

# Redes Neurais com Aprendizado Profundo para Detecção de Fake News

Caio Silas de Araujo Amaro<sup>1</sup>,  
Mauro Lúcio Afonso Paulino dos Santos Filho<sup>1</sup>

<sup>1</sup> Departamento de Computação  
Universidade Federal de Ouro Preto (UFOP) – Ouro Preto, MG – Brazil

mauro.paulino@aluno.ufop.edu.br, caio.amaro@aluno.ufop.edu.br

**Abstract.** *This article addresses the detection of fake news through two deep learning neural networks. We explore strategies using artificial intelligence in Python and evaluate performance on a Kaggle dataset. The first model achieved 97.57% accuracy, while the more complex second model exhibited overfitting. The results indicate opportunities for improvement and emphasize the need for effective strategies in the face of the constant evolution of misinformation dissemination tactics. Finally, it underscores the importance of interdisciplinary collaboration among experts in artificial intelligence, linguistics, and ethics to address persistent challenges in fake news detection. The pursuit of more efficient and adaptable approaches across different contexts remains a crucial area of research.*

**Resumo.** *Este artigo aborda a detecção de fake news por meio de duas redes neurais com aprendizado profundo. Exploramos estratégias utilizando inteligência artificial em Python e avaliamos o desempenho em uma base de dados do Kaggle. O primeiro modelo atingiu 97.57% de acurácia, enquanto o segundo, mais complexo, apresentou overfitting. Os resultados indicam oportunidades de aprimoramento e ressaltam a necessidade de estratégias eficazes diante da constante evolução das táticas de disseminação de desinformação. Por fim ressalta a importância da colaboração interdisciplinar entre especialistas em inteligência artificial, linguistas e ética para enfrentar os desafios persistentes na detecção de fake news. A busca por abordagens mais eficientes e adaptáveis a diferentes contextos permanece como uma área crucial de pesquisa.*

## 1. Introdução

Fake news é definido pelo jornal americano New York Times como "histórias inventadas feitas para enganar" e essas notícias são publicadas em formatos e meios similares a notícias verdadeiras. [Pan et al. 2018]. A propagação de Fake news é um problema que vem crescendo cada vez mais na sociedade do século XXI, com a globalização e os avanços tecnológicos ficou ainda mais difícil se dizer o que é verdadeiro ou não na internet.

Além dessa dificuldade de discernir a realidade da ficção, as fake news também espalham uma grande quantidade de desinformação e como diz o autor do seguinte artigo [Oliveira and Gomes 2019] essa desinformação é um processo complexo que envolve riscos concretos à democracia, assim como viola a noção de liberdade de expressão, comunicação e direito à informação, que tendem a organizar ações coletivas ou difundir juízos desejados pelos desinformados.

Segundo os autores de [Lorenceti and Salton 2021] , uma das grandes problemáticas é que o volume de fake news sendo compartilhada, vem aumentando e os usuários não possuem tempo hábil para verificar todas as notícias por se tratar de um processo demorado. A partir dos potenciais danos causados pelas fake news, evidencia-se a necessidade de estratégias para detecção e combate destas.

Este trabalho tem como objetivo apresentar estratégias para detecção de fake news, com vistas a indicar caminhos e possibilidades para tratar o problema de forma mais sistemática em textos da Base de Dados encontrado no Kaggle , levando em consideração o uso de Inteligência Artificial a partir de codificação em Python. Implementaremos duas redes neurais diferentes, baseando-se nas principais abordagens identificadas na revisão de literatura e criaremos um cenário de avaliação de desempenho para testar a acurácia dos algoritmos, executar simulações e obter resultados relevantes, por fim iremos disponibilizar um repositório de software contendo os códigos-fontes desenvolvidos durante este trabalho, facilitando o acesso para pesquisas futuras.

## **2. Revisão Bibliográfica**

Este capítulo tem como objetivo ser uma base teórica para nossa proposta, além de apresentar trabalhos relacionados.

### **2.1. Artigos Relacionados**

Abaixo seguirão artigos relacionados ao tema deste trabalho.

#### **2.1.1. Artigo 1**

O estudo de [Kaliyar et al. 2020] utiliza a rede neural convolucional profunda FNDNet para detecção de notícias falsas. A pesquisa aplica técnicas de inteligência artificial para identificar e classificar automaticamente notícias falsas, visando combater a disseminação de desinformação. O FNDNet extrai automaticamente características discriminatórias de notícias falsas e verdadeiras, sendo treinado em um conjunto de dados rotulado. Utiliza embeddings de palavras pré-treinados para representar o conteúdo das notícias, sem depender de características pré-definidas. O desempenho do modelo é avaliado em precisão, revocação e medida F1. A arquitetura do FNDNet, inspirada em avanços recentes na detecção de notícias falsas, inclui camadas convolucionais para extrair características discriminatórias da sequência de embeddings de palavras.

Os resultados experimentais demonstram que o FNDNet supera os métodos anteriores de detecção de notícias falsas, alcançando uma precisão de 98,36%. A pesquisa destaca a importância da seleção de hiperparâmetros e da arquitetura da rede neural convolucional profunda para a detecção eficaz de notícias falsas.

#### **2.1.2. Artigo 2**

O artigo de [Song et al. 2021] propõe um modelo para detectar notícias falsas em mídias sociais, visando combater seus impactos negativos na sociedade. O modelo utiliza atenção cruzada e redes neurais convolucionais multicanal, adotando uma abordagem multimodal

que incorpora informações de texto e imagem para avaliar a veracidade das notícias. Essa abordagem permite uma análise mais abrangente, considerando não apenas o conteúdo textual, mas também elementos visuais associados.

O modelo é composto por diversos componentes, incluindo a Rede de Resíduos de Atenção Cruzada (CARN), a Rede Neural Convolucional Multicanal (MCN), a Incorporação de Entrada e o Componente de Previsão de Notícias Falsas. Esses elementos são projetados para extrair características relevantes de diferentes modalidades e integrá-las para uma decisão final sobre a veracidade da notícia.

Experimentos conduzidos em conjuntos de dados do mundo real destacam a eficácia do modelo proposto na detecção de notícias falsas em ambientes de mídia social. Comparado a métodos convencionais, o modelo demonstra desempenho superior na integração de informações textuais e visuais, preservando as propriedades únicas de cada modalidade e reduzindo o impacto de informações ruidosas.

### **2.1.3. Artigo 3**

O estudo de [Monti et al. 2019] aborda a detecção automática de notícias falsas em mídias sociais por meio do uso de aprendizado profundo geométrico. O artigo explora técnicas desse campo para identificar padrões de propagação de notícias falsas e reais em ambientes sociais online.

O aprendizado geométrico, focado em dados estruturados, é destacado como eficaz para lidar com dados não euclidianos, como grafos. A abordagem baseada em propagação para detecção de notícias falsas é destacada, oferecendo vantagens como independência de idioma, resistência a ataques adversariais e capacidade de detecção precoce.

O modelo utilizado, um Graph Convolutional Neural Network (Graph CNN), é detalhado com suas camadas convolucionais e totalmente conectadas, além do uso de atenção de grafo e pooling médio para redução de dimensionalidade. O desempenho do modelo é avaliado com uma precisão satisfatória e uma área sob a curva ROC (AUC) próxima a 90%, demonstrando eficácia na detecção de notícias falsas em um curto período de tempo após a propagação.

### **2.1.4. Artigo 4**

O estudo de [Kumar et al. 2020] aborda a detecção de notícias falsas em redes sociais utilizando abordagens de aprendizado profundo. O trabalho coletou 1356 exemplos de notícias de usuários do Twitter e fontes de mídia, criando conjuntos de dados para notícias verdadeiras e falsas. Diversas arquiteturas de modelos de aprendizado profundo foram exploradas, incluindo Redes Neurais Convolucionais (CNN), Memórias de Longo Prazo (LSTMs), combinações de CNN e LSTMs, e modelos ensembledos com mecanismos de atenção.

Os resultados médios no conjunto de dados PolitiFact revelaram desempenhos variados das arquiteturas:

CNN: 73,29% de precisão LSTM: 80,62% de precisão LSTM bidirecional:

83,81% de precisão CNN + LSTM ensemble: 86,14% de precisão LSTM bidirecional + LSTM ensemble: 86,89% de precisão CNN + LSTM ensemble com mecanismo de atenção: 86,57% de precisão CNN + LSTM bidirecional ensemble com mecanismo de atenção: 88,78% de precisão Os resultados destacam que as redes ensembledas superaram as arquiteturas mais simples, sendo a combinação de CNN e LSTM bidirecional ensemble com mecanismo de atenção a mais precisa, atingindo 88,78

### **2.1.5. Artigo 5**

O artigo de [Machado ] explora a aplicação da inteligência artificial (IA) no combate à disseminação de notícias falsas na internet. Ele começa com uma breve história da IA e destaca a importância da conscientização sobre o uso adequado da internet. O trabalho aborda estudos que buscam melhorar a capacidade de aprendizagem da máquina para identificar e mitigar o avanço das fake news, discutindo tanto as utilidades quanto os desafios na implementação de sistemas de detecção baseados em IA.

Os benefícios da utilização da IA na detecção de fake news são enfatizados, incluindo o processamento eficiente de grandes volumes de dados e a identificação de padrões distintivos entre notícias falsas e verdadeiras. Destaca-se também o papel da IA na conscientização da população sobre a importância de verificar a veracidade das informações compartilhadas, visando reduzir a propagação de notícias falsas.

Contudo, a implementação de sistemas de detecção de fake news baseados em IA apresenta desafios, como a constante necessidade de aprimorar algoritmos, lidar com a rápida evolução das estratégias de disseminação de fake news e garantir transparência e ética no desenvolvimento e uso dessas tecnologias.

O método proposto no artigo utiliza técnicas de IA para a detecção de fake news, analisando características dos textos, como tamanho médio de palavras e riqueza do vocabulário. Destaca-se a criação de um conjunto de informações com fontes falsas e verdadeiras para alimentar a IA, melhorando sua capacidade de identificar e filtrar notícias falsas.

## **2.2. Análise dos Artigos**

Ao analisar os 5 artigos relacionados, fica evidente que as abordagens para a detecção de notícias falsas envolvem uma variedade de técnicas, desde o uso de redes neurais convolucionais (CNNs) e memórias de longo prazo (LSTMs) até a aplicação de aprendizado profundo geométrico e métodos ensembledos com mecanismos de atenção. As pesquisas destacam a eficácia dessas técnicas, evidenciando resultados promissores na identificação de notícias falsas em diferentes contextos, como mídias sociais e internet em geral. Contudo, observa-se que, apesar do progresso significativo, desafios persistentes persistem, como a necessidade de lidar com a evolução rápida das estratégias de disseminação de fake news, garantir transparência e ética na implementação de sistemas de detecção, e aprimorar a capacidade de generalização para diferentes idiomas e contextos geográficos.

Essas análises apontam para lacunas que ainda precisam ser preenchidas na pesquisa sobre detecção de notícias falsas. Questões importantes incluem como a inteligência artificial pode melhorar sua resistência a ataques adversariais e como os modelos podem ser adaptados a contextos linguísticos e culturais diversos. Além disso, a

necessidade de desenvolver métodos mais eficazes para lidar com o rápido surgimento de novas estratégias de disseminação de desinformação é uma área que merece atenção. A colaboração interdisciplinar entre especialistas em IA, linguistas, e especialistas em ética e transparência pode ser essencial para enfrentar esses desafios.

Por isso utilizaremos uma estratégia utilizada no artigo [Machado ], e implementaremos duas redes neurais com estratégias diferentes para que seja possível realizar uma análise detalhada sobre cada uma delas.

### 3. Metodologia

Neste estudo criamos duas redes neurais diferentes para que seja possível comparar seus resultados e desempenho numa mesma base de dados. Base de Dados que fez parte de uma competição do Kaggle na qual o melhor competidor alcançou 98.598% de acurácia.

Com a compreensão de todos os conceitos relacionados às arquiteturas já explicados e a revisão da literatura já realizada, o próximo passo é iniciar o processo detalhado de descrever os procedimentos e técnicas utilizados para abordar o problema de pesquisa proposto.

Esta seção aborda três subdivisões a respeito dos modelos desenvolvidos, descritas a seguir, que estão divididas em arquitetura do de cada modelo, treinamento dos modelos, e avaliação dos modelos.

#### 3.1. Arquitetura dos Modelos

##### 3.1.1. Modelo 1

Segue abaixo uma imagem da descrição da arquitetura do primeiro modelo.

Layer (type)	Output Shape	Param #
embedding (Embedding)	(None, 5000, 16)	1600000
flatten (Flatten)	(None, 80000)	0
dense (Dense)	(None, 32)	2560032
dense_1 (Dense)	(None, 1)	33
Total params: 4160065 (15.87 MB)		
Trainable params: 4160065 (15.87 MB)		
Non-trainable params: 0 (0.00 Byte)		

Figure 1. Fig1 - Estrutura do Modelo 1

Sendo obviamente um modelo mais "simples", visto que possui somente uma camada de embedding, um flatten e uma única camada totalmente conectada oculta antes da saída do modelo.

### 3.1.2. Modelo 2

Segue abaixo uma imagem da descrição da arquitetura do segundo modelo.

Layer (type)	Output Shape	Param #
embedding_2 (Embedding)	(None, 5000, 16)	1600000
lstm_1 (LSTM)	(None, 64)	20736
flatten_2 (Flatten)	(None, 64)	0
dense_4 (Dense)	(None, 32)	2080
dense_5 (Dense)	(None, 1)	33
Total params: 1622849 (6.19 MB)		
Trainable params: 1622849 (6.19 MB)		
Non-trainable params: 0 (0.00 Byte)		

**Figure 2. Fig2 - Estrutura do Modelo 2**

Sendo obviamente um modelo mais "complexo" que o primeiro, visto que além das camadas de embedding, flatten e também a camada densa, este modelo também possui uma camada LSTM.

### 3.2. Treinamento do Modelo

O modelo foi treinado utilizando o conjunto de dados rotulados do kaggle Base de Dados, importante ressaltar que apesar da base de dados ter um exemplo de dados não rotulados para teste que deve ser submetido ao site, por falta de tempo utilizamos os próprios dados de treino e dividimos em conjuntos de treino e teste. E utilizamos o padrão de 80/20 de treinamento e teste, ou seja, 80% dos dados foram separados e selecionados de maneira aleatória para essa etapa de treinamento do modelo. Importante mencionar também que o modelo foi treinado por um número fixo de épocas, sendo 10.

### 3.3. Avaliação

Finalmente, o modelo é submetido a teste utilizando um conjunto de dados separado, que não foi previamente utilizado em nenhuma etapa anterior. A avaliação do desempenho do modelo inclui o uso de métricas como precisão, recall, acurácia e loss.

Essas métricas oferecem uma medida objetiva da capacidade do modelo em classificar corretamente os diferentes tipos de tumores cerebrais. Neste estudo, 20% dos dados são reservados para testes, seguindo o padrão estabelecido pelo método de divisão entre conjuntos de treinamento e teste.

Para contextualizar os resultados obtidos, faremos uma comparação com outras abordagens de detecção de fake news presentes na literatura e previamente mencionadas nos artigos relacionados. Tais como modelos do ranking de competidores presente na base de dados.

### 3.4. Implementação

Para a implementação do trabalho descrito, utilizamos uma sessão do Google Colab, a qual possui um processador com dois núcleos, 12 GBytes de memória RAM e cache L3 de 40-50 Mbytes. E utilizamos o acelerador de hardware T4 GPU.

## 4. Experimentos

Os experimentos foram feitos utilizando a função *fit* da biblioteca *tf.keras.Model*, na qual passamos a base de dados para treinamento e o conjunto de dados de validação.

Ambos os modelos foram treinados por 10 épocas, e após isso avaliados no conjunto de dados de teste com a função *evaluate* também da biblioteca *tf.keras.Model*.

Para além disso, utilizamos a função *predict* da mesma biblioteca para aumentar o número de métricas de avaliação.

## 5. Resultados

### 5.1. Primeiro Modelo

O desempenho do primeiro modelo revelou-se notavelmente eficaz, com uma perda nos dados de teste registrada em 0.0677 e uma acurácia igual a 0.9776.

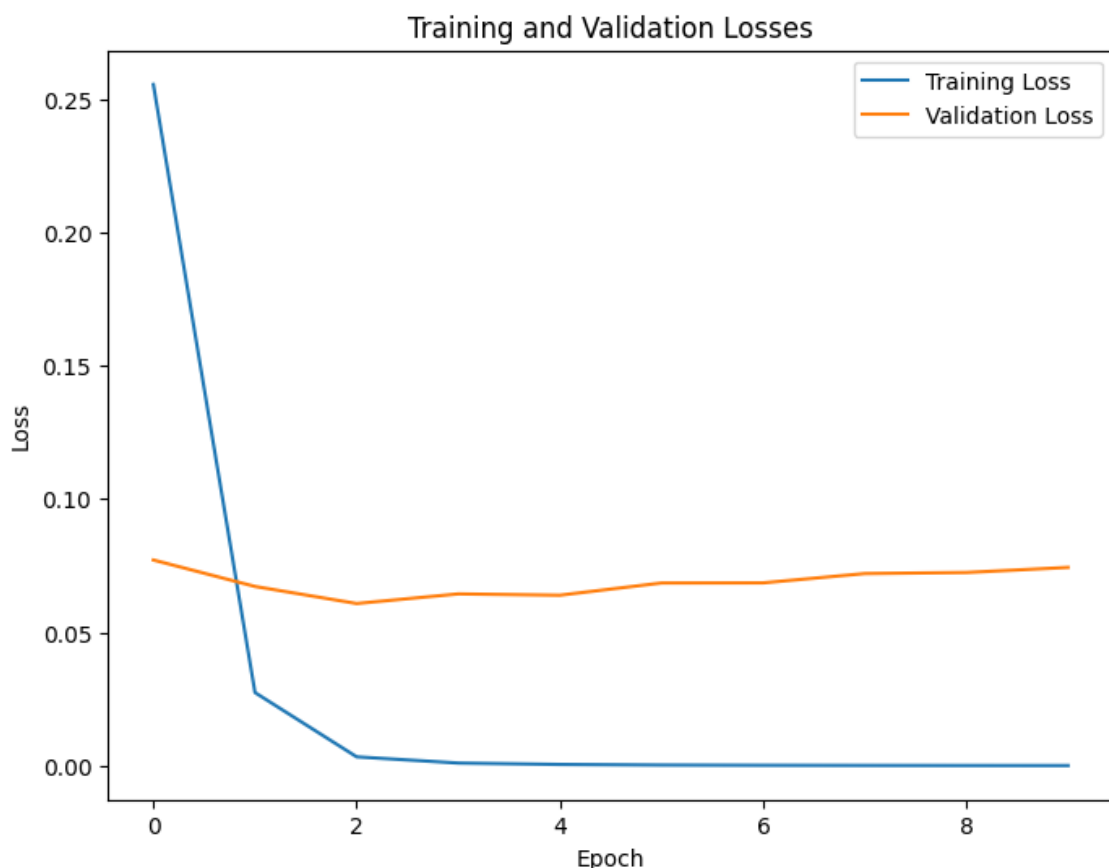


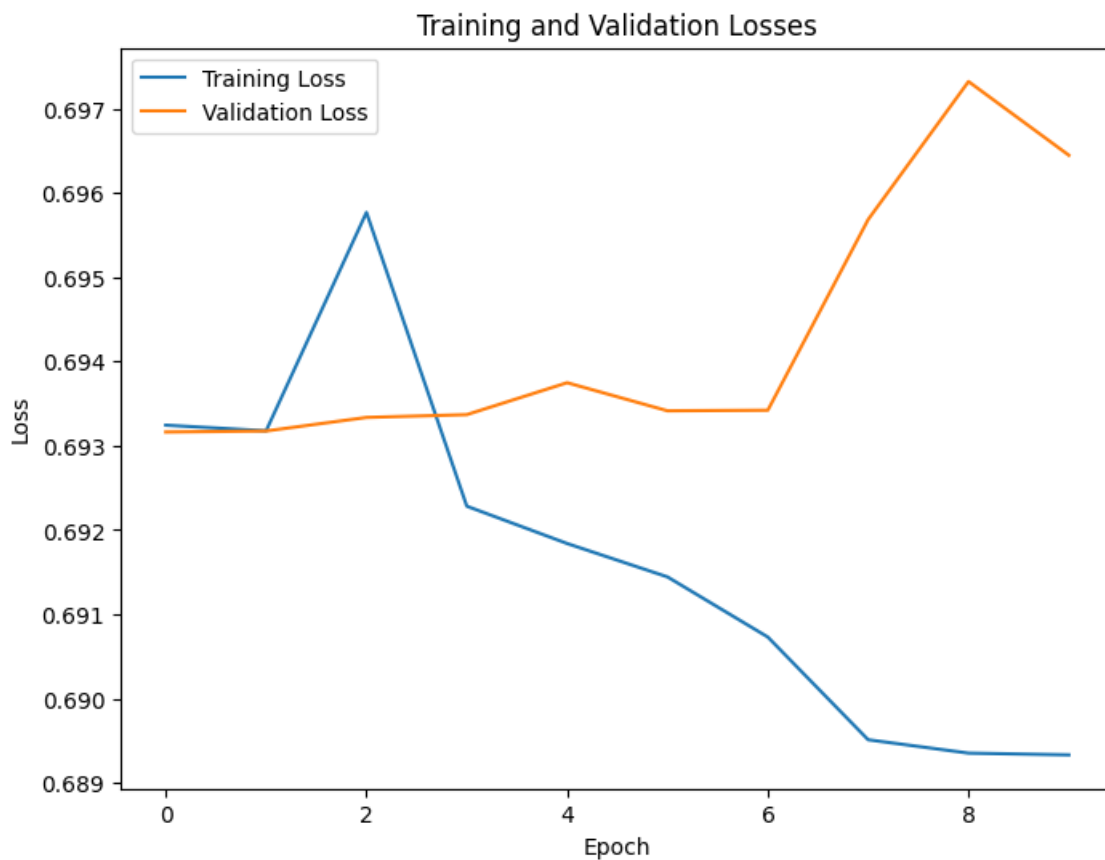
Figure 3. Primeiro modelo - Resultado Validação

Abaixo segue a tabela de métricas do primeiro modelo. **Class 0** é o equivalente a previsão de notícias verdadeiras e **Class 1** é o equivalente a previsão de notícias falsas.

	Class 0	Class 1
<b>Precision</b>	0.9781	0.9733
<b>Recall</b>	0.9728	0.9785
<b>F1-Score</b>	0.9754	0.9759

## 5.2. Segundo Modelo

No entanto, o segundo modelo não obteve um desempenho tão satisfatório, apresentando uma perda nos dados de teste de 0.6953 e uma acurácia nos dados de teste de 0.4913.



**Figure 4. Segundo modelo - Resultado Validação**

Abaixo segue a tabela de métricas do segundo modelo. **Class 0** é o equivalente a previsão de notícias verdadeiras e **Class 1** é o equivalente a previsão de notícias falsas.

	Class 0	Class 1
<b>Precision</b>	0.4971	0.8333
<b>Recall</b>	0.9995	0.0023
<b>F1-Score</b>	0.6639	0.0047

## 6. Conclusão

Após análise dos resultados, foi possível observar que embora o segundo modelo seja mais complexo, ele apresentou um desempenho pior e provavelmente sofreu de overfitting



com relação à base de dados. Essa ocorrência decorre, provavelmente, da inadequação do tamanho da base de dados em relação à complexidade do modelo. Ficou evidente que, ao utilizar o modelo mais simples (Primeiro Modelo), observamos uma melhoria nos resultados, indicando uma melhor adequação à simplicidade da base de dados escolhida.

Comparando com os resultados presentes na Base de Dados que fez parte de uma competição do Kaggle fica claro que por mais que o Primeiro Modelo implementado tenha alcançado um ótimo desempenho de 97.57% de acurácia, ainda não alcançou o estado da arte, visto que o melhor competidor alcançou 98.59% de acurácia.

Tendo isso em vista, para projetos futuros é interessante procurar formas de melhorar o Primeiro Modelo sem torná-lo complexo demais - como foi o caso do Segundo Modelo implementado - seja utilizando de mais camadas densamente conectadas, ou utilizando de estratégias diferentes, como por exemplo o aprendizado geométrico já citado nesse artigo.

## References

- [Kaliyar et al. 2020] Kaliyar, R. K., Goswami, A., Narang, P., and Sinha, S. (2020). Fndnet—a deep convolutional neural network for fake news detection. *Cognitive Systems Research*, 61:32–44.
- [Kumar et al. 2020] Kumar, S., Asthana, R., Upadhyay, S., Upreti, N., and Akbar, M. (2020). Fake news detection using deep learning models: A novel approach. *Transactions on Emerging Telecommunications Technologies*, 31(2):e3767.
- [Lorenceti and Salton 2021] Lorenceti, A. and Salton, G. (2021). Detecção de fake news em um tweet utilizando machine learning e processamento de linguagem natural. *Seminário de Iniciação Científica e Seminário Integrado de Ensino, Pesquisa e Extensão (SIEPE)*, page e29111.
- [Machado ] Machado, L. d. F. A inteligência artificial e sua interferência na detecção de fake news.
- [Monti et al. 2019] Monti, F., Frasca, F., Eynard, D., Mannion, D., and Bronstein, M. M. (2019). Fake news detection on social media using geometric deep learning. *arXiv preprint arXiv:1902.06673*.
- [Oliveira and Gomes 2019] Oliveira, A. S. and Gomes, P. O. (2019). Os limites da liberdade de expressão: fake news como ameaça a democracia. *Revista de Direitos e Garantias Fundamentais*, 20(2):93–118.
- [Pan et al. 2018] Pan, J. Z., Pavlova, S., Li, C., Li, N., Li, Y., and Liu, J. (2018). Content based fake news detection using knowledge graphs. In Vrandečić, D., Bontcheva, K., Suárez-Figueroa, M. C., Presutti, V., Celino, I., Sabou, M., Kaffee, L.-A., and Simperl, E., editors, *The Semantic Web – ISWC 2018*, pages 669–683, Cham. Springer International Publishing.
- [Song et al. 2021] Song, C., Ning, N., Zhang, Y., and Wu, B. (2021). A multimodal fake news detection model based on crossmodal attention residual and multichannel convolutional neural networks. *Information Processing & Management*, 58(1):102437.