JORGE ANDRÉS CRUZ SILVA

CREDIBILIDADE E VERACIDADE A PARTIR DA RELAÇÃO ENTRE INTELIGÊNCIA ARTIFICIAL E FACT-CHECKING

Relatório de Pós-doutorado realizado na Universidade Estadual Paulista (UNESP), Faculdade de Arquitetura, Artes, Comunicação e Design de Bauru.

Supervisor(a): Prof. Dr. Denis Porto Renó

Agência de fomento (se houver) - Nº do Processo

Bauru

2025

Credibilidade e veracidade a partir da relação entre inteligência artificial e fact-checking

JORGE CRUZ-SILVA

Resumo

Este estudo explora a relação entre inteligência artificial (IA) e credibilidade digital, com foco em fact-checking e noticiabilidade. São analisadas ferramentas automatizadas baseadas em algoritmos de processamento de linguagem natural (NLP) e aprendizado de máquina (Potthast et al., 2018; Lahoti & Prasad, 2022), que permitem detectar padrões discursivos enganosos e verificar dados em tempo real. Também é apresentada uma proposta que utiliza web scraping e a API do ChatGPT para facilitar a análise do conteúdo jurídico publicado na Gazeta do Registro Oficial do Equador. Conclui-se que a combinação de IA, verificação humana e educação midiática é essencial para combater a desinformação e fortalecer a confiança digital (Eysenbach, 2007; Li et al., 2023).

Palavras-clave: Inteligência Artificial, Fact-checking, Credibilidade digital, Web scraping, Desinformação

Abstract

This study explores the relationship between artificial intelligence (AI) and digital credibility, focusing on fact-checking and newsworthiness. Automated tools based on natural language processing (NLP) algorithms and machine learning (Potthast et al., 2018; Lahoti & Prasad, 2022) are analyzed, which allow the detection of misleading discursive patterns and real-time data verification. The study also presents a proposal that employs web scraping and the ChatGPT API to facilitate the analysis of legal content published in the Official Gazette of Ecuador. It concludes that combining AI, human verification, and media literacy is essential to fight disinformation and strengthening digital trust (Eysenbach, 2007; Li et al., 2023).

Keywords: Artificial Intelligence, Fact-checking, Digital credibility, Web scraping, Disinformation

Introdução

A crescente presença da inteligência artificial (IA) no ecossistema digital transformou profundamente a forma como a informação é produzida, distribuída e consumida. Essa evolução trouxe consigo importantes implicações para a construção da credibilidade e da veracidade nos ambientes digitais, particularmente no jornalismo e na verificação de fatos (fact-checking). A proliferação de conteúdos gerados por IA, juntamente com o aumento da desinformação, apresenta desafios inéditos para a sociedade contemporânea, destacando a necessidade de desenvolver ferramentas e estratégias que garantam a integridade informativa.

A credibilidade no ambiente digital é um conceito complexo que se constrói a partir de diversos fatores, incluindo a reputação das fontes, a estrutura da mensagem e as plataformas onde a informação é divulgada. Nesse contexto, a IA tornou-se um elemento-chave que pode tanto fortalecer quanto enfraquecer a percepção de confiabilidade nos conteúdos digitais. Por um lado, a IA oferece novas possibilidades para automatizar processos de verificação, detectar padrões discursivos enganosos e analisar grandes volumes de dados em tempo real. Por outro lado, a mesma tecnologia tem sido utilizada para gerar conteúdos falsos, como os deepfakes, que desafiam os métodos tradicionais de fact-checking e dificultam a distinção entre informações verdadeiras e manipuladas.

O conceito de noticiabilidade também passou por mudanças significativas com a incorporação de sistemas automatizados na seleção de conteúdos informativos. Algoritmos de aprendizado de máquina tornaram-se ferramentas essenciais para identificar temas relevantes, prever tendências informativas e personalizar o fluxo de notícias conforme os interesses do usuário. No entanto, essa automação também apresenta riscos, como a amplificação de vieses algorítmicos, a criação de “câmaras de eco” e a exclusão de vozes minoritárias na esfera pública.

Nesse contexto, o desenvolvimento de ferramentas de veracidade baseadas em IA tem ganhado grande relevância. Propostas inovadoras, como chatbots de verificação em tempo real, plataformas especializadas na detecção de manipulações visuais e ferramentas que analisam o discurso emocional evidenciam os esforços para responder aos desafios informativos atuais. A participação ativa da sociedade civil, dos meios de comunicação e da academia no desenho dessas ferramentas é fundamental para garantir que o desenvolvimento tecnológico esteja orientado a fortalecer a democracia e a informação veraz.

Este trabalho explora a relação entre a inteligência artificial e a credibilidade informativa, destacando os avanços tecnológicos na verificação de fatos e suas implicações na construção da confiança digital. Para isso, apresenta-se uma análise das principais ferramentas de fact-checking automatizado, os fatores que determinam a noticiabilidade no ambiente digital e as iniciativas acadêmicas e tecnológicas que buscam mitigar os efeitos da desinformação. Por fim, expõe-se uma proposta inovadora que integra a IA na análise do conteúdo publicado na Gazeta do Registro Oficial do Equador, como uma ferramenta de consulta e verificação jurídica que promove a transparência e o acesso à informação pública.

A exploração desses aspectos é crucial em um contexto global marcado pela sobrecarga informativa e pela crescente sofisticação das técnicas de manipulação digital. O desenvolvimento de sistemas de verificação baseados em IA, complementados pelo critério humano e pela educação midiática, projeta-se como uma estratégia integral para fortalecer a veracidade informativa e fomentar uma cidadania digital crítica e comprometida com a construção de sociedades democráticas mais justas e equitativas.

ESTADO DA ARTE

Credibilidade no contexto digital

Para definir o conceito de credibilidade nos ambientes digitais em que a sociedade está imersa nas últimas três décadas, é importante destacar que, embora muitos aspectos estejam relacionados aos conceitos clássicos de verdade, veracidade e verossimilhança, também é essencial considerar que vivemos em tempos de realidades sintéticas (Knorr-Cetina, 2009), que nos expõem a uma espécie de "grande desordem" (Stalder, 2018).

Ao abordar o fenômeno complexo da credibilidade na digitalidade, é fundamental reconhecer diversos fatores que se inter-relacionam. Os usuários estão diante de um fluxo constante de informações e dados que provêm de plataformas nas quais fontes e usuários ativam notícias, comentários, reações, etc. Esse rio de informações dificulta o processo de discernimento entre conteúdo confiável e não confiável.

Assim como no processo comunicativo (particularmente no jornalismo) de obtenção de informações, há elementos que influenciam na construção da credibilidade digital: reconhecer a confiabilidade das fontes, identificar a confiabilidade da mensagem, encontrar ambientes que promovam a verificação, utilizar tecnologias que auxiliem na detecção de elementos desinformativos — tanto em texto quanto em áudio e vídeo. Além disso, as dinâmicas nas redes sociais influenciam como um conteúdo é percebido, a partir das diversas interações geradas em torno dessa informação.

O primeiro fator determinante é avaliar a confiabilidade das fontes e suas mensagens. A confiabilidade refere-se à capacidade de produzir resultados semelhantes em condições semelhantes, de forma consistente. Essa confiança a médio e longo prazo está relacionada ao acompanhamento prévio e contínuo dos emissores de informação. Li et al. (2023) mencionam que os usuários tendem a confiar na informação fornecida por autores com verificação nas plataformas de redes sociais ou com trajetória consolidada em seu campo — particularmente em ciência, política e saúde. Segundo os autores, esse é o primeiro mecanismo de defesa contra a propagação de informações erradas ou falsificadas. Assim, uma das variáveis

fundamentais para futuras automatizações nos processos de fact-checking será a capacidade de medir a reputação, a experiência e a profissão (ou ofício) do emissor.

Por outro lado, a credibilidade não depende apenas da fonte, mas também das características da mensagem. A argumentação estruturada, a imparcialidade na apresentação de dados, a inclusão de referências verificáveis e o respaldo em evidências empíricas são elementos essenciais que reforçam a percepção de veracidade em um conteúdo digital. Uma mensagem que apresenta informações claras, objetivas e apoiadas por fontes confiáveis gera maior confiança do que aquela que utiliza afirmações infundadas, emoções exageradas ou estratégias discursivas manipuladoras (Li & Suh, 2015; Abrar et al., 2023).

O processo de construção da credibilidade digital também foi influenciado pela transformação dos intermediários tradicionais nos ecossistemas informativos. O conceito de apomediação, desenvolvido por Eysenbach (2007), descreve a transição de modelos de comunicação em que especialistas e instituições centralizadas validavam a informação, para um sistema em que os usuários recorrem a comunidades online, ferramentas tecnológicas e redes de confiança para avaliar a veracidade dos conteúdos. Nesse contexto, os apomediadores digitais, como fóruns especializados, plataformas de fact-checking e influenciadores nas redes sociais, desempenham um papel crucial na validação do conhecimento.

A crescente dependência da inteligência coletiva para a verificação de dados gerou uma mudança na forma como a confiança é construída nos ambientes digitais, deslocando a autoridade das fontes institucionais para estruturas de validação descentralizadas. Esse fenômeno tem sido particularmente relevante na luta contra a desinformação, onde comunidades organizadas se dedicam a analisar e refutar informações errôneas de forma colaborativa.

A automação da verificação de fatos tem sido outra das estratégias implementadas para fortalecer a credibilidade nos ambientes digitais. Graças ao desenvolvimento da inteligência artificial, surgiram sistemas capazes de detectar padrões na desinformação, verificar

afirmações em tempo real e comparar informações com bancos de dados verificados. No entanto, apesar dos avanços tecnológicos nessa área, a intervenção humana continua sendo essencial para evitar vieses algorítmicos e garantir a precisão na avaliação dos dados (Castellón, 2021).

Os algoritmos de fact-checking automatizado podem identificar inconsistências e analisar grandes volumes de conteúdo de forma eficiente, mas sua eficácia é limitada pela necessidade de interpretação contextual e análise crítica — aspectos nos quais o julgamento humano continua sendo indispensável. Nesse sentido, a combinação de ferramentas automatizadas com processos de verificação conduzidos por especialistas permite construir um sistema mais robusto para a detecção e mitigação da desinformação.

O envolvimento dos usuários na geração e validação de conteúdos digitais tem sido outro fator determinante na construção da credibilidade online. As plataformas que permitem a co-criação de conteúdo geram um maior senso de transparência e confiança, pois os usuários tornam-se agentes ativos na produção da informação.

Os espaços onde as comunidades podem debater, corrigir erros e fornecer evidências para sustentar ou refutar afirmações fomentam um ecossistema informativo mais confiável. Exemplos disso são fóruns e subreddits dedicados à verificação de fatos, onde usuários com diferentes níveis de conhecimento colaboram para analisar a veracidade de publicações virais. Essa dinâmica não só fortalece a credibilidade dos conteúdos, como também promove a alfabetização midiática ao incentivar os usuários a adotar uma postura crítica diante da informação que consomem (Eysenbach, 2007; He & He, 2022). O design visual dos ambientes digitais é outro elemento que influencia a percepção de credibilidade. Ao interagir com plataformas e sites, os usuários recorrem a heurísticas visuais para avaliar a confiabilidade de um conteúdo. A aparência profissional de uma página, o uso de tipografias legíveis, a organização estruturada do conteúdo e a ausência de erros ortográficos e tipográficos são fatores que influenciam na percepção de confiança de um site (Chandrashekar & Hockema, 2009).

Em contraste, sites com designs caóticos, excesso de publicidade invasiva ou estrutura desordenada costumam ser percebidos como menos confiáveis, independentemente da qualidade do conteúdo que apresentam.

Por fim, as dinâmicas próprias das redes sociais desempenham um papel crucial na percepção de credibilidade nos ambientes digitais. Fatores como a quantidade de "curtidas", comentários e compartilhamentos que uma publicação recebe podem afetar a percepção que os usuários têm sobre sua veracidade. Publicações com alto nível de engagement tendem a ser consideradas mais confiáveis, embora esse fenômeno possa levar à viralização de informações falsas quando os usuários interpretam a popularidade como um indicador de autenticidade (Li et al., 2023).

Além disso, a interatividade e a dependência dos usuários de uma plataforma específica influenciam a forma como avaliam a informação. Redes que facilitam o debate e o acesso a fontes diversas tendem a gerar maior confiança do que aquelas que restringem a interação e limitam o acesso a múltiplas perspectivas (Li & Suh, 2015).

A credibilidade digital é o resultado de um processo complexo que envolve múltiplas dimensões, desde a avaliação da fonte e da mensagem até a participação comunitária, a verificação automatizada, o design das plataformas e as dinâmicas de interação nas redes sociais.

Em um contexto de crescente desinformação, é fundamental promover estratégias de alfabetização midiática que permitam aos usuários desenvolver habilidades críticas para avaliar a veracidade dos conteúdos digitais. Da mesma forma, o design de plataformas deve favorecer a transparência e a participação ativa dos usuários na verificação de informações, reforçando assim a confiança nos ecossistemas digitais. Somente por meio de uma abordagem integrada, que combine tecnologia, comunidade e educação, será possível fortalecer a credibilidade no ambiente digital e mitigar o impacto da desinformação na sociedade contemporânea..

Veracidade e Noticiabilidade

Os conceitos de veracidade e noticiabilidade são fundamentais na prática jornalística e têm sido objeto de amplo debate acadêmico. A busca pela verdade é frequentemente considerada o pilar essencial do jornalismo. O Código de Ética da Society of Professional Journalists estabelece que "o jornalismo ético deve ser preciso e justo" (Society of Professional Journalists, 2014). Por outro lado, a noticiabilidade depende do critério de jornalistas e editores, que determinam quais eventos ou temas merecem destaque, influenciando assim a percepção pública sobre a relevância desses temas (Park & Kaye, 2023). Embora esses conceitos tenham sido historicamente essenciais no jornalismo, o surgimento da inteligência artificial (IA) nos últimos anos gerou novos desafios e oportunidades.

Por um lado, a IA tem o potencial de melhorar a precisão informativa por meio da automatização da detecção de informações falsas, aumentando a velocidade e a exatidão dos processos de verificação. Por exemplo, algoritmos de IA podem ser treinados para identificar notícias ou imagens falsas e prevenir sua disseminação (Shu et al., 2017). Isso é particularmente valioso no ambiente midiático atual, onde a propagação de desinformação é uma ameaça constante. A IA também pode ajudar a detectar inconsistências ou contradições em declarações públicas de políticos, figuras públicas ou fontes informativas, o que fortalece a confiabilidade do jornalismo (Kanagavalli & Priya, 2022; Farooq & Makhdomi, 2021; Fletcher et al., 2020). Ao automatizar o processo de verificação, os jornalistas podem se concentrar em outras tarefas como análises mais profundas ou investigações jornalísticas.

No entanto, existem preocupações sobre a confiabilidade total da IA na detecção da verdade. O uso de IA na produção de notícias pode agravar problemas de viés e discriminação. Por exemplo, algoritmos de IA podem ser suscetíveis a erros e vieses, especialmente se forem treinados com conjuntos de dados parciais ou se faltarem critérios de transparência (Mehrabi et al., 2021). Isso destaca a importância de projetar e treinar sistemas de IA de forma ética e transparente, garantindo que os processos de decisão sejam baseados em critérios equilibrados e livres de subjetividade. Além disso, sistemas de IA têm potencial para gerar informações falsas, como ocorre com os vídeos deepfake (Shin & Lee, 2022). Isso reforça a

necessidade de que jornalistas e organizações de notícias utilizem a IA com cautela na produção de informações.

O uso da IA para identificar conteúdo noticioso pode melhorar a eficiência e a diversidade na seleção de notícias, mas também pode reforçar a hegemonia de fontes tradicionais, excluindo outras vozes que não têm recursos ou capacidades para competir no ambiente midiático. Algoritmos de IA podem analisar grandes volumes de dados provenientes de redes sociais ou artigos da web para identificar tendências emergentes ou temas de interesse público, permitindo que os meios de comunicação abordem temas que, de outra forma, passariam despercebidos. Outra aplicação possível é a busca automatizada em bases de dados governamentais para identificar tendências, resultados, contratos ou outras informações geradas por e para a sociedade (Schütz, 2023). No entanto, também existe o risco de que a IA reforce vieses na seleção de notícias, especialmente se os algoritmos forem projetados para priorizar fontes consolidadas ou excluir vozes menos representadas (Diakopoulos, 2016). Isso destaca a necessidade de considerar cuidadosamente o design e o uso de sistemas de IA na produção jornalística, bem como de garantir que perspectivas diversas estejam representadas na cobertura midiática.

Por outro lado, o uso da IA na seleção de conteúdos noticiosos pode contribuir para o fenômeno da "câmara de eco", em que indivíduos são expostos apenas a notícias que confirmam suas crenças preexistentes (Bromley, 2012). Isso ocorre porque os algoritmos de IA geralmente priorizam conteúdos semelhantes aos que o usuário já consumiu anteriormente, com o objetivo de personalizar sua experiência informativa. Embora isso possa ser benéfico para oferecer conteúdos relevantes aos interesses do usuário, também pode reforçar preconceitos e limitar a exposição a perspectivas alternativas.

Caracterização da Noticiabilidade

A noticiabilidade é determinada por diversos fatores que permitem aos meios de comunicação avaliar a importância e o interesse de um evento ou tema. Embora existam

múltiplos elementos que influenciam essa avaliação, algumas características são amplamente reconhecidas como determinantes-chave no jornalismo.

Uma das principais características é a atualidade, ou seja, a temporalidade e novidade dos acontecimentos. As notícias recentes ou de última hora tendem a atrair mais atenção do que aquelas que já foram abordadas anteriormente. A imediatidade costuma aumentar a relevância do conteúdo, já que o público valoriza mais informações recentes. Shoemaker e Reese (2014) destacam que "a frescura ou a novidade do evento é um fator crítico na determinação da noticiabilidade" (p. 60).

Outra característica-chave é a relevância, entendida como a relação direta que a notícia tem com o público. A relevância é determinada por fatores como proximidade geográfica, interesse cultural ou conexão pessoal que o público tenha com o tema abordado. McQuail (2010) afirma que "as notícias devem ter relevância para algum segmento da audiência, especialmente em relação aos seus interesses culturais, sociais, econômicos ou políticos" (p. 97).

O impacto ou as consequências que uma notícia pode gerar também influenciam sua noticiabilidade. Notícias que têm um efeito significativo na sociedade ou nos indivíduos tendem a ganhar maior visibilidade. Esse impacto pode se manifestar em áreas econômicas, políticas ou sociais (Shoemaker & Reese, 2014).

A novidade é outro fator determinante. Notícias que apresentam informações únicas ou inesperadas são mais noticiáveis do que aquelas que seguem padrões convencionais. Isso é especialmente verdadeiro quando os eventos desafiam suposições estabelecidas ou rompem com as expectativas do público. Como afirmam Kovach e Rosenstiel (2014), "a natureza inesperada ou surpreendente de um evento frequentemente contribui para sua noticiabilidade" (p. 32).

Por fim, a proeminência ou o reconhecimento social das pessoas envolvidas na notícia também desempenha um papel importante. Personagens influentes ou com alta visibilidade

midiática costumam atrair maior cobertura jornalística do que aqueles com menor notoriedade pública. O Código de Ética da Society of Professional Journalists (2014) enfatiza que "os jornalistas devem ser vigilantes ao questionar seus próprios motivos, assim como os das pessoas que cobrem" (p. 4).

Inteligência Artificial e Desinformação

A integração da inteligência artificial nos processos de geração e difusão de informações transformou radicalmente o ecossistema midiático global. Embora essa tecnologia tenha permitido avanços significativos na criação de conteúdos educativos, jornalísticos e de entretenimento, também facilitou o desenvolvimento de estratégias de manipulação e desinformação com um impacto sem precedentes. A capacidade da inteligência artificial de automatizar a produção de informações, amplificar vieses algorítmicos, gerar conteúdos falsificados hiper-realistas e distorcer a percepção da realidade em contextos políticos e eleitorais apresenta desafios urgentes na luta contra a desinformação. Nesse sentido, é fundamental analisar os mecanismos pelos quais a inteligência artificial influencia a propagação de narrativas enganosas e avaliar as estratégias de mitigação que poderiam contrabalançar esses efeitos.

Um dos principais âmbitos em que a inteligência artificial impacta a circulação de informações é a automatização do conteúdo. Por meio de sistemas de aprendizado de máquina e processamento de linguagem natural, a IA pode gerar textos, imagens, áudios e vídeos com um grau de sofisticação que torna cada vez mais difícil distinguir entre informações autênticas e manipuladas. Essa automatização tem aplicações legítimas, como a criação de conteúdo jornalístico assistido por IA ou a otimização de estratégias de comunicação em plataformas digitais, mas também pode ser utilizada com fins mal-intencionados.

Um exemplo alarmante desse fenômeno é o desenvolvimento de tecnologias de deepfake, que permitem a criação de mídias audiovisuais falsificadas com uma precisão sem precedentes. Essas ferramentas podem ser empregadas para distorcer a realidade, desacreditar figuras públicas, fabricar provas falsas e manipular a opinião pública, o que

representa um desafio significativo para a integridade do ecossistema informativo (Boonprakong et al., 2023; Tagarev et al., 2021; Shin & Shin, 2023). A capacidade dos deepfakes de alterar narrativas por meio da manipulação de imagens e vídeos introduz uma nova dimensão na propagação da desinformação, difuminando as fronteiras entre o real e o fictício (Celiktutan et al., 2024; Gao et al., 2024).

Além da automatização do conteúdo, outro aspecto crítico da interação entre IA e desinformação é a manipulação algorítmica. Os sistemas de recomendação e curadoria de conteúdo operam sob princípios que priorizam a personalização da experiência do usuário, o que pode resultar na amplificação de vieses cognitivos e na criação de câmaras de eco. Os algoritmos das plataformas digitais analisam as preferências dos usuários e otimizam a visibilidade do conteúdo com base em padrões de interação anteriores, reforçando assim a exposição a informações que confirmam suas crenças preexistentes. Esse fenômeno, conhecido como filter bubble, não só reduz a diversidade informativa, como também aumenta a suscetibilidade à desinformação ao limitar a exposição a pontos de vista alternativos (Kaswan et al., 2023; Rajput & Arora, 2024).

Ao mesmo tempo, os vieses inerentes aos sistemas de inteligência artificial podem amplificar desigualdades sociais e promover narrativas discriminatórias, já que os modelos de aprendizado de máquina dependem de dados de treinamento que refletem padrões históricos de informação, muitas vezes carregados de preconceitos sistêmicos (Abbas & Taeihagh, 2024). Em consequência, o design e a implementação desses algoritmos requerem um escrutínio rigoroso para evitar que a inteligência artificial se transforme em um catalisador da desinformação.

O desenvolvimento e a proliferação de conteúdo falsificado por meio da inteligência artificial têm tido um impacto considerável na estabilidade social e política, especialmente em contextos eleitorais. Durante processos democráticos, a difusão massiva de informações errôneas ou manipuladas pode alterar a percepção do eleitorado e minar a confiança nas instituições.

Os deepfakes, por exemplo, têm sido utilizados para fabricar declarações falsas de candidatos, desacreditar campanhas políticas e gerar desconfiança nos resultados eleitorais (Gao et al., 2024; Naitali et al., 2023). Essa capacidade de manipulação midiática levanta sérias preocupações em torno da integridade da democracia, já que a desinformação estratégica pode polarizar ainda mais as sociedades e enfraquecer a legitimidade dos processos eleitorais.

Nesse contexto, as redes sociais se tornaram uma das principais ferramentas de difusão de desinformação eleitoral, com algoritmos projetados para maximizar o alcance de conteúdos virais sem considerar sua veracidade (Mauk & Grömping, 2024; Ergüney, 2023; Pérez-Curiel et al., 2021). As plataformas digitais têm sido empregadas como veículos de campanhas de manipulação dirigidas, nas quais atores políticos e grupos de interesse utilizam estratégias algorítmicas para influenciar o comportamento dos eleitores e gerar narrativas polarizadoras.

Diante desses desafios, surgiram diversas iniciativas para conter o impacto da desinformação gerada por IA em processos políticos e sociais. Uma das estratégias mais relevantes foi o desenvolvimento de tecnologias de detecção de deepfakes, baseadas em redes neurais convolucionais (CNN) e modelos avançados como os Vision Transformers (ViT), que buscam identificar inconsistências nos padrões visuais e auditivos dos conteúdos falsificados (Celiktutan et al., 2024; Naitali et al., 2023).

Contudo, a constante evolução das técnicas de geração de deepfakes transformou essa corrida tecnológica em um desafio contínuo, já que os modelos de inteligência artificial empregados para detectar conteúdo manipulado devem ser constantemente atualizados para acompanhar as inovações na falsificação digital.

Além disso, iniciativas de verificação de informações e fact-checking ganharam maior relevância na luta contra a desinformação. Organismos de supervisão, acadêmicos e plataformas digitais implementaram mecanismos de transparência e prestação de contas para mitigar a propagação de conteúdos enganosos, embora essas medidas ainda enfrentem limitações em termos de escalabilidade e eficácia.

A interseção entre inteligência artificial e desinformação representa um desafio crítico para a construção de um ecossistema informativo confiável. A capacidade da IA de gerar conteúdo automatizado, amplificar vieses algorítmicos, produzir mídias hiper-realistas manipuladas e distorcer a percepção pública em contextos políticos transformou a maneira como a informação circula e é consumida.

Diante desse panorama, é essencial desenvolver estratégias multidimensionais que abordem tanto o design ético de algoritmos quanto a implementação de tecnologias de verificação e educação midiática.

A regulação dessas tecnologias, juntamente com a promoção de uma cidadania digital crítica e informada, pode contribuir para mitigar os efeitos nocivos da desinformação na era da inteligência artificial. A evolução desses desafios exige uma resposta coordenada entre pesquisadores, legisladores, tecnólogos e a sociedade em geral para garantir que a inteligência artificial seja uma ferramenta para a democratização do conhecimento e não um instrumento de manipulação informativa.

Verificação de Fatos (Fact-checking)

A verificação de fatos é um componente essencial do jornalismo que tem ganhado crescente relevância na era dos meios digitais, onde a desinformação e as notícias falsas podem se disseminar com facilidade (Mena, 2019). Os métodos tradicionais de fact-checking são realizados manualmente, o que implica verificar a informação por meio da comparação com múltiplas fontes. Esse processo costuma ser trabalhoso e, em muitos casos, não abrange a totalidade dos dados disponíveis (Wardle, 2017).

O surgimento da inteligência artificial (IA) permitiu que o fact-checking se tornasse mais eficiente e preciso. Os sistemas automatizados baseados em IA utilizam algoritmos de processamento de linguagem natural (Natural Language Processing — NLP) para analisar textos e detectar informações falsas ou enganosas (Potthast et al., 2018). Esses algoritmos identificam padrões e inconsistências em grandes conjuntos de dados, facilitando a detecção

de desinformação. A verificação automatizada por meio de IA pode ser realizada em tempo real, permitindo que os meios de comunicação respondam rapidamente diante da disseminação de informações falsas (Lahoti & Prasad, 2022).

Existem diversas ferramentas de fact-checking automatizado que foram desenvolvidas nos últimos anos. Exemplos notáveis incluem o Duke Reporters' Lab, Tech & Check Cooperative, Full Fact Live e Trends, assim como o Chequeabot do Chequeado (Sittmann et al., 2020). Essas ferramentas permitem monitorar eventos políticos e detectar padrões de desinformação nas redes sociais.

No entanto, essas tecnologias apresentam limitações significativas. Embora sejam particularmente eficazes para monitorar temas políticos e verificar conteúdo viral, costumam ter dificuldades para analisar afirmações complexas ou informações que ainda não foram previamente verificadas. Além disso, a tecnologia NLP enfrenta dificuldades em ambientes autoritários ou onde a informação não está estruturada para análise computacional.

É importante destacar que a automação do fact-checking não representa uma solução independente, mas deve ser compreendida como um complemento a outras abordagens. Os sistemas automatizados requerem um investimento significativo em pesquisa e desenvolvimento, o que pode dificultar sua adoção por parte de organizações jornalísticas menores.

Apesar de suas vantagens, o uso de IA para a verificação de fatos também apresenta desafios importantes. Um dos principais problemas está na falta de transparência dos algoritmos utilizados nesses sistemas. É fundamental que os meios de comunicação expliquem como esses algoritmos funcionam para garantir que não introduzam vieses nem perpetuem informações falsas (Diakopoulos, 2016).

IA e Noticiabilidade: Potencial e Desafios

A incorporação da inteligência artificial (IA) no jornalismo introduziu novas dinâmicas na identificação de conteúdo noticioso. Para compreender como a IA pode potencializar a

detecção de informações relevantes, é importante analisar as características-chave da noticiabilidade e como essas se aplicam por meio de sistemas automatizados.

A IA, por ser capaz de processar grandes volumes de dados em tempo real, transforma os critérios tradicionais de seleção de notícias e permite gerar conteúdo mais preciso, personalizado e oportuno.

Uma das principais características é a atualidade, que se refere à capacidade das notícias de informar sobre eventos recentes ou de última hora. A IA se destaca nesse aspecto graças à sua habilidade de rastrear múltiplas fontes de dados e detectar fatos relevantes com rapidez. Por exemplo, a Associated Press implementou sistemas de IA que geram automaticamente relatórios financeiros de empresas em questão de segundos após sua publicação. Essa velocidade é crucial para garantir que a audiência receba informações oportunas sobre eventos de relevância imediata.

Além disso, os sistemas de IA podem ser programados para rastrear a evolução de eventos em tempo real, identificando as fontes mais confiáveis, verificando a precisão da informação e acompanhando atualizações para oferecer uma narrativa informativa completa e precisa.

Em relação à relevância, a IA pode analisar dados de usuários, interações em redes sociais e padrões de consumo informativo para personalizar as recomendações de notícias. Essa personalização permite que os usuários recebam conteúdos ajustados a seus interesses específicos, facilitando o acesso a informações que impactam diretamente suas vidas. Para otimizar esse aspecto, os algoritmos podem aprender a detectar as preferências dos usuários, seu histórico de leitura e a análise semântica das afirmações para avaliar o valor informativo de cada notícia.

Ao mesmo tempo, esses sistemas podem identificar possíveis vieses no conteúdo informativo e sugerir fontes alternativas que ofereçam uma visão mais equilibrada dos fatos.

A capacidade preditiva da IA também permite identificar conteúdo com alto potencial de impacto. Ao monitorar redes sociais, fóruns digitais e outras plataformas online, os algoritmos podem detectar tendências emergentes que possam ter consequências sociais,

políticas ou econômicas significativas. Essa detecção precoce permite que os jornalistas abordem esses temas de forma oportuna, contribuindo para uma cobertura midiática mais eficaz.

Além disso, os sistemas de IA podem ser programados para rastrear a atividade em redes sociais, analisar os padrões de participação dos usuários e detectar possíveis fontes de desinformação ou campanhas de manipulação informativa, ajudando assim a prevenir a propagação de notícias falsas.

A novidade é outro elemento essencial no qual a IA pode se destacar. Ao empregar técnicas de aprendizado não supervisionado e modelagem temática, os sistemas automatizados podem identificar histórias que ainda não foram amplamente divulgadas. Essa capacidade é especialmente útil para detectar eventos locais, temas de nicho ou situações que não receberam atenção midiática suficiente.

Para otimizar essa função, os algoritmos podem aprender a identificar mudanças na agenda informativa, detectar variações no discurso público e reconhecer fontes emergentes de informação que ofereçam perspectivas inovadoras sobre temas-chave. Essa capacidade de exploração constante permite que as redações ampliem suas fontes informativas e se mantenham atentas a tendências que poderiam passar despercebidas.

Por fim, a IA pode analisar a proeminência dos atores envolvidos nas notícias. Por meio da análise de redes sociais, algoritmos de detecção de influência e sistemas de análise de sentimento, é possível identificar figuras influentes, avaliar seu poder social ou político e determinar seu impacto na agenda informativa.

Os sistemas automatizados podem monitorar as conexões sociais dos atores relevantes, analisar suas afiliações políticas e medir seu nível de interação em meios digitais para detectar possíveis vieses ou padrões de manipulação. Essa capacidade é essencial para que os meios de comunicação possam contextualizar adequadamente a informação, evitando cair no sensacionalismo ou na propagação de conteúdo que distorce a realidade.

Para que esses sistemas de IA sejam eficazes na avaliação da noticiabilidade, é necessário que sejam programados para reconhecer variáveis específicas que reforcem o processo de verificação de fatos.

Em relação à **atualidade**, os sistemas devem aprender a identificar eventos em andamento, avaliar a confiabilidade das fontes que informam sobre eles e monitorar a velocidade com que a informação se espalha para evitar a propagação de dados errôneos.

Quanto à **relevância**, é fundamental que os algoritmos possam avaliar o contexto e o histórico das notícias para identificar possíveis vieses ou propaganda.

Além disso, os sistemas de IA podem avaliar o impacto das notícias por meio da análise do sentimento nas reações do público, detectando possíveis consequências sociais ou políticas.

No caso da **novidade**, a IA pode identificar mudanças na agenda informativa e detectar inconsistências no conteúdo para evitar a difusão de informações manipuladas.

Finalmente, em relação à **proeminência**, os sistemas podem identificar o poder social, a reputação e a influência de figuras públicas envolvidas nos eventos noticiosos.

Modelo Executivo de IA na Noticiabilidade

Característica	**Aplicação da IA**	**Variáveis-chave**
Atualidade	Detecção e verificação em tempo real	Fontes confiáveis, velocidade de difusão
Relevância	Personalização de conteúdo	Preferências, histórico do usuário
Impacto	Identificação de eventos com grande alcance	Redes sociais, padrões de interação
Novidade	Detecção de temas emergentes	Mudanças na agenda, variação discursiva
Proeminência	Análise de figuras influentes	Reputação, poder político, influência

HACKATON

Tive a oportunidade de participar como expositor na mesa-redonda intitulada "A IA e a desinformação nos processos eleitorais", realizada no âmbito do evento MediaHack: Tecnologias Cidadãs para Decisões Informadas nas Eleições Equador 2025. Esse espaço reuniu especialistas em comunicação, tecnologia e inteligência artificial, e foi organizado pela UNESCO, OpenLab, Hacks/Hackers e pelo Observatório de Comunicação da PUCE. O objetivo foi oferecer conhecimentos técnicos e metodológicos para enfrentar a crescente problemática da desinformação no contexto eleitoral.

Durante minha intervenção, abordei temas-chave como o impacto da IA na manipulação informativa, a proliferação de notícias falsas e o uso de ferramentas tecnológicas para mitigar seus efeitos. Falei sobre casos internacionais relevantes, como o uso de sistemas automatizados por veículos como The Washington Post (Heliograf) e Bloomberg (Cyborg), que ilustram como a IA está sendo progressivamente incorporada na criação de conteúdo informativo. Enfatizei a crescente ameaça representada pelos deepfakes e pelo astroturfing, estratégias que utilizam bots e manipulação algorítmica para distorcer a opinião pública.

Na minha apresentação, destaquei três linhas de ação prioritárias que considero fundamentais para enfrentar a manipulação digital:

1. Promover a alfabetização midiática, fortalecendo o pensamento crítico na audiência;
2. Impulsionar o desenvolvimento e implementação de ferramentas de verificação digital como InVID, Botometer e FactCheck.org;
3. Apoiar a criação de marcos regulatórios que promovam a transparência no uso da IA em ambientes informativos, seguindo exemplos como o AI Act europeu.

Figura 1: Cartaz do Conversatório Hackaton



Fonte: Openlab (2025)

Evento Hackaton

O hackathon foi organizado pela UNESCO, OpenLab, Hacks/Hackers e pelo Observatório de Comunicação (OdeCom), que uniram esforços para promover a criação de ferramentas inovadoras que contribuam para combater a desinformação. Este espaço reuniu, entre os dias 22 e 23 de fevereiro de 2025, programadores, designers, comunicadores e especialistas em jornalismo, que trabalharam na construção de soluções tecnológicas para verificar informações de forma automatizada e eficiente. O autor deste trabalho fez parte do comitê organizador e do júri.

Figura 2: Participantes do Hackaton



Fonte: Rubén Zavala, Openlab (2025)

Figura 3: Participantes dp Hackaton



Fonte: Rubén Zavala, Openlab (2025)

Figura 4: Participantes do Hackaton



Fuente: Rubén Zavala, Openlab (2025)

Os grupos apresentaram diversas propostas que abordaram diferentes dimensões do problema da desinformação. A seguir, apresenta-se uma descrição detalhada dos projetos destacados, ressaltando suas características técnicas, pontos fortes e áreas de melhoria.

Croak Cierto: Este projeto propôs a criação de um aplicativo móvel complementado por um chatbot no WhatsApp. A ferramenta se destacou por sua capacidade de verificação imediata, fornecendo aos usuários alertas em tempo real sobre desinformação que circula nas redes sociais. O sistema, alimentado por algoritmos de inteligência artificial e aprendizado de máquina, permitia que os usuários enviassem links ou imagens para que o chatbot analisasse o conteúdo e fornecesse um parecer de veracidade. Entre seus principais méritos destacam-se seu design intuitivo e a capacidade de gerar relatórios detalhados sobre tendências de desinformação. No entanto, um aspecto que exigirá mais desenvolvimento é a incorporação de critérios transparentes na seleção de fontes e a melhoria do sistema de alertas para evitar falsos positivos.

Figura 5: Ferramenta Croak Cierto



Fonte: Openlab (2025)

Plataforma RunaCheck: Esta proposta destacou-se por seu foco na alfabetização midiática. O RunaCheck operava por meio de um chatbot que permitia verificar informações utilizando texto, imagens ou links. Seu sistema de categorização incluía etiquetas como "Verdadeiro", "Provavelmente Verdadeiro", "Não Verificável" e "Falso". Além disso, seu modelo incentivava a participação ativa do público no processo de verificação, promovendo a criação de uma comunidade de usuários comprometidos com a identificação de notícias falsas. Seu sistema de incentivos foi um aspecto particularmente inovador, evidenciando um esforço para fomentar a participação cidadã.

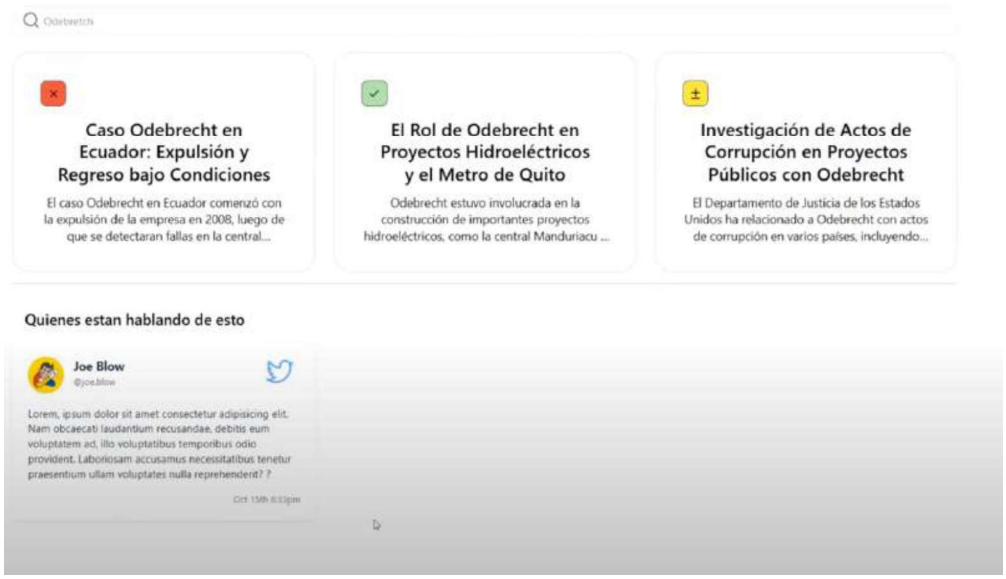
Figura 6: Ferramenta Runa Check



Fonte: Openlab (2025)

VerificIA: Esta ferramenta destacou-se pelo uso avançado de algoritmos de Reinforcement Learning, que permitiram monitorar discursos políticos e verificar afirmações públicas em tempo real. Seu foco na análise contextual foi um ponto-chave, pois possibilitou identificar padrões discursivos em redes sociais e comparar essas informações com fontes verificadas. A capacidade do VerificIA de detectar manipulação informativa foi especialmente valiosa em ambientes de alta polarização política.

Figura 7: Ferramenta VerificIA

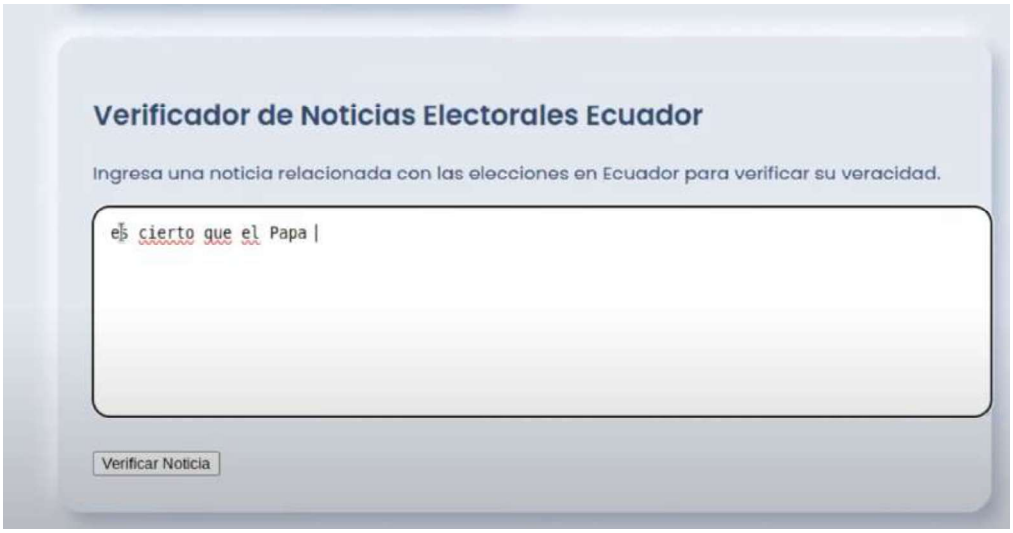


Fonte: Openlab (2025)

Veritas AI: O Veritas AI apresentou uma abordagem técnica sólida, combinando técnicas avançadas de web scraping com um banco de dados que avaliava a confiabilidade de veículos informativos. Seu modelo incluiu análise de sentimento e detecção de polarização nos textos, facilitando assim a identificação de padrões discursivos tendenciosos.

O desenvolvimento de um painel interativo foi um de seus principais diferenciais, permitindo apresentar tendências informativas em tempo real. Essa abordagem visual foi essencial para facilitar a compreensão da dinâmica da desinformação, especialmente para públicos não especializados em análise de dados.

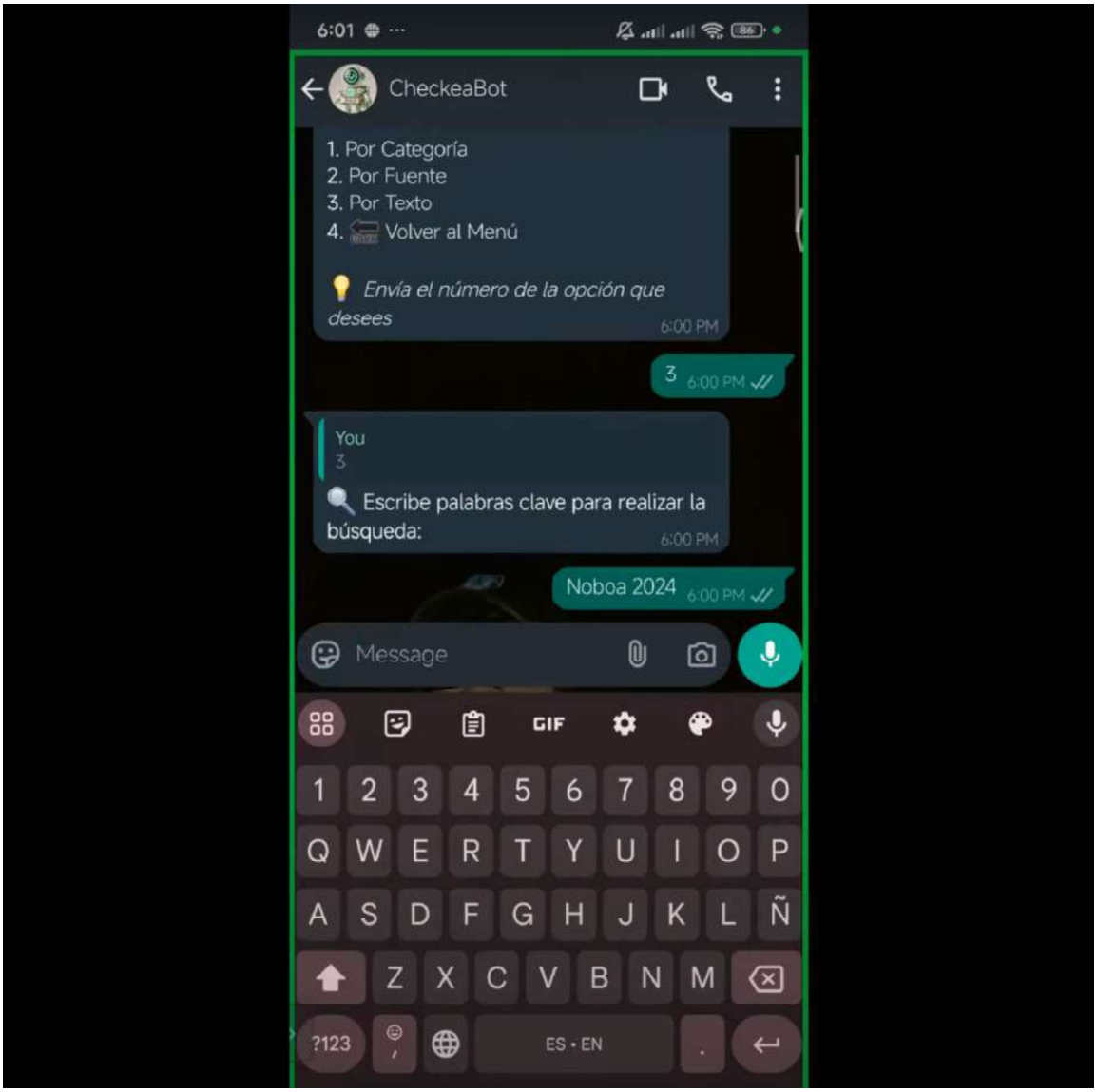
Figura 8: Ferramenta Veritas AI



Fonte: Openlab (2025)

ChequeaBot Este projeto destacou-se por seu caráter inclusivo, já que foi projetado especificamente para alcançar comunidades indígenas e idosos. Programado para interagir em línguas locais como quíchua e shuar, o ChequeaBot contribuiu para reduzir as barreiras digitais e facilitar o acesso à verificação de informações em populações vulneráveis. Seu design simples e a implementação de uma interface via WhatsApp permitiram que esta ferramenta se adaptasse facilmente a usuários com pouca experiência tecnológica.

Figura 9: Ferramenta Chequea Bot

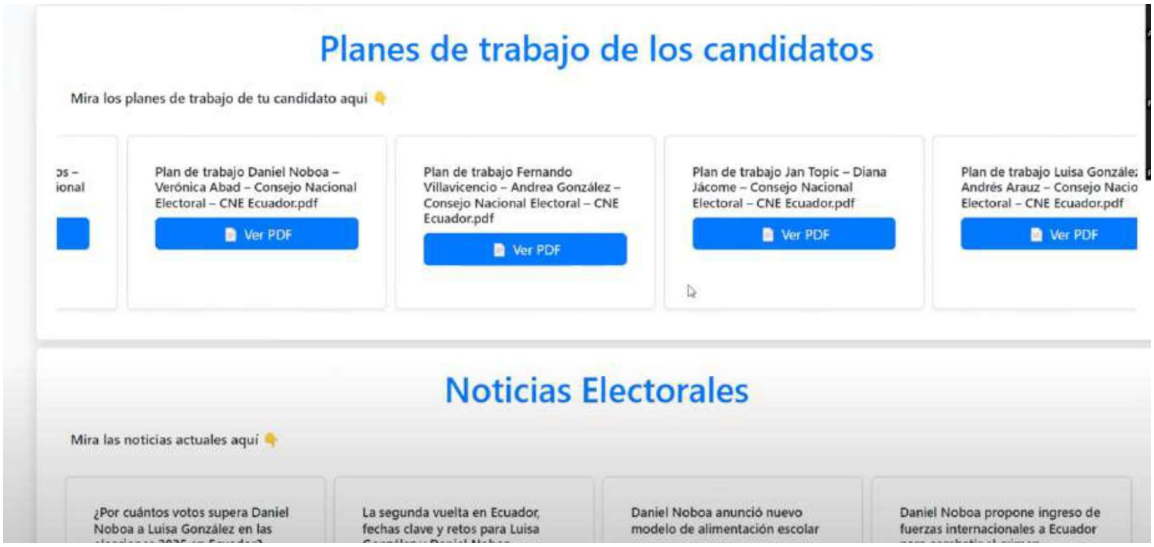


Fonte: Openlab (2025)

Ecuador Abierto: Com uma abordagem claramente política, este projeto concentrou-se na verificação de informações eleitorais durante processos democráticos. Utilizando técnicas avançadas de processamento de linguagem natural (NLP), o Ecuador Abierto comparou declarações públicas, entrevistas e debates com bases de dados verificadas. Seu sistema de alertas em tempo real foi uma inovação importante que demonstrou potencial para reduzir a

disseminação de campanhas de manipulação política. No entanto, um dos desafios pendentes é garantir a diversidade de fontes, evitando assim o viés na informação verificada..

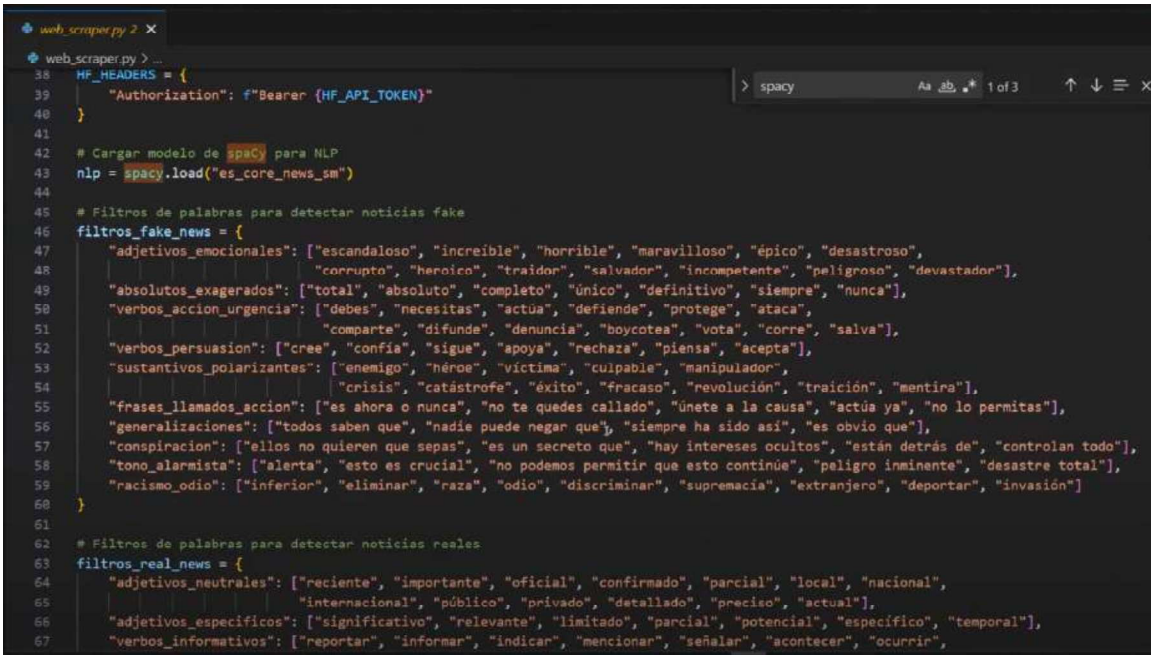
Figura 10: Ferramenta Ecuador Abierto



Fonte: Openlab (2025)

Godart Este projeto utilizou a metáfora do "cão de guarda" como símbolo de vigilância informativa. Seu principal foco foi a análise do impacto emocional nas notícias, utilizando algoritmos de análise de sentimento para identificar conteúdos projetados para gerar medo, raiva ou manipulação emocional. O projeto Godart destacou-se por sua abordagem na prevenção da viralização de notícias falsas, o que representa uma valiosa contribuição para a alfabetização emocional no ambiente digital.

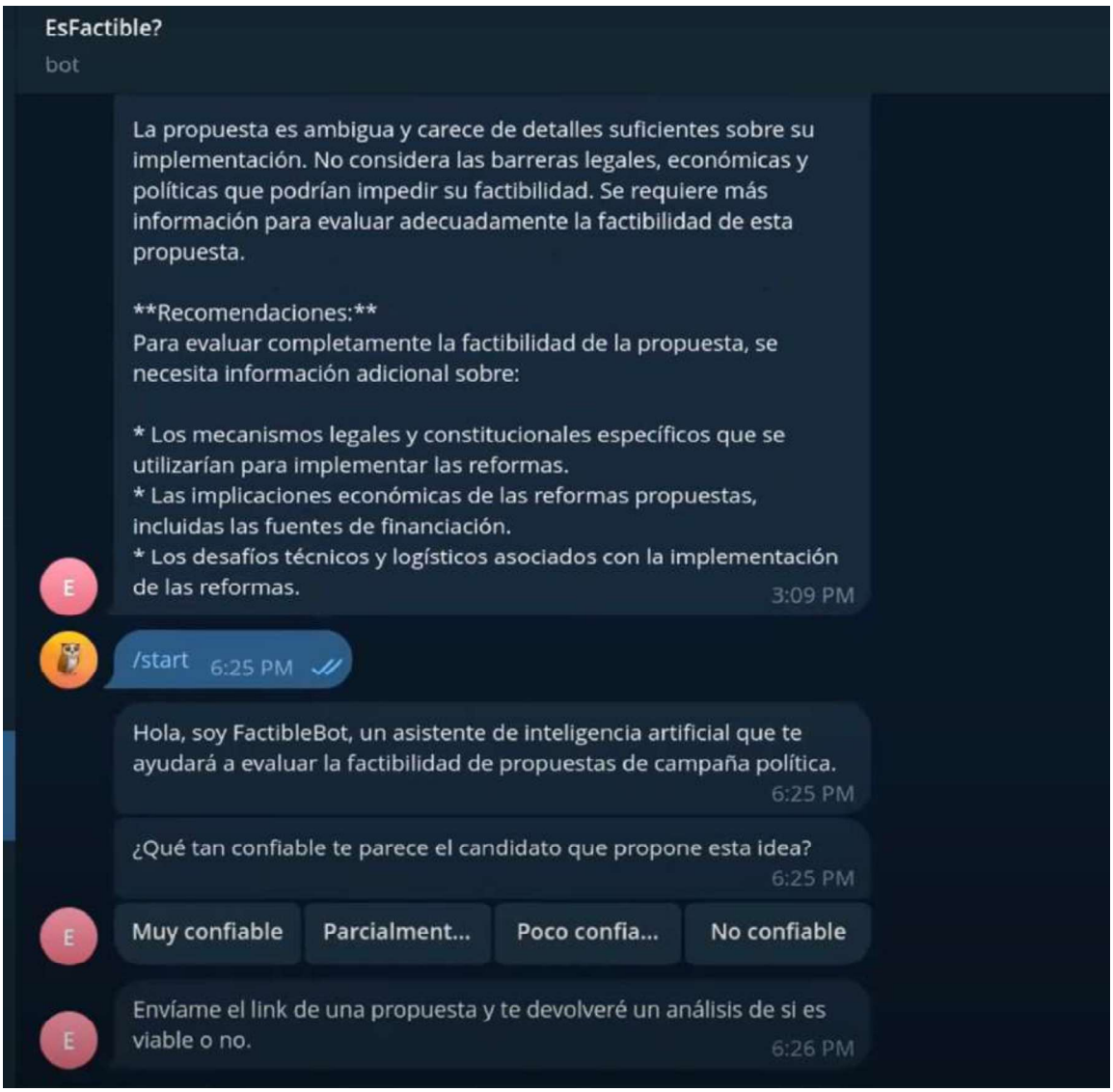
Figura 11: Ferramenta Goddart



Fonte: Openlab (2025)

Pillmind APP Este projeto apresentou um bot de Telegram projetado para avaliar a viabilidade de propostas eleitorais. A ferramenta analisava o contexto socioeconômico e político de cada proposta, emitindo um parecer que indicava se era "Viável", "Pouco Viável" ou "Inviável". Sua principal contribuição foi o uso de técnicas de fine-tuning para adaptar o modelo de IA às particularidades do contexto equatoriano. No entanto, um de seus principais desafios será otimizar a precisão de suas previsões e garantir a transparência de seus critérios de avaliação.

Figura 12: Ferramenta Pillmind APP



Fonte: Openlab (2025)

O hackathon evidenciou que o combate à desinformação requer uma abordagem integral que combine o desenvolvimento tecnológico com estratégias educativas que promovam a alfabetização midiática. A capacidade de integrar algoritmos éticos, escaláveis e eficientes consolidou-se como um dos principais desafios para o design de ferramentas automatizadas de fact-checking.

Os projetos mais destacados compartilharam uma série de características comuns que permitiram identificar padrões de sucesso:

• A integração de interfaces intuitivas que favorecem o acesso de públicos diversos.

• O design de sistemas que permitam a participação ativa da cidadania no processo de verificação.

• A incorporação de técnicas avançadas como a análise de sentimento, a detecção de polarização discursiva e o monitoramento em tempo real.

Resultados do Hackathon como Observação Não Participante

A análise das propostas apresentadas no hackathon revela uma série de padrões metodológicos e tendências tecnológicas que permitem interpretar o evento como um espaço experimental para o desenvolvimento de ferramentas de fact-checking. Sob essa perspectiva, o hackathon se torna um ambiente propício para identificar como a inteligência artificial (IA) se articula com as práticas jornalísticas contemporâneas e como as metodologias de verificação estão evoluindo diante do ecossistema digital.

Essa abordagem revela que o hackathon funcionou como um laboratório social no qual os participantes testaram diferentes hipóteses tecnológicas orientadas para enfrentar um dos maiores desafios do ecossistema informativo contemporâneo: a crescente manipulação informativa e o avanço das notícias falsas. As propostas apresentadas nesse espaço não se limitaram apenas ao desenvolvimento de soluções técnicas, mas também refletiram um esforço para articular metodologias de verificação mais inclusivas, transparentes e escaláveis.

Assim, o hackathon funcionou como um espaço onde convergiram abordagens jornalísticas, desenvolvimentos algorítmicos e metodologias de design centradas no usuário. Essa interação interdisciplinar contribuiu para que as ferramentas apresentadas se destacassem não

apenas por suas capacidades tecnológicas, mas também por sua pertinência social e comunicacional.

Categorias Emergentes no Hackathon

A análise dos projetos permite identificar três categorias fundamentais que explicam as tendências observadas no evento: a automação da verificação, a abordagem contextualizada para o público e a prevenção da manipulação informativa. Cada uma dessas categorias reflete uma forma particular de abordar os desafios que o ecossistema midiático contemporâneo e a luta contra a desinformação apresentam.

- **Automação da Verificação**

Projetos como a Ferramenta de Verificação em Tempo Real e o RunaCheck evidenciaram uma abordagem orientada para otimizar a rapidez do fact-checking por meio do uso de chatbots e processos automatizados de verificação. Esses projetos apresentaram um modelo que prioriza a imediatidade na verificação de dados, permitindo que os usuários obtenham informações verificadas em tempo real. Esse tipo de ferramenta revelou uma alta dependência de sistemas de processamento de linguagem natural (NLP) para contrastar dados e verificar a veracidade das informações.

A automação da verificação configurou-se como uma tendência emergente que responde à crescente velocidade com que a informação circula nos ambientes digitais. Ao reduzir a necessidade de intervenção humana direta no primeiro filtro de verificação, essas ferramentas otimizam o tempo de resposta das redações jornalísticas e permitem que os jornalistas se concentrem na análise crítica e contextual da informação. No entanto, o sucesso desses sistemas depende de sua capacidade de diferenciar padrões linguísticos enganosos e lidar adequadamente com o fenômeno da ambiguidade discursiva, especialmente em temas complexos ou polêmicos.

- **Abordagem Contextualizada para o Público**

Ferramentas como ChequeaBot e Ecuador Abierto destacaram-se pelos seus esforços em adaptar a verificação de dados a públicos específicos. Em vez de se concentrarem apenas na

precisão algorítmica, essas propostas incorporaram uma dimensão social que buscou integrar comunidades historicamente vulneráveis à manipulação informativa.

A integração de línguas indígenas, a atenção a populações vulneráveis e o design de sistemas acessíveis marcaram uma abordagem inclusiva que responde à diversidade do ambiente midiático.

O valor diferencial dessas propostas residiu no fato de que não apenas responderam à dinâmica tecnológica do fact-checking, mas também reconheceram a importância de adaptar as ferramentas de verificação às realidades socioculturais de seus usuários. Essa abordagem foi especialmente relevante no caso do ChequeaBot, que permitiu oferecer serviços de verificação de dados em línguas como o quíchua e o shuar, dois idiomas que costumam ficar fora do circuito digital tradicional.

Da mesma forma, o Ecuador Abierto demonstrou que a verificação contextualizada é uma ferramenta poderosa para combater a manipulação política durante períodos eleitorais.

- **Prevenção da Manipulação Informativa**

Projetos como VERA AI, Godart e Filming UPM propuseram mecanismos avançados para a detecção de discursos manipuladores. A incorporação da análise de sentimento, da previsão de impacto emocional e da detecção de padrões discursivos polarizados mostrou uma tendência ao fortalecimento do jornalismo preventivo diante da manipulação informativa.

Essas ferramentas basearam-se na premissa de que a desinformação não se limita apenas à propagação de falsidades, mas também se expressa por meio do uso estratégico de emoções para influenciar o comportamento do público.

Os modelos preditivos incorporados em ferramentas como o VERA AI permitiram antecipar como certas mensagens poderiam gerar climas sociais de tensão, medo ou polarização. Do mesmo modo, o Godart enfatizou a identificação de padrões discursivos projetados para manipular emocionalmente as audiências, o que representa uma dimensão inovadora para o fact-checking preventivo.

Observações Metodológicas

Sob uma perspectiva científica, o hackathon pode ser interpretado como um experimento social no qual se evidenciou uma notável correlação entre a eficácia das ferramentas e os seguintes fatores:

- **Acessibilidade do design:** Os projetos mais intuitivos tiveram melhor recepção por parte dos avaliadores, destacando que a tecnologia deve se concentrar em facilitar a experiência do usuário. A usabilidade foi um fator determinante para que as ferramentas fossem consideradas viáveis para sua implementação em redações ou em ambientes comunitários.
- **Capacidade de integração com fontes abertas:** Os sistemas que conseguiram vincular informações em tempo real a partir de fontes públicas verificáveis mostraram maior precisão na verificação de dados. Essa integração foi essencial para que os sistemas de *fact-checking* oferecessem resultados mais amplos, contextualizados e relevantes para o usuário final.
- **Diversidade na equipe de trabalho:** Os projetos que integraram equipes interdisciplinares mostraram maior solidez tanto na lógica algorítmica quanto na abordagem jornalística. A combinação de desenvolvedores, comunicadores e designers contribuiu para que as ferramentas não apenas fossem tecnicamente robustas, mas também compreensíveis e acessíveis para diversos públicos.

Abordagens Estratégicas para o Desenvolvimento de Ferramentas de Verificação

A análise dos projetos apresentados no hackathon, complementada pelo **Modelo Executivo de IA na Noticiabilidade**, permitiu identificar caminhos estratégicos para que a sociedade civil, a academia e os meios de comunicação colaborem de forma eficiente no design de ferramentas eficazes de *fact-checking*. A seguir, detalham-se as áreas-chave que devem orientar esses esforços colaborativos:

Atualidade: Detecção e verificação em tempo real

A verificação em tempo real é uma prioridade no contexto midiático contemporâneo, especialmente durante eventos de alto impacto social, como crises políticas, catástrofes naturais ou eleições. Nesse sentido, recomenda-se que:

- **Meios de comunicação:** Desenvolvam sistemas automatizados que permitam detectar e verificar informações em tempo real, priorizando a implementação de algoritmos de Processamento de Linguagem Natural (NLP) capazes de identificar padrões discursivos enganosos.
- **Academia**: Estabeleçam projetos de pesquisa focados na melhoria de sistemas de NLP para lidar com a ambiguidade discursiva, o uso de metáforas e as variações linguísticas que costumam dificultar a verificação automatizada.
- **Sociedade civil**: Forneça dados abertos e bases de informações verificáveis que permitam fortalecer o treinamento desses sistemas e aumentar sua precisão em tempo real.

Relevância: Personalização de conteúdo

A personalização da informação representa uma ferramenta essencial para responder aos interesses e necessidades das audiências. Para otimizar esse aspecto, sugere-se que:

- **Meios de comunicação:** Desenvolvam plataformas que ofereçam recomendações informativas personalizadas, garantindo que essas recomendações estejam livres de vieses algorítmicos que reforcem bolhas informativas.
- **Academia**: Desenvolvam estudos sobre padrões de consumo informativo, identificando mecanismos que permitam mitigar os efeitos do filtro personalizado.
- **Sociedade civil**: Promova espaços educativos que fortaleçam a capacidade crítica das audiências para interpretar a informação personalizada.

Impacto: Identificação de eventos com grande alcance

O uso de algoritmos preditivos permite antecipar tendências informativas e prevenir a manipulação em massa. Nessa linha, sugere-se que:

- **Meios de comunicação:** Incorporem sistemas de monitoramento de redes sociais capazes de detectar sinais precoces de desinformação viral.
- **Academia**: Fortaleçam os modelos preditivos que permitem identificar eventos que possam ganhar destaque na agenda midiática.
- **Sociedade civil**: Participe ativamente como observadora de dinâmicas sociais que indiquem a propagação coordenada de informações manipuladoras.

Novidade: Detecção de temas emergentes

A rápida evolução do discurso digital exige ferramentas que permitam identificar novos temas em circulação. Para abordar essa necessidade, propõe-se que:

- **Meios de comunicação:** Desenvolvam sistemas que reconheçam mudanças nas narrativas informativas, permitindo identificar temas emergentes antes que se tornem virais.
- **Academia**: Impulsione pesquisas sobre padrões discursivos em transformação nos ambientes digitais, facilitando a criação de modelos que identifiquem variações discursivas relevantes.
- **Sociedade civil:** Fortaleça redes de colaboração comunitária que deem visibilidade a temas de interesse social que possam ficar invisibilizados na agenda midiática dominante.

Proeminência: Análise de figuras influentes

O monitoramento do impacto discursivo de figuras públicas é essencial para prevenir a manipulação informativa. Para isso, recomenda-se que:

- **Meios de comunicação:** Desenvolvam sistemas que rastreiem as declarações de atores políticos, sociais e midiáticos para prevenir a viralização de discursos manipuladores.
- **Academia**: Desenvolva modelos de análise discursiva que avaliem o impacto social e político dessas figuras.

- **Sociedade civil**: Promova o debate público sobre o papel que os líderes de opinião têm na disseminação de informações falsas ou manipuladoras.

PARTICIPAÇÃO NO INCOIN

Durante o ano de 2024, participei como palestrante em duas edições do Workshop INCOIN, um evento acadêmico centrado na aplicação da inteligência artificial em diversos campos do conhecimento. Em ambas as apresentações, abordou-se o impacto da IA no campo da transcrição de entrevistas, destacando seu valor como ferramenta para otimizar a análise qualitativa nas ciências sociais e no jornalismo.

A primeira apresentação ocorreu em 23 de agosto de 2024, no Workshop INCOIN: Inteligência Artificial no Ensino Superior, organizado pela PUCE. Nesse espaço, apresentei o trabalho intitulado "Transcrição e Inteligência Artificial", no qual expus os resultados de um estudo que avaliou o desempenho de quatro plataformas de IA para a transcrição automática de entrevistas em espanhol: Office 365 Transcribe, Amazon Transcribe, Notta e Whisper.

Essa pesquisa abordou aspectos como a eficiência (velocidade de transcrição) e a fidelidade textual, destacando que ferramentas como o Whisper demonstraram alta precisão em contextos com ruído ambiental, enquanto o Notta destacou-se pela rapidez, embora com limitações na precisão diante de sotaques regionais marcados. Enfatizou-se que, apesar dos avanços tecnológicos, a intervenção humana continua sendo crucial para garantir a qualidade do conteúdo transcrito.

Figura 13: Certificado INCOIN Ecuador



Fonte: Documentação pessoal

Posteriormente, nos dias 15 e 16 de novembro de 2024, essa mesma apresentação foi exposta no Workshop INCOIN: Inteligência Artificial em Multiciências, organizado pela Universidade César Vallejo, em Trujillo, Peru. Nesta segunda participação, reforçaram-se as conclusões do primeiro encontro, destacando a importância de que os pesquisadores selecionem cuidadosamente as ferramentas de transcrição de acordo com as necessidades específicas de seus projetos.

Figura 14: Certificado INCOIN Ecuador



Fuente: Documentação pessoal

O estudo apresentado em ambos os eventos teve como objetivo avaliar o desempenho de quatro plataformas de IA para a transcrição automática de entrevistas em espanhol: Office 365 Transcribe, Amazon Transcribe, Notta e Whisper. Para isso, utilizou-se um corpus de 450 entrevistas curtas, selecionadas para representar uma variedade de contextos linguísticos e sotaques. A análise avaliou dois critérios principais:

- Eficiência: A rapidez com que cada ferramenta completou a transcrição.
- Fidelidade textual: A precisão na captura dos enunciados, especialmente em situações de ruído de fundo ou com sotaques marcados.

Os resultados destacaram que o Whisper apresentou o melhor desempenho em termos de precisão, mesmo em ambientes ruidosos, embora sua implementação técnica tenha se mostrado mais complexa. Por outro lado, o Notta destacou-se pela velocidade, embora com limitações na detecção de sotaques regionais. As ferramentas Office 365 Transcribe e Amazon Transcribe ofereceram facilidade de uso e integração com sistemas na nuvem, mas apresentaram dificuldades para identificar corretamente as mudanças de locutor em entrevistas em grupo.

O estudo concluiu que a IA representa um avanço significativo para agilizar a transcrição de entrevistas, mas que a revisão humana continua sendo indispensável para corrigir erros e garantir a precisão do conteúdo. Além disso, enfatizou-se que a escolha da ferramenta adequada deve considerar fatores como o tipo de sotaque, a qualidade do áudio e o propósito da análise qualitativa.

As conclusões desse estudo guardam estreita relação com os desafios mais amplos enfrentados pelo jornalismo contemporâneo no que diz respeito à desinformação e à manipulação digital. Assim como a IA facilitou a automatização do processo de transcrição, ela também tem sido integrada à geração e distribuição de conteúdos informativos, o que apresenta riscos significativos na verificação de dados e na detecção de informações falsas.

A implementação de ferramentas de IA em redações jornalísticas, plataformas digitais e redes sociais pode acelerar a propagação de informações manipuladas, especialmente em períodos eleitorais ou durante eventos de grande impacto midiático.

Por esse motivo, a análise da IA como ferramenta para a transcrição automatizada oferece uma valiosa perspectiva que pode ser aplicada ao campo da noticiabilidade e do fact-checking. A necessidade de combinar a automação com a revisão humana ressalta a importância da colaboração interdisciplinar entre profissionais do jornalismo, da academia e da sociedade civil. Essa convergência é essencial para garantir que as tecnologias emergentes contribuam para fortalecer a precisão informativa, evitando que a automação se torne um fator que amplifique a manipulação informativa.

PROPOSTA: PLATAFORMA DE FACT-CHECKING DA GAZETA DO REGISTRO OFICIAL DO EQUADOR

A digitalização da informação governamental tornou-se um eixo fundamental para a transparência e o acesso cidadão aos dados públicos. Nesse contexto, o desenvolvimento de uma ferramenta tecnológica que facilite a análise, pesquisa e consulta do conteúdo publicado na Gazeta do Registro Oficial do Equador representa uma contribuição significativa para pesquisadores, jornalistas, acadêmicos e cidadãos interessados no marco legal do país.

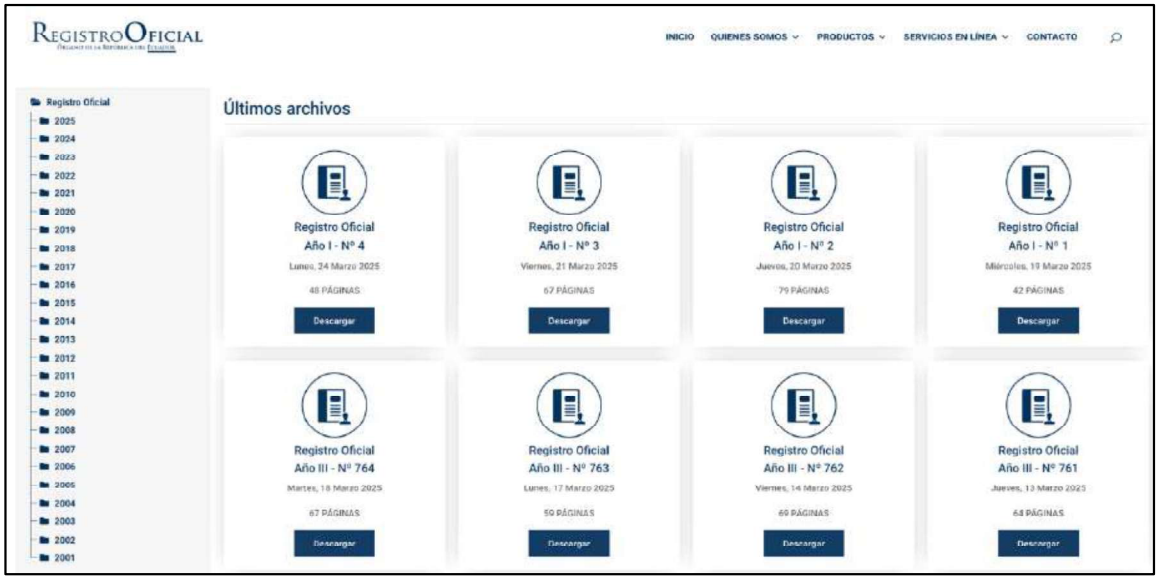
Este projeto busca desenvolver um sistema que integre técnicas avançadas de web scraping, a construção de uma base de dados estruturada (pool), o treinamento de um modelo baseado na API do ChatGPT, e o design de uma plataforma web ou aplicativo móvel que facilite a interação eficiente com a informação extraída.

O Registro Oficial do Equador contém informações legais fundamentais, como decretos executivos, leis, regulamentos, resoluções e acordos que impactam diretamente a vida política, econômica e social do país. No entanto, o acesso a essas informações costuma ser complexo devido à natureza extensa, dispersa e pouco estruturada dos dados publicados.

Mais do que uma ferramenta para consultar normas legais, esta proposta busca permitir que os cidadãos explorem a evolução histórica do marco normativo, destacando quais elementos foram incorporados, modificados ou eliminados nas leis vigentes.

A plataforma facilitará a identificação de lacunas ou mudanças relevantes, proporcionando assim uma ferramenta útil tanto para a sociedade civil quanto para profissionais do direito, pesquisadores e jornalistas.

Figura 15: Plataforma do Registro Oficial



Fonte: Corte Constitucional do Equador, 2025

Componentes do Projeto

1. Web Scraping: A primeira fase do projeto consiste no desenvolvimento de um processo automatizado de extração de dados por meio de web scraping. Essa técnica permitirá coletar de forma sistemática o conteúdo da Gazeta do Registro Oficial, identificando elementos-chave como datas de publicação, temáticas, atores envolvidos e o texto completo de cada documento.

O uso de ferramentas de *scraping* garantirá a coleta eficiente de informações e facilitará sua posterior análise. Esse desenvolvimento é realizado como parte do trabalho do OdeCom e de seu pesquisador associado, o Dr. Denis Chávez Ordóñez, cujo trabalho orienta a metodologia desta iniciativa (Chávez e Chávez, 2024).

Figura 16: Código proposto para iniciar o Web Scraping

```python
import requests
from bs4 import BeautifulSoup
import pandas as pd
import time

# URL base del Registro Oficial del Ecuador
BASE_URL = 'https://www.registroficial.gob.ec/'

# Función para obtener el contenido de una página
def obtener_contenido(url):
    headers = {'User-Agent': 'Mozilla/5.0'}  # Para evitar bloqueos
    try:
        respuesta = requests.get(url, headers=headers)
        respuesta.raise_for_status()
        return BeautifulSoup(respuesta.content, 'html.parser')
    except requests.exceptions.RequestException as e:
        print(f"Error al acceder a {url}: {e}")
        return None

# Función para extraer enlaces de ediciones recientes
def obtener_enlaces_ediciones():
    sopa = obtener_contenido(BASE_URL)
    if not sopa:
        return []

    enlaces = []
    for item in sopa.find_all('a', class_='link_edicion'):
        enlace = item.get('href')
        if enlace and enlace.startswith('https://'):
            enlaces.append(enlace)
    return enlaces

# Función para extraer información de una edición específica
def extraer_informacion_edicion(url):
    sopa = obtener_contenido(url)
    if not sopa:
        return None

    data = []
    for articulo in sopa.find_all('div', class_='contenedor-articulo'):
        titulo = articulo.find('h2').get_text(strip=True)
        fecha = articulo.find('span', class_='fecha-publicacion').get_text(strip=True)
        contenido = articulo.find('div', class_='contenido-articulo').get_text(strip=True)
        data.append({
            'Título': titulo,
            'Fecha': fecha,
            'Contenido': contenido,
            'URL': url
        })
    return data

# Proceso principal de scraping
def main():
    enlaces_ediciones = obtener_enlaces_ediciones()
    todas_las_ediciones = []

    for enlace in enlaces_ediciones:
        print(f"Extrayendo datos de: {enlace}")
        datos_edicion = extraer_informacion_edicion(enlace)
        if datos_edicion:
            todas_las_ediciones.extend(datos_edicion)
        time.sleep(2)  # Respetar los tiempos de respuesta del servidor

    # Guardar la información en un archivo CSV
    df = pd.DataFrame(todas_las_ediciones)
    df.to_csv('registro_oficial_ecuador.csv', index=False, encoding='utf-8-sig')
    print("Proceso finalizado. Los datos se guardaron en 'registro_oficial_ecuador.csv'")

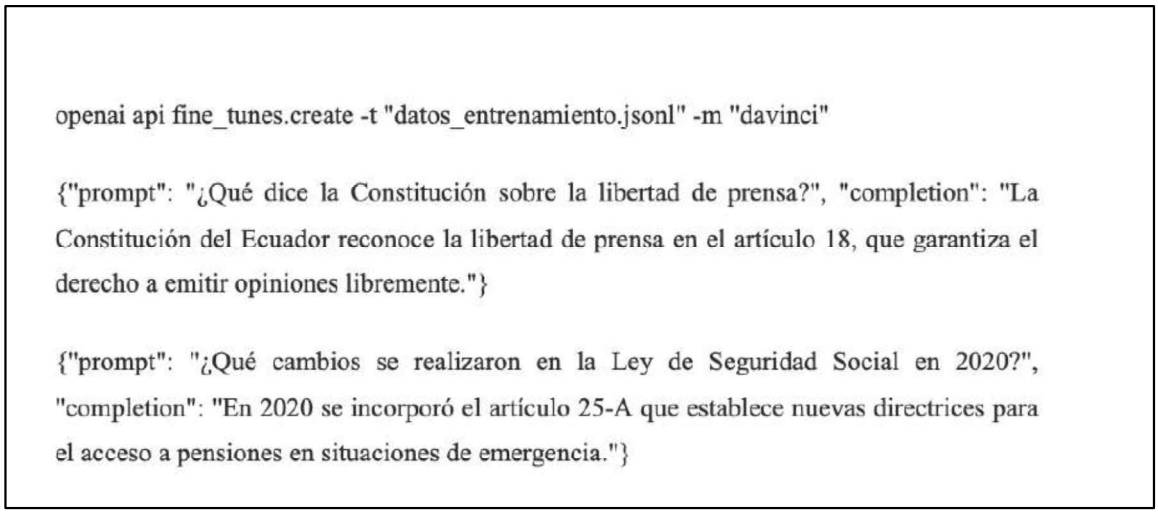
if __name__ == "__main__":
    main()
```

Fonte: Elaboração própria por Chávez e Cruz-Silva

2. Construção de Base de Dados (Pool): Após a extração das informações, será implementada uma **base de dados relacional** que permitirá armazenar, organizar e consultar os textos legais de forma eficiente. Esse *pool* de dados será projetado para facilitar a busca por palavras-chave, temáticas, autores, datas ou até mesmo por padrões discursivos que reflitam tendências normativas.

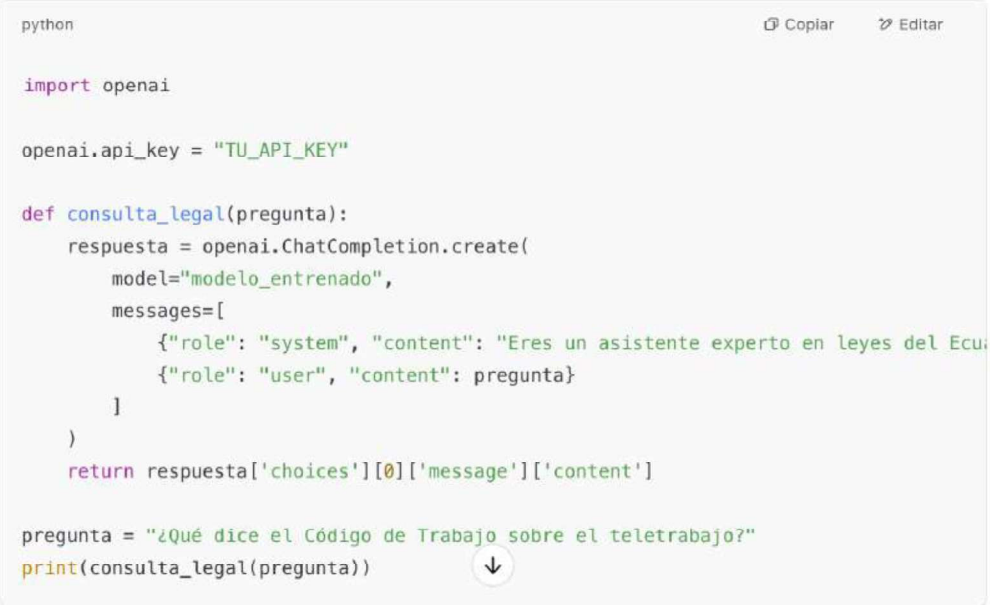
3. Treinamento da API do ChatGPT: C om a base de dados estruturada, proceder-se-á ao treinamento do modelo da API do ChatGPT para que ele seja capaz de responder a consultas específicas relacionadas ao conteúdo da Gazeta. Esse sistema será treinado utilizando prompts especializados que permitam reconhecer terminologia jurídica e fornecer respostas claras, contextualizadas e confiáveis. A adoção desse enfoque responde à necessidade de criar ferramentas que integrem a automação com a compreensão da linguagem natural sobre IA aplicada à verificação de fatos.

Figura 17: Exemplo de treinamento da API do ChatGPT



Fonte: Elaboração própria por Chávez e Cruz-Silva

Figura 17: E Exemplo de integração em Python



Fonte: Elaboração própria por Chávez e Cruz-Silva

4. Design da Plataforma: O desenvolvimento de uma plataforma web ou de um aplicativo móvel facilitará o acesso do público geral e especializado à informação do Registro Oficial. Essa ferramenta contará com uma interface intuitiva que permita realizar buscas avançadas, receber alertas sobre novas publicações relevantes e gerar resumos automatizados por meio do modelo treinado. Além disso, buscar-se-á que essa plataforma seja acessível tanto para especialistas em direito quanto para cidadãos sem experiência prévia na análise jurídica.

Relação com o Marco Teórico e a Inteligência Artificial

O design e a implementação desta proposta baseiam-se em conceitos-chave desenvolvidos no marco teórico sobre credibilidade digital, noticiabilidade e verificação de fatos em ambientes digitais.

A construção da credibilidade digital é sustentada pela combinação de três fatores principais:

1. Avaliação das fontes
2. Estruturação da mensagem

3. Integração de ferramentas tecnológicas para detectar elementos desinformativos

O projeto aqui apresentado responde a essas premissas ao implementar um sistema automatizado que garante a extração precisa de dados legais a partir de fontes oficiais, assegurando assim a confiabilidade da informação processada.

Além disso, a proposta alinha-se com os princípios de noticiabilidade descritos por Shoemaker e Reese (2014), que destacam a importância de fatores como atualidade, relevância, impacto, novidade e proeminência na identificação de informações significativas. A plataforma proposta se constrói com base nesses elementos ao oferecer aos usuários a possibilidade de acessar normativas recentes, localizar conteúdo relevante a temas de interesse social e receber alertas em tempo real sobre publicações importantes.

O projeto também incorpora práticas avançadas de fact-checking automatizado, elemento essencial na luta contra a desinformação. Pesquisas como as de Potthast et al. (2018) enfatizam que a integração de sistemas baseados em IA para a verificação de fatos é essencial para enfrentar a crescente complexidade da manipulação informativa. Nesse sentido, o modelo do ChatGPT treinado facilitará a identificação de contradições discursivas, a validação de afirmações e a geração de resumos analíticos que simplificam a compreensão de textos jurídicos complexos.

Vantagens do Projeto

- **Acesso otimizado à informação legal:** A ferramenta facilitará a busca eficiente por informações jurídicas relevantes, reduzindo o tempo e melhorando a precisão dos resultados.
- **Automatização da análise legal:** A integração do modelo de IA permitirá a geração de resumos, interpretação de termos complexos e análise comparativa entre leis e regulamentos.
- **Facilidade de uso:** A plataforma será projetada para que pessoas com diferentes níveis de experiência possam acessar e interpretar as informações publicadas no Registro Oficial.

Desafios e Dificuldades

O desenvolvimento do projeto envolve diversos desafios técnicos e metodológicos, entre os quais se destacam:

- **Precisão do Web Scraping:** A variabilidade no formato de publicação do Registro Oficial pode dificultar a extração automática de dados limpos e estruturados.
- **Treinamento da API do ChatGPT:** A complexidade da linguagem jurídica exige um treinamento minucioso do modelo para que suas respostas sejam precisas e contextualizadas.
- **Design da interface:** A plataforma deve garantir uma navegação intuitiva que permita o acesso a dados jurídicos complexos de forma clara e simples.

CONCLUSÕES FINAIS

A integração da inteligência artificial (IA) no âmbito da credibilidade digital e do fact-checking apresenta tanto oportunidades significativas quanto desafios importantes. Como destacado neste estudo, as ferramentas baseadas em IA oferecem vantagens substanciais para automatizar o processo de verificação, permitir a detecção rápida de conteúdo enganoso e melhorar a eficiência na análise de dados. Por meio do uso de algoritmos de processamento de linguagem natural (NLP) e de modelos de aprendizado de máquina, essas ferramentas permitem identificar padrões discursivos enganosos, rastrear a propagação de informações e analisar tendências comunicativas em tempo real (Potthast et al., 2018; Lahoti & Prasad, 2022).

Uma das principais contribuições desta pesquisa é o sistema proposto que integra técnicas de web scraping e a API do ChatGPT para facilitar o acesso à informação jurídica publicada na Gazeta do Registro Oficial do Equador. Este projeto aborda uma problemática chave no acesso à informação governamental, ao simplificar a busca por atualizações legais e permitir que cidadãos, pesquisadores e jornalistas acessem de forma eficiente dados cruciais. Ao combinar processos automatizados com a construção de uma base de dados estruturada, esta

proposta representa uma contribuição valiosa para melhorar a transparência e fortalecer a confiança pública nos marcos legais.

O marco teórico apresentado neste estudo reforça a ideia de que a credibilidade digital é um fenômeno complexo que envolve múltiplas dimensões. Autores como Eysenbach (2007) e Li et al. (2023) destacam que a credibilidade se constrói não apenas a partir da fonte da informação, mas também por meio da estrutura da mensagem e do contexto em que ela é difundida. Isso se alinha à proposta desenvolvida neste projeto, onde a implementação da IA não só melhora o acesso a dados jurídicos, mas também introduz critérios de veracidade baseados na rastreabilidade do conteúdo e na identificação de possíveis manipulações informativas.

O evento Hackathon "MediaHack: Tecnologias Cidadãs para Decisões Informadas nas Eleições Equador 2025", no qual foram exploradas diversas propostas tecnológicas para combater a desinformação, trouxe descobertas relevantes para esta pesquisa. Nesse evento foram apresentados projetos que integraram IA em diferentes áreas, destacando ferramentas que facilitaram a verificação automatizada em redes sociais, a análise de discursos políticos e a detecção de conteúdo manipulado por meio de técnicas avançadas, como a análise de sentimento e o monitoramento de padrões discursivos polarizados. Esse tipo de experiência evidencia que a IA pode desempenhar um papel essencial na proteção do ecossistema informativo e na luta contra a manipulação digital.

No contexto do hackathon, ferramentas como o "VerificIA" destacaram-se pelo uso de algoritmos de Reinforcement Learning para monitorar discursos políticos e verificar afirmações públicas em tempo real. Isso demonstra que a IA pode ser aplicada com sucesso para identificar padrões discursivos que indiquem manipulação informativa. Por outro lado, projetos como o "ChequeaBot", desenvolvido para comunidades indígenas e idosos, evidenciaram que a IA não deve se limitar apenas à detecção de conteúdo enganoso, mas também deve garantir a acessibilidade e a inclusão digital.

A IA, no entanto, não está isenta de limitações. Embora essas ferramentas sejam eficientes para identificar padrões enganadores e verificar dados em tempo real, elas apresentam dificuldades para compreender o contexto cultural e político de determinadas mensagens.

Essa limitação ressalta a necessidade de que o fact-checking automatizado seja complementado com a intervenção humana, especialmente em situações onde a ambiguidade discursiva ou a manipulação emocional possam dificultar uma avaliação precisa. Como destacado em pesquisas anteriores (Eysenbach, 2007; Castellón, 2021), a colaboração interdisciplinar entre jornalistas, desenvolvedores e especialistas em IA é essencial para garantir a precisão e o impacto social dessas ferramentas.

Por outro lado, esta pesquisa destaca a necessidade de promover a alfabetização midiática como parte essencial da luta contra a desinformação. Iniciativas educativas que incentivem o pensamento crítico, a análise de fontes e a participação ativa na verificação de dados são fundamentais para fortalecer a resiliência social frente às narrativas manipuladoras. Essa abordagem está alinhada à proposta desenvolvida nesta pesquisa, onde a plataforma projetada não só facilita o acesso à informação jurídica, mas também permite que os usuários se envolvam ativamente no processo de análise informativa.

Nesse sentido, é importante destacar que a IA também tem demonstrado seu potencial em outros campos além do fact-checking. Ferramentas baseadas em IA estão sendo amplamente utilizadas na medicina para o diagnóstico precoce de doenças, na educação para personalizar o aprendizado e na gestão ambiental para o monitoramento de ecossistemas. Essa versatilidade evidencia que a IA, quando aplicada de forma ética e transparente, pode ser uma poderosa aliada na resolução de problemas sociais complexos.

Em conclusão, este estudo destaca que a combinação de sistemas automatizados de verificação, a intervenção humana e a promoção da educação midiática são fundamentais para construir um ecossistema informativo resiliente. Integrar a IA de forma responsável, assegurando critérios éticos e de transparência, contribuirá para mitigar os efeitos nocivos da desinformação e para fortalecer a confiança pública nos ambientes digitais.

A evolução dessas ferramentas deve estar acompanhada de um compromisso ativo por parte da sociedade civil, dos meios de comunicação e da academia, garantindo assim que a inteligência artificial se torne uma aliada para a democratização do conhecimento e a proteção da verdade informativa.

REFERENCIAS

Abbas, F., & Taeihagh, A. (2024). Unmasking deepfakes: A systematic review of deepfake detection and generation techniques using artificial intelligence. Expert Systems with Applications, 252. https://doi.org/10.1016/j.eswa.2024.124260

Abrar, M. F., Khan, M. S., Khan, I., ElAffendi, M., & Ahmad, S. (2023). Towards Fake News Detection: A Multivocal Literature Review of Credibility Factors in Online News Stories and Analysis Using Analytical Hierarchical Process. Electronics (Switzerland), 12(15). https://doi.org/10.3390/electronics12153280

Boonprakong, N., Tag, B., & Dingler, T. (2023). Designing Technologies to Support Critical Thinking in an Age of Misinformation. IEEE Pervasive Computing, 22(3), 8–17. https://doi.org/10.1109/MPRV.2023.3275514

Bromley, M. (2012). *From spotlight to echo chamber? Citizen journalism and international news*. En *International News in the Digital Age* (pp. 23-40). Routledge.

Castellón, L. B. (2021). Audiovisual Verification in the Evolution of Television Newsrooms: Al Jazeera and the Transition from Satellite to the Cloud. Analisi, 64, 85–102. https://doi.org/10.5565/REV/ANALISI.3414

Celiktutan, B., Cadario, R., & Morewedge, C. K. (2024). People see more of their biases in algorithms. Proceedings of the National Academy of Sciences of the United States of America, 121(16). https://doi.org/10.1073/pnas.2317602121

Chandrashekar, S., & Hockema, S. (2009). Online access, participation and information credibility assessment. TIC-STH'09: 2009 IEEE Toronto International Conference - Science and Technology for Humanity, 755–760. https://doi.org/10.1109/TIC-STH.2009.5444397

Chavez, D. I., & Chavez, H. (2024). Data Mining and Interaction Design Applied to the Research Work Activity. En Communication and Applied Technologies (Vol. 375,

pp. 93-103). Springer Nature. https://link.springer.com/chapter/10.1007/978-981-99-7210-4_9

Diakopoulos, N. (2016). *Accountability in algorithmic decision making*. *Communications of the ACM, 59*(2), 56-62.

Ergüney, M. (2023). The impact of disinformation on the electoral process in the digital age. In Management in the Digital Era: Different Perspectives.

Eysenbach, G. (2007). From intermediation to disintermediation and apomediation: new models for consumers to access and assess the credibility of health information in the age of Web2.0. Medinfo. MEDINFO, 12(Pt 1), 162–166.

Farooq, M., & Makhdomi, A. A. (2021). *Blockchain-Based Incentive Mechanism to Combat Fake News*. *Communications in Computer and Information Science, 1487*, 62-77.

Fletcher, R., Schifferes, S., & Thurman, N. (2020). *Building the 'Truthmeter': Training algorithms to help journalists assess the credibility of social media sources*. *Convergence, 26*(1), 19-34.

Gao, J., Concas, S., Orrù, G., Feng, X., Marcialis, G. L., & Roli, F. (2024). Generalized Deepfake Detection Algorithm Based on Inconsistency Between Inner and Outer Faces. In Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics): Vol. 14365 LNCS. https://doi.org/10.1007/978-3-031-51023-6_29

Garg, D., & Gill, R. (2023). Deepfake Generation and Detection - An Exploratory Study. 2023 10th IEEE Uttar Pradesh Section International Conference on Electrical, Electronics and Computer Engineering, UPCON 2023, 888–893. https://doi.org/10.1109/UPCON59197.2023.10434896

He, L., & He, C. (2022). Help Me #DebunkThis: Unpacking Individual and Community's Collaborative Work in Information Credibility Assessment. Proceedings of the ACM on Human-Computer Interaction, 6(CSCW2). https://doi.org/10.1145/3555138

Kanagavalli, N., & Priya, S. B. (2022). *Social Networks Fake Account and Fake News Identification with Reliable Deep Learning*. *Intelligent Automation & Soft Computing, 33*(1), 191-205.

Kaswan, K. S., Malik, K., Dhatterwal, J. S., Naruka, M. S., & Govardhan, D. (2023). Deepfakes: A Review on Technologies, Applications and Strategies. 2023 International Conference on Power Energy, Environment and Intelligent Control, PEEIC 2023, 292–297. https://doi.org/10.1109/PEEIC59336.2023.10450604

Knorr-Cetina, K. (2009). The Synthetic Situation: Interactionism for a Global World. Symbolic Interaction, 32(1), 61–87. https://doi.org/https://doi.org/10.1525/si.2009.32.1.61

Kovach, B., & Rosenstiel, T. (2014). The Elements of Journalism: What Newspeople Should Know and the Public Should Expect. New York: Three Rivers Press.

Lahoti, P., & Prasad, S. (2022). *Using machine learning and robotic process automation for misinformation detection on Twitter: Analysing the tweets on Covid-19 pandemic*. *AIP Conference Proceedings*. Coimbatore.

Li, J., Kuutila, M., Huusko, E., Kariyakarawana, N., Savic, M., Ahooie, N. N., Hosio, S., & Mäntylä, M. (2023). Assessing Credibility Factors of Short-Form Social Media Posts: A Crowdsourced Online Experiment. ACM International Conference Proceeding Series. https://doi.org/10.1145/3605390.3605406

Li, R., & Suh, A. (2015). Factors Influencing Information credibility on Social Media Platforms: Evidence from Facebook Pages. Procedia Computer Science, 72, 314–328. https://doi.org/10.1016/j.procs.2015.12.146

Mauk, M., & Grömping, M. (2024). Online Disinformation Predicts Inaccurate Beliefs About Election Fairness Among Both Winners and Losers. Comparative Political Studies, 57(6), 965–998. https://doi.org/10.1177/00104140231193008

Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., & Galstyan, A. (2021). *A survey on bias and fairness in machine learning*. *ACM Computing Surveys (CSUR), 54*(6), 1-35.

Mena, P. (2019). *Principles and Boundaries of Fact-checking: Journalists' Perceptions*. *Journalism Practice, 13*(6), 657-672.

McQuail, D. (2010). McQuail's Mass Communication Theory. London: Sage Publications.

Naitali, A., Ridouani, M., Salahdine, F., & Kaabouch, N. (2023). Deepfake Attacks: Generation, Detection, Datasets, Challenges, and Research Directions. Computers, 12(10). https://doi.org/10.3390/computers12100216

Park, C. S., & Kaye, B. K. (2023). *Applying news values theory to liking, commenting and sharing mainstream news articles on Facebook*. *Journalism, 24*(3), 633-653.

Pérez-Curiel, C., Domínguez-García, R., & Jiménez-Marín, G. (2021). Public Sphere and Misinformation in the U.S. Election: Trump's Audience and Populism Indicators in the COVID-19 Context. Journalism and Media, 2(3), 335–350. https://doi.org/10.3390/journalmedia2030020

Potthast, M., Kiesel, J., Reinartz, K., Bevendorff, J., & Stein, B. (2018). *A Stylometric Inquiry into Hyperpartisan and Fake News*. *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics*. Melbourne.

Rajput, T., & Arora, B. (2024). A Systematic Review of Deepfake Detection Using Learning Techniques and Vision Transformer. In Lecture Notes in Networks and Systems: Vol. 991 LNNS. https://doi.org/10.1007/978-981-97-2550-2_17

Schütz, M. (2023). *Disinformation Detection: Knowledge Infusion with Transfer Learning and Visualizations*. *ECIR 2023. Lecture Notes in Computer Science*. Cham.

Shin, D., & Shin, E. Y. (2023). Data's Impact on Algorithmic Bias. Computer, 56(6), 90–94. https://doi.org/10.1109/MC.2023.3262909

Shin, D., & Lee, J. (2022). *The Effect of Deepfake Video on News Credibility and Corrective Influence of Cost-Based Knowledge about Deepfakes*. *Digital Journalism, 10*(3), 412-432.

Shoemaker, P., & Reese, S. D. (2014). Mediating the message in the 21st century: A media sociology perspective. London: Routledge.

Shu, K., Sliva, A., Wang, S., Tang, J., & Liu, H. (2017). *Fake News Detection on Social Media: A Data Mining Perspective*. *ACM SIGKDD Explorations Newsletter, 19*(1), 22-36.

Society of Professional Journalists. (2014). SPJ Code of Ethics. Recuperado de https://www.spj.org/ethicscode.asp

Stalder, F. (2018). The Digital Condition (Polity Press). https://doi.org/10.1177/0267323118799184a

Sittmann, J., Tompkins, & Andrew. (2020). *DW Akademie*. Recuperado de https://p.dw.com/p/3eOfW.

Tagarev, A., Bozhanova, K., Nikolova-Koleva, I., & Ivanov, I. (2021). Tackling Multilinguality and Internationality in Fake News. In R. Mitkov & G. Angelova (Eds.), Proceedings of the International Conference on Recent Advances in Natural Language Processing (RANLP 2021) (pp. 1380–1386). INCOMA Ltd. https://aclanthology.org/2021.ranlp-1.154/

Trithara, D. (2020). Securitizing Disinformation: The Case of Westminster's Digital, Culture, Media and Sport Committee. Democracy and Security. https://doi.org/10.1080/17419166.2020.1740685

Wardle, C. (2017). *Information Disorder*. *Policy Commons*.

Yilma, K. M. (2021). On Disinformation, Elections and Ethiopian Law. Journal of African Law, 65(3), 351–375.