

I.G.R.O. Duarte¹

*¹Faculdade Tecnológica do Estado de São Paulo/Unidade São José dos Campos/Análise e Desenvolvimento de Sistemas.
ivan.duarte01@fatec.sp.gov.br*

COMO A INTELIGÊNCIA ARTIFICIAL PODE AJUDAR NO COMBATE A DESINFORMAÇÃO

I.G.R.O. Duarte¹

Caçapava
2023

¹*Faculdade Tecnológica do Estado de São Paulo/Unidade São José dos Campos/Análise e Desenvolvimento de Sistemas.*
ivan.duarte01@fatec.sp.gov.br

COMO A INTELIGÊNCIA ARTIFICIAL PODE AJUDAR NO COMBATE A DESINFORMAÇÃO

Trabalho apresentado à disciplina de Interação Humano Computador do Curso de Análise e Desenvolvimento de Sistemas da Faculdade de Tecnologia de São José dos Campos.

Orientador: Prof: Giuliano Araujo Bertoti

Professor Orientador: Giuliano Araujo Bertoti

Caçapava
2023

COMO A INTELIGÊNCIA ARTIFICIAL PODE AJUDAR NO COMBATE A DESINFORMAÇÃO

I.G.R.O. Duarte¹

¹Faculdade Tecnológica do Estado de São Paulo/Unidade São José dos Campos/Análise e Desenvolvimento de Sistemas.

ivan.duarte01@fatec.sp.gov.br

RESUMO

Este artigo aborda a importância e o potencial da inteligência artificial (IA) no combate à desinformação. A disseminação rápida e generalizada de informações falsas e enganosas tem se tornado um desafio crítico em nossas sociedades contemporâneas. Com base em uma revisão bibliográfica dos artigos citados, este estudo investiga o papel da IA no combate à desinformação e a utilização de suas técnicas para aumentar a segurança na navegação da internet.

Palavras-chave: Inteligência Artificial, Desinformação, Métodos de Checagem.

1. INTRODUÇÃO

A chamada “Era da Informação” é o termo utilizado para se referir ao período histórico contemporâneo, é caracterizada pelo rápido desenvolvimento do avanço tecnológico e da comunicação. Entende-se por comunicação o processo de troca de informações, que na nossa sociedade é fundamental para a interação humana e desempenha um papel crucial em todos os aspectos da vida, seja pessoal, profissional ou social.

Com a chegada da “Era da Informação” a internet trouxe consigo uma quebra de paradigma para as mídias tradicionais, transformando a maneira como consumimos, produzimos e compartilhamos conteúdo, bem como a forma como nos envolvemos com a informação e a comunicação em geral. Junto a ela foi anexado um novo desafio, com a democratização do acesso a informação e compartilhamento de dados, a internet também se tornou um ambiente fértil para o que chamamos de Fake News, também conhecido como notícias falsas, informações enganosas ou tiradas de contexto que são apresentadas como notícias reais. Essas notícias geralmente são divulgadas por meio de canais de mídia tradicionais ou plataformas online como as redes sociais. Elas podem assumir diferentes formas, como histórias fabricadas, imagens ou vídeos manipulados, manchetes enganosas ou relatos tendenciosos.

A intenção por trás das Fake News pode variar, incluindo a geração de lucro, a promoção de determinadas agendas ou a causação de desordem social ou política. tornaram uma preocupação significativa nos últimos anos devido à facilidade com que podem ser criadas e disseminadas através das redes sociais. A natureza viral desse conteúdo pode levar à crença generalizada em informações falsas, potencialmente influenciando a opinião pública e até mesmo afetando processos políticos. Após as eleições presidenciais nos Estados Unidos, as redes sociais frequentemente se tornaram um veículo treinado para disseminar desinformação e boatos.[3]

Fica evidente os problemas que um amplo acesso a informação pode causar, entretanto a busca pela abordagem mais apropriada para o tema Fake News e Controle de Informação ainda é bem polêmica, enquanto alguns grupos defendem uma regulamentação mais rígida da internet através de leis e agências de regulamentação estatais, outros consideram uma regulamentação estatal como uma oportunidade para o estado utilizar essa ferramenta para censurar opositores ideológicos e suprimir a liberdade de expressão.

A detecção de informações falsas, ainda é um desafio complexo, mas a inteligência artificial (IA) pode desempenhar um papel importante nesse processo utilizando técnicas como: Processamento de linguagem natural (NLP), análise de fontes e fatos, aprendizado de máquina e algoritmos de detecção de padrões, redes neurais e deep learning (Seção 2).

2. INTELIGÊNCIA ARTIFICIAL

A inteligência artificial (IA) refere-se à capacidade de um sistema ou máquina de imitar ou simular a inteligência humana. É um campo interdisciplinar que combina conhecimentos de ciência da computação, matemática, psicologia, neurociência e outras áreas para desenvolver algoritmos e sistemas capazes de aprender, raciocinar, reconhecer padrões, resolver problemas e tomar decisões autônomas. A IA tem sido amplamente aplicada em diversos setores como na medicina por exemplo, auxilia no diagnóstico médico e melhoria dos sistemas de saúde.

Existem várias abordagens e técnicas utilizadas na IA. O aprendizado de máquina (Machine Learning) treina sistemas em grandes conjuntos de dados para reconhecer padrões e tomar decisões. Redes neurais artificiais são modelos computacionais inspirados no cérebro humano, processando informações e aprendendo a partir dos dados. O processamento de linguagem natural – PLN (Seção 2.1), capacita sistemas a entender e interpretar a linguagem humana, usado em chatbots, assistentes virtuais e tradução automática. A visão computacional permite

que sistemas compreendam informações visuais, como imagens e vídeos, sendo aplicada em reconhecimento facial e detecção de objetos. Na robótica, a IA capacita robôs a agir e interagir autonomamente, aperfeiçoando movimentação, navegação e interação homem-máquina.

Outro exemplo de aplicação da inteligência artificial são os chatbots, aplicações que podem ser programadas com uma base de dados limitada, ou então treinados como um modelo de linguagem ampliando sua base de informações que fornecem soluções para diversos tipos de problemas. Um dos chatbots mais conhecidos atualmente é o ChatGPT, que é capaz de gerar respostas complexas em poucos segundos, como é explorado pelo autor Reid Hoffman em *“Impromptu – Amplifying Our Humanity Through AI”* [1], neste livro Reid Hoffman aborda temas do cotidiano e complexos debates éticos e morais, justamente para explorar o potencial da ferramenta conhecida com GPT-4.

Os algoritmos de inteligência artificial (IA) são usados em uma variedade de aplicações online, desde recomendação de produtos e publicidade direcionada até previsão de taxas de empréstimo e seguro. No entanto, à medida que a tomada de decisão baseada em IA afeta diretamente a vida das pessoas, a responsabilidade e a imparcialidade dos algoritmos avançados de IA estão sendo questionadas. Nos últimos anos, a necessidade de transparência algorítmica está recebendo mais atenção para possibilitar sistemas de tomada de decisão baseados em IA responsáveis. Para isso, técnicas de IA Explicável (XAI) foram introduzidas para adicionar transparência a algoritmos de aprendizado de máquina em caixa-preta facilitando sua interpretação. No domínio das redes sociais, os algoritmos de feed de notícias e busca funcionam de forma semelhante aos algoritmos de tomada de decisão, pois os usuários são expostos a conteúdo selecionado algorítmicamente. Confiar totalmente em notícias curadas algorítmicamente poderia levar potencialmente à propagação em larga escala de informações falsas e fabricadas [4]

2.1 Algoritmos PLN

Algoritmos de Processamento de Linguagem Natural (NLP) são métodos computacionais utilizados para a interpretação e manipulação de textos e linguagem humana por sistemas de inteligência artificial. Esses algoritmos permitem que computadores processem e compreendam a linguagem humana de maneira eficiente, permitindo uma ampla gama de aplicações práticas, como tradução automática, resumo de texto, chatbots e análise de sentimentos. A complexidade dos algoritmos de NLP reside em sua capacidade de lidar com a ambiguidade e a riqueza da linguagem natural. A linguagem humana é cheia de nuances, ambiguidades, sarcasmo, contexto cultural e outras características que tornam seu processamento uma tarefa desafiadora. Os algoritmos de NLP devem ser capazes de lidar com essas complexidades para obter resultados precisos e úteis. Uma das principais abordagens em algoritmos de NLP é a representação da linguagem por meio de modelos estatísticos e computacionais. Esses modelos são treinados em grandes quantidades de dados textuais, a fim de aprender padrões e relações entre palavras, frases e documentos. Um exemplo de modelo amplamente utilizado é o modelo de linguagem, que estima a probabilidade de uma sequência de palavras ocorrer em um texto específico.

Outra técnica importante em algoritmos de NLP é a extração de recursos linguísticos. Esses recursos incluem palavras-chave, entidades nomeadas, relações semânticas e outras informações relevantes presentes no texto. A extração desses recursos é

fundamental para várias tarefas de NLP, como classificação de texto, sumarização automática e análise de sentimentos. A compreensão do contexto é um aspecto crítico nos algoritmos de NLP. Muitas vezes, o significado de uma palavra ou frase depende do contexto em que é usada. Por exemplo, a palavra "banco" pode se referir a uma instituição financeira ou a uma peça de mobília, dependendo do contexto. Os algoritmos de NLP precisam capturar e interpretar esse contexto adequadamente para fornecer resultados precisos.

Uma técnica comum em algoritmos de NLP para lidar com o contexto é a análise sintática, que envolve a análise da estrutura gramatical de uma sentença. Isso permite identificar as relações entre palavras e a função que cada uma desempenha na frase.

Outra área importante em algoritmos de NLP é o processamento de sentimentos. Esses algoritmos são projetados para identificar e classificar a polaridade dos sentimentos expressos em textos, como positivo, negativo ou neutro. Isso é amplamente utilizado em análise de sentimentos de mídias sociais, avaliações de produtos e outras aplicações que requerem a compreensão do sentimento humano. Daniel Jurafsky e James H. Martin no livro *“Speech and Language Processing – Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition”*, dizem que: “O que diferencia as aplicações de processamento de linguagem de outros sistemas de processamento de dados é o uso do conhecimento da linguagem.”. Esta afirmação é fundamental para o combate a desinformação pois para que uma IA consiga distinguir os fatos, primeiro ela precisa ser capaz de analisar o que está escrito para depois comparar se aquelas informações são condizentes com a realidade dos fatos.

2.2 Redes Neurais e Deep Learning

As redes neurais e o deep learning baseiam-se em conceitos e algoritmos inspirados no funcionamento do cérebro humano, permitindo a criação de modelos computacionais altamente eficazes e versáteis.

Uma rede neural é uma estrutura composta por unidades interconectadas chamadas de neurônios artificiais. Cada neurônio recebe entradas ponderadas, realiza um cálculo e produz uma saída que é transmitida para outros neurônios na rede. Essas conexões são organizadas em camadas, sendo a primeira a camada de entrada, a última a camada de saída e as intermediárias conhecidas como camadas ocultas. O treinamento de uma rede neural é um processo crucial, em que os pesos das conexões entre os neurônios são ajustados com base em dados de treinamento. Esse ajuste é realizado através de algoritmos de aprendizado, utilizam técnicas de otimização para minimizar o erro entre as saídas da rede e os valores desejados. À medida que a rede é exposta a mais dados de treinamento, ela se torna mais capaz de generalizar e tomar decisões corretas em novos casos.

Já o Deep Learning é uma abordagem de aprendizado de máquina que utiliza redes neurais profundas, compostas por várias camadas ocultas. Essas redes têm a capacidade de aprender representações hierárquicas dos dados, permitindo a extração de características complexas em diferentes níveis de abstração. Dessa forma, o deep learning é capaz de lidar com problemas de alta dimensionalidade e encontrar padrões sutis em conjuntos de dados massivos.

Uma das principais vantagens das redes neurais e do deep learning é a capacidade de lidar com dados não estruturados, como imagens, áudio e texto. São amplamente utilizadas para realizar tarefas como reconhecimento de objetos, segmentação de imagens e detecção de padrões. Já em processamento de

linguagem natural (Seção 2.1), as redes neurais recorrentes e os modelos de transformers têm se destacado em tarefas como tradução automática, resumo de texto e geração de texto. No entanto, é importante destacar que o treinamento de redes neurais profundas requer uma grande quantidade de dados rotulados e um poder computacional significativo. Além disso, a interpretabilidade dos modelos de deep learning ainda é um desafio, pois eles tendem a ser caixas-pretas, dificultando a compreensão dos processos internos que levam a uma determinada decisão. A utilização das Redes Neurais e do Deep Learning associados aos Métodos de Checagem (Seção 3), podem ser uma alternativa para identificar de forma mais precisa informações falsas ou enganosas.

3. MÉTODOS DE CHECAGEM

O processo de checagem de notícias envolve um conjunto de técnicas e abordagens para verificar a veracidade e precisão das informações em diversos meios, como notícias, artigos e posts em mídias sociais. Essas técnicas variam em métodos específicos, como a verificação de fontes primárias e a pesquisa em bases de dados.

Na detecção de boatos, o objetivo é classificar uma informação como boato ou não. Um modelo de detecção envolve quatro etapas: detecção, rastreamento, postura e veracidade, que auxiliam na identificação dos boatos. As postagens desempenham um papel importante na determinação da autenticidade dos boatos. A detecção de boatos pode ser dividida em quatro subtarefas: classificação de postura, classificação de veracidade, rastreamento de boatos e classificação de boatos (Arkaitz et al., 2017). No entanto, ainda há pontos que requerem mais detalhes para entender completamente o problema e podemos aprender com os resultados se um boato é verdadeiro ou não e, se for um boato, até que ponto é verdadeiro.

A comparação de múltiplas fontes de informações de diferentes veículos e canais de notícias é outra técnica utilizada. Ao verificar se várias fontes confiáveis estão relatando um evento de maneira semelhante, aumenta-se a probabilidade de que a informação seja precisa.

A detecção de clickbait tem como objetivo atrair a atenção dos visitantes e incentivá-los a clicar em um link específico. As abordagens existentes para o clickbait utilizam características de extração de mensagens teaser, páginas da web vinculadas e metainformações de tweets. Portanto, em situações semelhantes, é possível notificar os leitores antes de ler qualquer notícia, indicando que ela pode ser falsa, para que os leitores estejam mais atentos.

Além disso, a verificação do contexto em que a informação é apresentada e a busca por evidências adicionais que confirmem ou refutem a afirmação são partes essenciais na checagem de fatos. A detecção de spam em e-mails também é um desafio significativo, causando prejuízos financeiros e incômodo aos usuários. Diferentes grupos estão trabalhando com abordagens de aprendizado de máquina para filtrar spam.

Para combater a disseminação de notícias falsas em redes sociais, são necessárias abordagens que considerem tanto o comportamento das entidades envolvidas quanto os aspectos sociais, combinando conhecimento e dados. É possível detectar notícias falsas com base em fatos conhecidos, como tempo, localização, qualidade e postura de outras pessoas. A medição dessas semelhanças pode ajudar a avaliar a qualidade das notícias.

4. CONCLUSÃO

Em conclusão, a utilização da inteligência artificial no combate à desinformação revela-se uma abordagem promissora e necessária. A capacidade da IA de identificar, rastrear e combater a disseminação de informações falsas e enganosas nas redes sociais e plataformas online é uma vantagem significativa no enfrentamento desse desafio atual. Além disso, os sistemas de IA podem desempenhar um papel crucial na verificação de fatos, analisando conteúdos em tempo real e auxiliando na identificação de informações confiáveis.

Outro aspecto importante é a capacidade da IA de educar e conscientizar o público sobre a desinformação. Por meio de chatbots e assistentes virtuais, as pessoas podem acessar informações confiáveis e verificadas instantaneamente, estimulando o desenvolvimento de um pensamento crítico e fornecendo recursos para combater a disseminação de notícias falsas. No entanto, é fundamental reconhecer que o uso da IA no combate à desinformação também apresenta desafios técnicos e éticos podendo surgir como uma possível alternativa aos conflitos entre grupos a favor e contra uma abordagem mais incisiva do estado no controle da internet. É necessário aprimorar constantemente os algoritmos de aprendizado de máquina para garantir a eficácia, precisão e imparcialidade na detecção da desinformação, ao mesmo tempo em que se protege a privacidade e a segurança dos usuários.

Dessa forma, a inteligência artificial oferece um arsenal de ferramentas e abordagens para enfrentar a desinformação, mas é importante lembrar que não se trata de uma solução única. A colaboração entre especialistas em IA, profissionais de comunicação e legisladores é essencial para desenvolver estratégias abrangentes e eficazes no combate à desinformação. Com esforços conjuntos, podemos aproveitar o potencial da IA para promover um ambiente de informação mais confiável e proteger a sociedade contra os perigos da desinformação.

5. REFERÊNCIAS

- [1] REID HOFFMAN. *Impromptu – Amplifying Our Humanity Through AI* – Dallepedia LLC, 2023.
- [2] DANIEL JURAFSKY and JAMES H. MARTIN. *Speech and Language Processing – Introduction to Natural Language Processing, Computation Linguistics, and Speech Recognition* – PEARSON Prentice Hall, 2008
- [3] Sajjad Ahmed, Knut Hinkelmann e Flavio Corradini, “Combining Machine Learning with Knowledge Engineering to detect Fake News in Social Networks-a survey” Arxiv, v1, 2022 (p 1-5).
- [4] Sina Mohseni, Fan Yang e Shiva Pentiyala, Mengnan Du, Yi Liu, Nic Lupfer, Xia Hu, Shuiwang Ji e Eric Ragan “Machine Learning Explanations to Prevent Overtrust in Fake News Detection” Arxiv, v2, 2020 (p 1-2).
- [5] Arkaitz Zubiaga, Ahmet Aker, Kalina Bontcheva, Maria Liakata, and Rob Procter. “Detection and resolution of rumours in social media: A survey”. arXiv, 2017.