Fakecheck*, que tinham relevância temática com o conteúdo atual. Os textos foram etiquetados em

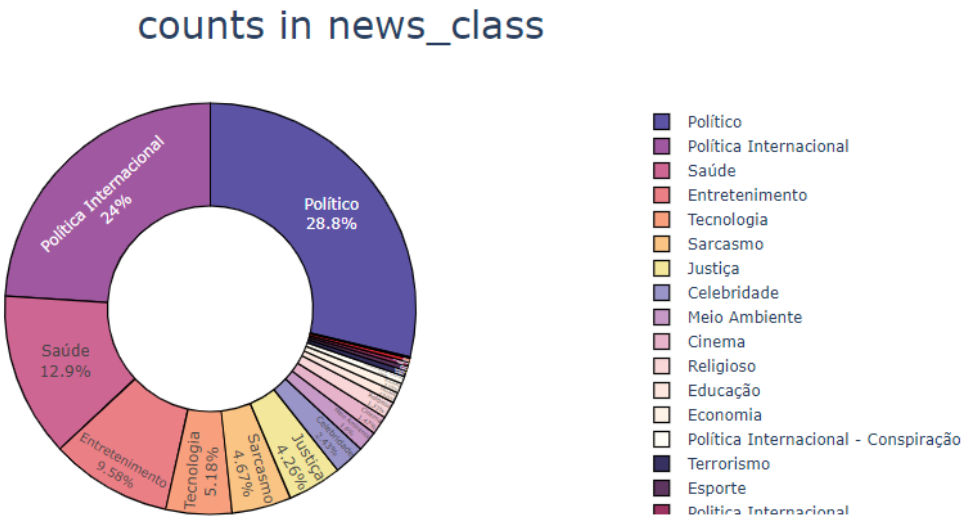
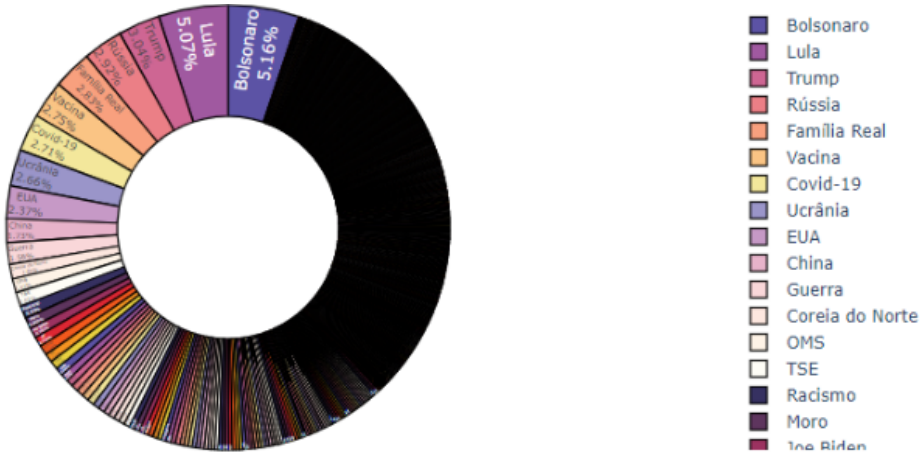



**Figura 2. Divisão de temas no *corpus*.**

*True* (Verdade) e *Fake* (Falso). Entre as notícias de política internacional, foram dadas preferências por fontes internacionais. A diversidade nas fontes foi visando evitar vieses políticos que poderiam comprometer a credibilidade do *corpus*. Um exemplo são as notícias da China. Sobre o país foram adicionadas o mesmo tema com fontes do *The Economist, Foreign Affair*, China *Daily*, CNN.

- **Etapa 03:** Pré-Processamento: é uma etapa importante no processamento de linguagem natural, uma vez que os textos contêm ruído e pontuações que podem afetar a análise. Nessa etapa de pré-processamento foi realizada a limpeza dos textos usando *lemmatização*, *stop-words*, símbolos e números e caracteres especiais. Após a limpeza, os dados foram convertidos de textos para recursos numéricos, uma vez que no aprendizado de máquina é necessário esse processo da conversão. No pré-processamento os dados são melhorados, tornando-os mais fiéis e adequados para o AM(Aprendizado de Máquina) [Faceli et al. 2021]. Na técnica de *Stemming*, geralmente se refere a um processo de cortar as extremidades das palavras para que ela possa atingir o objetivo corretamente, na maioria das vezes, inclui a remoção de afixos derivacionais[Manning et al. 2008]. Reduzi-las significa diminuir o tempo no processamento [Baarir e Djeffal 2021].
  Destaca-se que, para melhorar o desempenho e a precisão dos algoritmos de aprendizado de máquina, é possível o uso de vetorizador TFIDF. O vetorizador, converte o documento de texto em matriz de recursos TFIDF - TF (Frequência de Termo) e IDF (Frequência Inversa de Documentos) [Jouhar et al. 2024].
- **Etapa 4:** Explorando os dados: Nessa fase é possível realizar uma análise mais detalhada, sendo possível visualizar entre as notícias falsas e verdadeiras, se há padrões ou alguma anormalidade, ou a porcentagem de palavras-chave que estão presentes, como é possível ver na Figura 3, onde foram definidas palavras chave para cada texto.Também foi observado que o tamanho das palavras entre as *corpóreas* são diferentes, como pode ser observado na Figura 4 e na Figura 5, em que se pode observar que o tamanho dos textos falsos costumam ser menores e com frases mais curtas, principalmente em textos de mídias sociais.

A seção a seguir descreve o processo de construção da interface da Ada.




Fonte: Autoria própria.

**Figura 3. Divisão de palavras chave no *corpus*.**

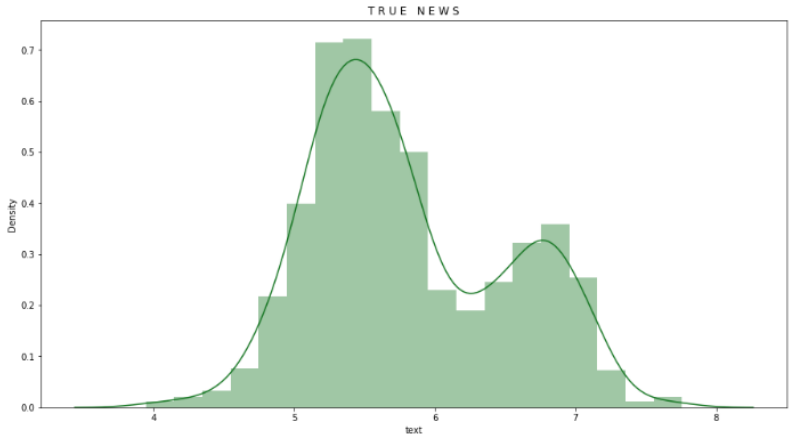



**Figura 4. Tamanho de textos com conteúdo real.**

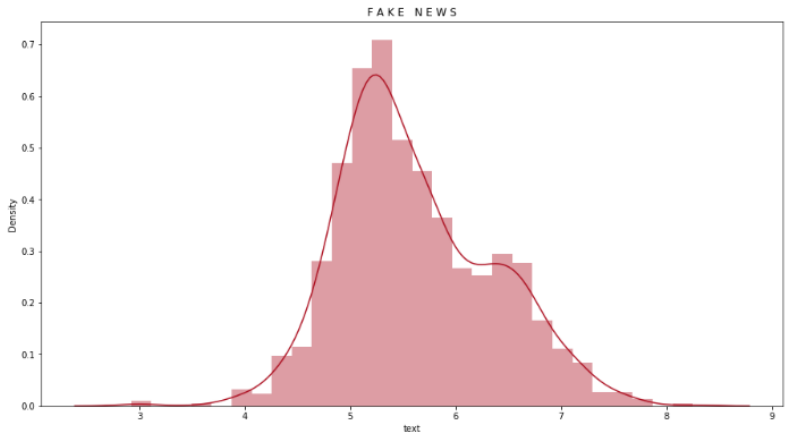



**Figura 5. Tamanho de textos com conteúdo falso.**

## 5. Processo de construção da ferramenta

As pesquisas bibliográficas direcionaram este projeto envolvendo idosos, pessoas mais vulneráveis às notícias falsas e ao seu compartilhamento, devido às questões cognitivas e de conhecimento de mídia. As personas ilustradas a seguir foram criadas para direcionar a construção da ferramenta Ada. Elas foram elaboradas com base em levantamento de perfis de artigos científicos, realizada na etapa de revisão da literatura, bem como por meio do perfil dos alunos do curso de Letramento Digital para o público 60+ realizado na USP.

As duas primeiras personas representam as pessoas com quem os idosos costumam ter mais contato, como seus familiares. Como pode ser observado na Figura 6, embora ambos os adultos com mais de 60 anos tenham perfis e escolaridade diferentes, não possuem habilidades com o uso de tecnologias. As personas são definidas por [Cooper et al. 2014] como arquétipos fictícios, representando um grupo de usuários com certas características e comportamentos semelhantes. Neste trabalho, elas foram criadas com base em análises de dados reais.


**Figura 6. Personas da Ferramenta Ada.**

Ainda em relação às personas, elas auxiliaram na compreensão do contexto dos idosos em relação às *fake news* e apoiaram o processo de design da Ada, utilizando dados do IBGE sobre o uso de tecnologia por esse público. Após a elaboração das personas e um estudo abrangendo 147 estudos acadêmicos de países como Canadá, Estados Unidos, Japão e Brasil, que visava compreender a viabilidade da detecção automática e a criação de uma ferramenta, um primeiro protótipo da ferramenta foi desenvolvido, com a inclusão de um *corpus* contendo 2549 notícias em Português [Lira e Rodrigues 2023].

O protótipo apresentava uma interface gráfica que permitia ao usuário inserir um trecho de notícia e receber imediatamente uma resposta sobre a veracidade do texto, indicando se era *"Fake"ou "True"*. A interface foi projetada com cores que facilitam a utilização, inclusive por usuários com baixa visão, e as imagens possuem legendas com

texto alternativo. Na Figura 7, é possível visualizar as interfaces iniciais da ferramenta Ada e a tela em uso na Figura 8.


**Figura 7. Ada - Página Inicial.**

Após o desenvolvimento do protótipo, idosos do curso de extensão da Universidade de São Paulo (USP)foram convidados a interagir para avaliar a solução. Oito se voluntariaram e o estudo foi realizado em São Carlos na sede do ICMC/USP. Eram 5 mulheres e 3 homens, com idades entre 73 e 64 anos.

Os idosos assinaram inicialmente o TCLE (Termo de Consentimento Livre e Esclarecido) e foram informados que poderiam desistir de participar a qualquer momento. Em seguida, foram apresentados à Ada e receberam instruções sobre o funcionamento da ferramenta. Inicialmente, a pesquisadora explicou aos idosos como e em que local poderiam escrever o texto na ferramenta e tocar para verificar. Foram utilizadas notícias pré-selecionadas. Em seguida, foi explicado quais eram as formas de respostas que poderiam receber da ferramenta.

Em um segundo momento, os idosos eram livres para verificar qualquer tema. Cada participante levou em média (depois da explicação sobre o uso da ferramenta) de 4 a 10 minutos na interação, conforme sua habilidade com o aparelho celular para digitar

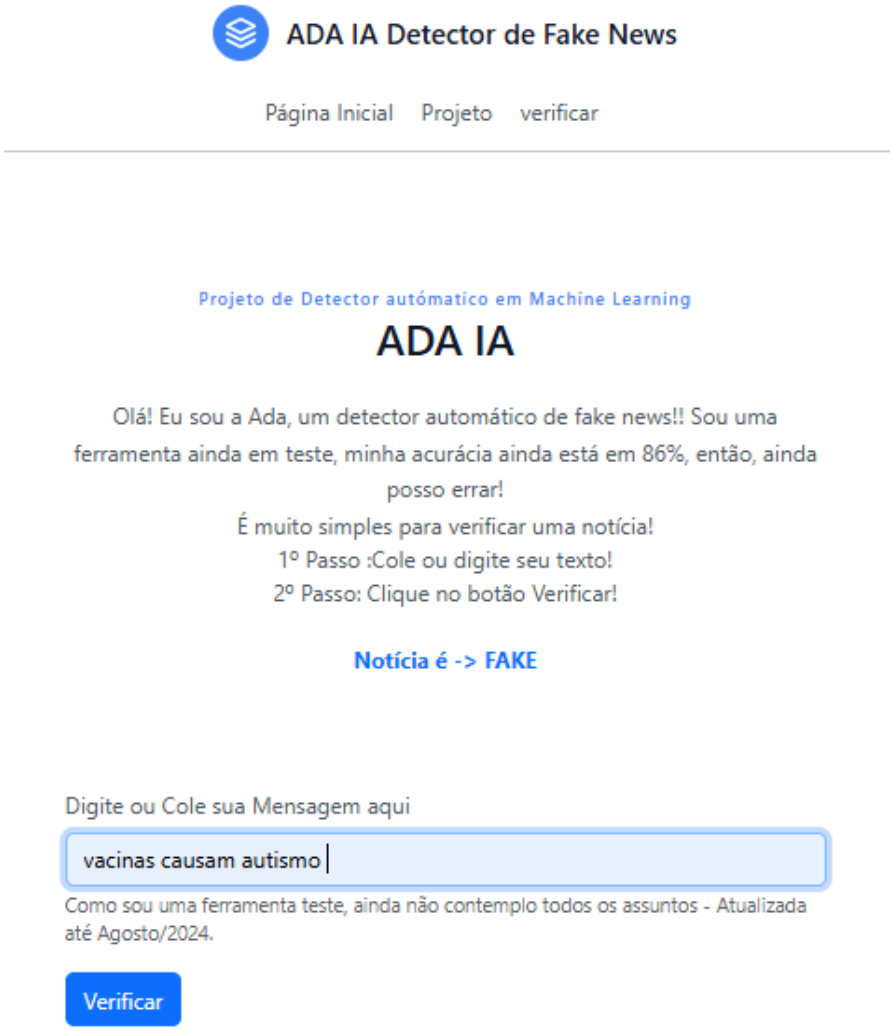



**Figura 8. Ada - Página de verificação.**

a mensagem. Os temas mais consultados foram relacionados a saúde. A Figura 9 ilustra um dos voluntários recebendo explicação do pesquisador.

Após o uso, os idosos deram seus *feedbacks* e responderam a um formulário de usabilidade (SUS - *System Usability Scale*) [Ng et al. 2011].

Nos resultados do SUS, quando perguntados se usariam com frequência, 50% responderam 'concordam totalmente' e 37,5% 'concordam'. Duas das dificuldades encontradas pelos usuários foram a necessidade de copiar e colar um texto e ter que digitar o endereço da plataforma. Isso justifica que, ao serem perguntados sobre a interface do sistema, as respostas tenham sido divididas, considerando que 25% responderam 'concordo' e 25% 'concordam totalmente', como demonstrado na Figura 10.

Todavia, ao serem questionados se estavam confortáveis com o sistema, 62,5% responderam "concordo"e 25% concordaram totalmente. Quando questionados se o sistema era fácil de usar, 62,5% responderam "concordo"e 25% concordaram totalmente. Ambas as respostas podem ser visualizadas no gráfico da Figura 11.

**Figura 9. Explicação sobre o uso da ferramenta Ada.**

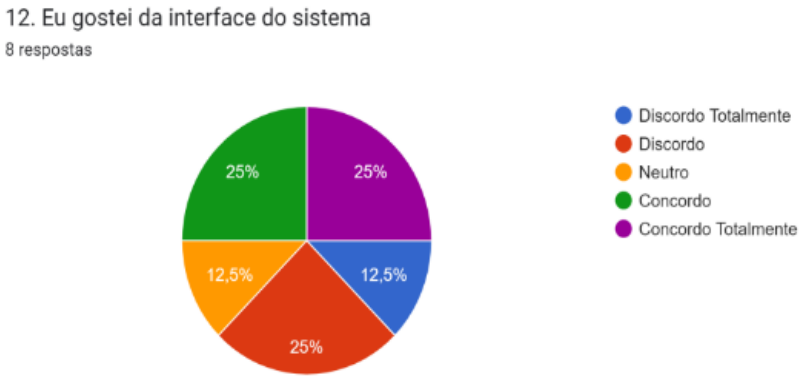



**Figura 10. Resposta ao questionário de avaliação da interface do sistema.**


**Figura 11. Resposta ao questionário de avaliação sobre os participantes estarem confortáveis no sistema**

Dada a dificuldade apontada pelos idosos com a ação de copiar o texto do local de origem e colar na ferramenta Ada para que a mesma verifique, decidiu-se criar um bot para tentar automatizar essa tarefa.

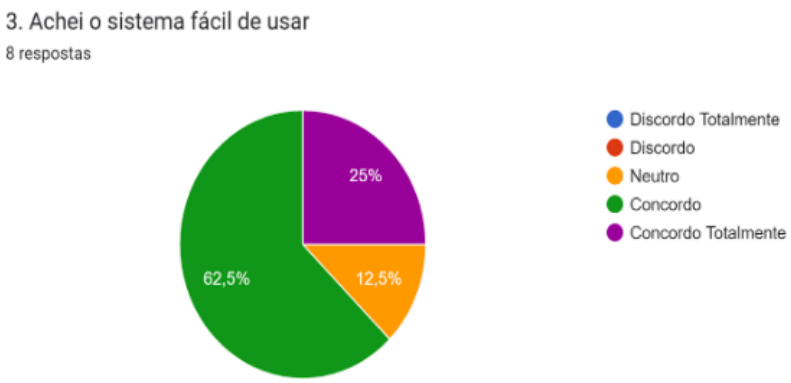



**Figura 12. Resposta ao questionário de avaliação sobre a Ada ser fácil de uso.**

Os idosos também apontaram em seus *feedbacks* que desejavam que a ferramenta fosse um aplicativo ou tivesse uma forma simplificada para verificação, assim, o *bot* seria um sistema mais simplificado para auxiliar os usuários na verificação de informações ou acesso à ferramenta.

A Ada IA, como ferramenta, mais do que detecção automática de notícias, também tem uma função educativa. Na página inicial foram adicionados 6 pontos importantes para boas práticas nas mídias sociais, independente da idade, grau de escolaridade ou conhecimento de letramento digital. Na Figura 13 é exibida a tela da ferramenta com dicas.


**Figura 13. Dicas e boas práticas sugeridas pela ferramenta Ada.**

A opção por ilustrações estáticas na ferramenta, em vez de fotografias, visa criar um ambiente lúdico e imparcial para a verificação de notícias, sem induzir o usuário a qualquer tipo de interferência partidária ou viés que possa afetar a confiança no resultado. Além disso, as figuras pretendem transmitir uma mensagem sobre o propósito da ferramenta, auxiliando no aspecto cognitivo. A ferramenta inclui uma página com

informações sobre a economia das notícias falsas, alertando para a existência de uma indústria que lucra com o discurso de ódio. Portanto, é necessário um esforço conjunto para evitar compartilhar esse tipo de conteúdo.

Tendo em vista esse cenário, este projeto adota uma metodologia híbrida para tentar reduzir a disseminação de notícias falsas pelos idosos. A abordagem considera o uso de um ferramental, tal qual a ferramenta Ada, mas também um suporte ao conhecimento e ao letramento do público alvo sobre o tema.

A seção a seguir descreve tal abordagem, sob a perspectiva da verificação por meio do conhecimento, uma vez que o ferramental foi supracitado.

### 5.1. Abordagem híbrida: Verificação por Conhecimento

Os recursos com aprendizado de máquina, e os métodos tradicionais, podem não ser eficazes com o público idoso uma vez que esses poderiam apresentar dificuldade em ter como apoio somente o recurso de uma ferramenta para detecção automática de *fake news*. O modelo híbrido adotado é ilustrado na Figura 14.

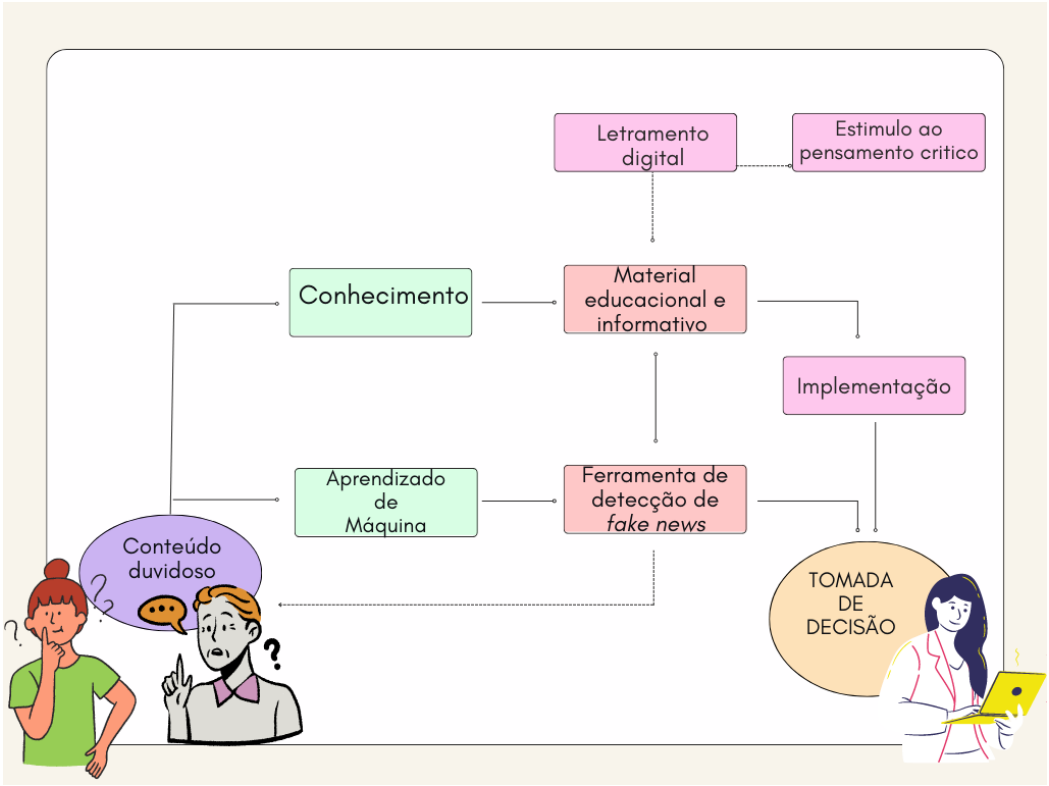



**Figura 14. Abordagem híbrida para diminuição de notícias falsas pelos idosos.**

[Della Vedova et al. 2018] propõem um modelo de detecção com aprendizado de máquina e sinais sociais, e de conteúdo. O segundo, que são os sinais sociais como o engajamento e a interação de um usuário ou conteúdo, por exemplo, curtidas e número de retuítes.

Um estudo apresentado por [Guess et al. 2020] avalia a questão da intervenção e alfabetização digital no mundo real dos americanos e indianos. Os resultados apresentados indicam que intervenções curtas e escaláveis podem ser eficazes para

combater a desinformação. Esse resultado é encorajador para formular estratégias. Os autores explicam que o experimento da pesquisa analisou a eficácia de apresentar às pessoas dicas para detectar notícias falsas. O estudo também destacou que as evidências apresentadas mostram que a deficiência na alfabetização de mídia digital (letramento digital) é um fator importante quando se trata das pessoas acreditarem na desinformação.

[Braddock 2022] afirmam que a inoculação pode ser eficaz em algumas ações preventivas específicas da desinformação, mas com o tempo, não melhora o discernimento na avaliação do mundo real. O autor apresenta resultado após expor um grupo a mensagens de inoculação: *"[...] A inoculação previu positivamente a reatância psicológica, o que, no que lhe concerne, reduziu a intenção de apoiar o grupo extremista"*.

A teoria da inoculação pode ser trabalhada dentro do modelo de conhecimento, uma vez que para [Van der Linden et al. 2020]: *"A inoculação baseia-se na ideia de que, se as pessoas forem avisadas de que podem estar mal-informadas e expostas a exemplos enfraquecidos das maneiras pelas quais podem ser enganadas, elas se tornarão mais imunes à desinformação"*.

[Talwar et al. 2020] comprovam em pesquisa que a conscientização teve efeito positivo sobre o compartilhamento de notícias falsas; porém, autenticar notícias antes de compartilhar não afetou o resultado sobre o compartilhamento, devido às crenças. O estudo sugere que usuários de mídia social que se envolvem em ações corretivas provavelmente não compartilharão notícias falsas, principalmente pela falta de tempo.

Ou seja, a abordagem por meio de verificação por conhecimento se mostra uma frente inevitável em uma sociedade conectada por algoritmos com alta performance.

Para fomentar esta etapa baseada em conhecimento, este projeto tem feito udo de cursos de extensão em que se ensina para o público alvo sobre *fake news*, formas de identificá-las, bem como um senso crítico sobre buscar por fontes confiáveis sempre antes de disseminar a notícia.

As atividades planejadas como demonstrado na Figura15 para dar continuidade a este projeto incluem novos cursos sobre notícias falsas, utilizando estratégias educativas para identificar conteúdo que circula nas mídias sociais, especialmente em redes sociais. O objetivo é promover o senso crítico, abordar noções de responsabilidade civil e criminal no compartilhamento de conteúdo falso, e fornecer instruções sobre o uso de ferramentas tecnológicas para identificação.

Esses cursos serão realizados tanto presencialmente quanto remotamente, e também está sendo considerada a possibilidade de oferecê-los de forma síncrona e assíncrona para alcançar um público mais amplo. Dada a variedade de temas abordados nos cursos, especialistas de outras áreas, como direito e política, serão convidados para enriquecer o debate com o público-alvo.

Além disso, novas rodadas de avaliações serão conduzidas com a ferramenta, utilizando um *corpus* atualizado e as novas implementações na ferramenta Ada.

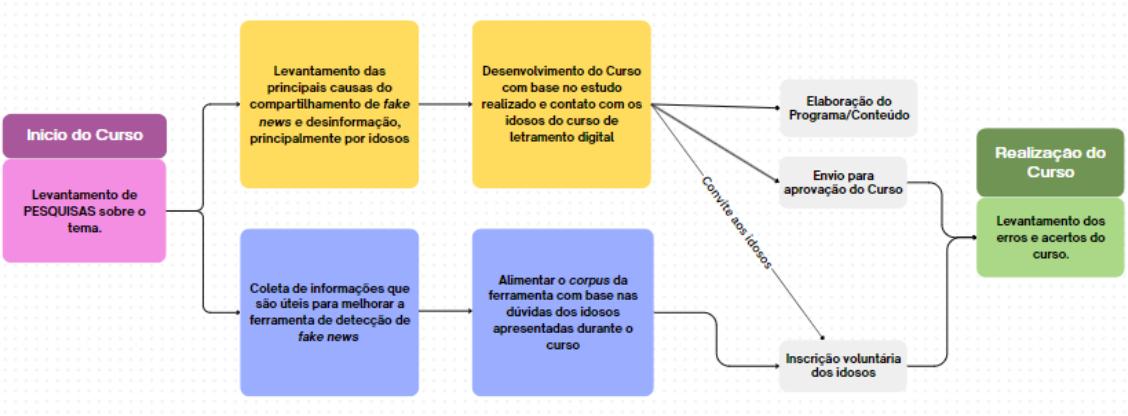



**Figura 15. Planejamento do curso e atividades**

## 6. Limitações do Trabalho

O trabalho apresenta limitações em termos de tempo e tamanho do *corpus*, uma vez que consiste em 3 mil textos atualizados até o final de 2023. Para uma análise mais aprofundada, seria benéfico dobrar esse número. Além disso, é crucial que as fontes sejam diversificadas.

Neste estágio do trabalho, ainda não é possível avaliar a eficácia do uso da ferramenta de detecção de *fake news*. No entanto, espera-se que seja viável determinar esses resultados nas próximas etapas.

Durante o desenvolvimento do trabalho, foi observado que a modelagem dinâmica de opinião pode ser uma ferramenta útil para compreender o papel de fatores sociais na aceitação de determinados temas, bem como na polarização e disseminação de conteúdo falso. Isso representa uma oportunidade para trabalhar com modelos preditivos em relação a temas específicos que possam desencadear ondas de desinformação, como ocorre em situações de pandemia e períodos eleitorais.

## 7. Cuidados Éticos

Este estudo foi submetido e aprovado pelo Comitê de Ética em Pesquisa sob o CAAE 76417223.5.0000.5504, em conformidade com a Resolução CNS n.º 466/2012, a Resolução CNS n.º 510/2016, a Norma Operacional 001/2013 e a Resolução CNS nº 674/2022. Todas as etapas envolvendo seres humanos foram conduzidas com rigor ético, respeitando os princípios da autonomia, beneficência, não maleficência e justiça.

A participação foi voluntária, mediante assinatura do Termo de Consentimento Livre e Esclarecido, com garantia de anonimato, confidencialidade das informações e liberdade de desistência a qualquer momento, sem qualquer prejuízo. Nenhum dado sensível, imagem identificável ou nome de participante foi incluído nesta versão. As imagens eventualmente utilizadas foram devidamente anonimizadas, conforme exigência legal.

As pessoas participantes foram recrutadas de forma ética, com base em critérios previamente definidos, incluindo idade mínima de 60 anos, capacidade cognitiva preservada e interesse em tecnologias digitais. Os possíveis riscos foram minimizados, com suporte psicológico e técnico disponível durante as interações. Não houve

remuneração financeira, respeitando os princípios da equidade e evitando incentivos indevidos.

Este trabalho também seguiu as diretrizes da Lei Geral de Proteção de Dados Pessoais (Lei nº 13.709/2018), com coleta, armazenamento e tratamento de dados pessoais exclusivamente para fins acadêmicos e científicos, de forma segura e transparente.

## 8. Considerações Finais

Este trabalho visa contribuir para reduzir a propagação de *fake news* e desinformação com uma abordagem híbrida para detecção de notícias falsas. Utilizando aprendizado de máquina e abordagem baseada em conhecimento, busca-se capacitar os indivíduos a distinguir entre notícias falsas e verdadeiras, sem torná-los negacionistas ou céticos, mas sim críticos.

Evidências de estudos indicam que a intervenção baseada em conhecimento reduz significativamente a propensão de compartilhar notícias falsas, atuando como uma vacina para mitigar os impactos de informações manipuladas ou enganosas. A proposta demonstra que recursos escaláveis e de baixo custo podem contribuir para tornar as mídias sociais um ambiente mais seguro.

Assim como em certos vírus, em que existem grupos mais vulneráveis, como os idosos que foram priorizados na imunização da COVID-19, é importante que os adultos também passem por esse processo, uma vez que os idosos muitas vezes compartilham notícias falsas por falta de conhecimento, enquanto outras faixas etárias se aproveitam dessa vulnerabilidade para disseminar desinformação e teorias da conspiração.

A Internet possibilitou a disseminação de conteúdo de ódio, manipulado e enganoso nos últimos anos, especialmente durante a pandemia, eleições e conflitos armados. Em um mundo conectado, não é viável garantir a segurança apenas com regulamentações ou leis mais rígidas; os usuários precisam se tornar agentes principais da pacificação, desenvolvendo hábitos de verificação.

*Fake News* e desinformação precisam ser tratados pelos governos como vírus com potenciais efeitos letais para suas democracias, estabilidade e economia, porém, não deve usar dessa justificativa para o autoritarismo e a censura serem instrumentos para salvar sistemas democráticos, uma vez que são opostos. Porém, entender a relação cérebro e computador e sua influência moderadora e modeladora no compartilhamento de notícias falsas, se torna essencial para criar estratégias educacionais.

Este trabalho pode ser visto como inovador porque propõe que a intervenção por meio de um ferramental aumente a eficácia quando utilizada com aprendizagem de máquina, por meio da detecção automática , mas também por meio de uma abordagem por conhecimento, respeitando caso a caso.

## 9. Agradecimentos

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) – Código de Financiamento 001.

*Nota:* Parte do conteúdo deste artigo foi revisada e traduzida com o auxílio de ferramentas de inteligência artificial (*SiderAI* Barra Lateral), com o objetivo de aprimorar

a coerência e a correção linguística. A ferramenta Gemini também foi usada para descrição das imagens.

## Referências

Abras, C., Maloney-Krichmar, D., Preece, J., et al. (2004). User-centered design. *Bainbridge, W. Encyclopedia of Human-Computer Interaction. Thousand Oaks: Sage Publications*, 37(4):445–456.

Baarir, N. F. e Djeffal, A. (2021). Fake news detection using machine learning. In *2020 2nd International Workshop on Human-Centric Smart Environments for Health and Well-being (IHSH)*, pages 125–130.

Barbosa, S. e Silva, B. (2010). *Interação humano-computador*. Elsevier Brasil.

Braddock, K. (2022). Vaccinating against hate: Using attitudinal inoculation to confer resistance to persuasion by extremist propaganda. *Terrorism and political violence*, 34(2):240–262.

Brashier, N. M. e Schacter, D. L. (2020). Aging in an era of fake news. *Current directions in psychological science*, 29(3):316–323. Disponível em: https://journals.sagepub.com/doi/full/10.1177/0963721420915872. Acesso em: Maio de 2025.

Cooper, A., Reimann, R., Cronin, D., e Noessel, C. (2014). *About face: the essentials of interaction design*. John Wiley & Sons.

da Silva Junior, D. P., Alves, D. D., Carneiro, N., Matos, E. d. S., Baranauskas, M. C. C., e Mendoza, Y. L. M. (2024). Grandihc-br 2025-2035 - gc1: New theoretical and methodological approaches in hci. In *Proceedings of the XXIII Brazilian Symposium on Human Factors in Computing Systems*, IHC '24, New York, NY, USA. Association for Computing Machinery.

Della Vedova, M. L., Tacchini, E., Moret, S., Ballarin, G., Di Pierro, M., e De Alfaro, L. (2018). Automatic online fake news detection combining content and social signals. In *22nd Conference of Open Innovations Association (FRUCT)*, pages 272–279. IEEE.

Faceli, K., Lorena, A. C., Gama, J., Almeida, T. A. d., e Carvalho, A. C. P. d. L. F. d. (2021). *Inteligência artificial: uma abordagem de aprendizado de máquina*. LTC. Disponível em: https://repositorio.usp.br/directbitstream/ff933d41-4c3d-4b57-80c2-b4f1c805b1dc/3128493.pdf. Acesso em: Maio 2025.

Galvão, M. C. T. (2023). O uso das ferramentas de comunicação instantânea na comunicação interna de uma instituição: um estudo na ufrn. Disponível em: https://repositorio.ufrn.br/server/api/core/bitstreams/aea3ca71-5520-4b29-966f-235b65d5df67/content. Acesso em: Maio 2025.

Gaspar, R. d. P., Bonacin, R., e Gonçalves, V. (2021). Um estudo sobre atividades participativas para soluções iot para o home care de pessoas idosas. *arXiv preprint arXiv:2103.01078*. Disponível em: https://arxiv.org/abs/2103.01078. Acesso em maio 2025.

Guess, A. M., Lerner, M., Lyons, B., Montgomery, J. M., Nyhan, B., Reifler, J., e Sircar, N. (2020). A digital media literacy intervention increases discernment between mainstream and false news in the united states and india. *Proceedings of the National Academy of Sciences*, 117(27):15536–15545.

Jouhar, J., Pratap, A., Tijo, N., e Mony, M. (2024). Fake news detection using python and machine learning. *Procedia Computer Science*, 233:763–771.

Kong, S. H., Tan, L. M., Gan, K. H., e Samsudin, N. H. (2020). Fake news detection using deep learning. In *2020 IEEE 10th symposium on computer applications & industrial electronics (ISCAIE)*, pages 102–107. IEEE.

Lira, C. e Rodrigues, K. (2023). Ada - ferramenta para detecção automática de notícias falsas. In *Anais Estendidos do XXII Simpósio Brasileiro sobre Fatores Humanos em Sistemas Computacionais*, pages 160–164, Porto Alegre, RS, Brasil. SBC.

Lowdermilk, T. (2019). *Design centrado no usuário: um guia para o desenvolvimento de aplicativos amigáveis*. Novatec Editora.

Manning, C. D., Raghavan, P., e Schütze, H. (2008). *Introduction to information retrieval*, volume 39. Cambridge University Press, Cambridge.

Melo, A. e Abelheira, R. (2015). *Design Thinking & Thinking Design: Metodologia, ferramentas e uma reflexão sobre o tema*. Novatec Editora.

Monteiro, R. A., Santos, R. L. S., Pardo, T. A. S., Almeida, T. A., Ruiz, E. E. S., e Vale, O. A. (2018). Contributions to the study of fake news in portuguese: New corpus and automatic detection results. In *International Conference on Computational Processing of the Portuguese Language*, pages 324–334. Springer.

Moreira Filho, J. L. (2021). *Python para linguística de corpus: guia prático*. Ilexis.net.it Editora.

Neris, V. P. A., Rosa, J. C. S., Maciel, C., Pereira, V. C., Galvão, V. F., e Arruda, I. L. (2024). Grandihc-br 2025-2035 - gc4: Sociocultural aspects in human-computer interaction. In *Proceedings of the XXIII Brazilian Symposium on Human Factors in Computing Systems*, IHC '24, New York, NY, USA. Association for Computing Machinery.

Ng, A. W., Lo, H. W., e Chan, A. H. (2011). Measuring the usability of safety signs: A use of system usability scale (sus). In *proceedings of the International MultiConference of Engineers and Computer Scientists*, volume 2, pages 1296–1301. IAENG Hong Kong. Disponível em: `https://measuringu.com/sus/`. Acesso em: maio 2025.

Rogers, Y., Sharp, H., e Preece, J. (2013). *Design de interação*. Bookman Editora.

Santos, R. L. d. S. (2022). *Detecção automática de notícias falsas em português*. PhD thesis, Universidade de São Paulo. Disponível em: `https://www.teses.usp.br/teses/disponiveis/55/55134/tde-14072022-165613/en.php`. Acesso em maio 2025.

Talwar, S., Dhir, A., Singh, D., Virk, G. S., e Salo, J. (2020). Sharing of fake news on social media: Application of the honeycomb framework and the third-person effect hypothesis. *Journal of Retailing and Consumer Services*, 57:102197.

Van der Linden, S., Roozenbeek, J., et al. (2020). Psychological inoculation against fake news. *The psychology of fake news: Accepting, sharing, and correcting misinformation*, pages 147–169.

Zervopoulos, A., Alvanou, A. G., Bezas, K., Papamichail, A., Maragoudakis, M., e Kermanidis, K. (2020). Hong kong protests: using natural language processing for fake news detection on twitter. In *IFIP international conference on artificial intelligence applications and innovations*, pages 408–419. Springer.