LLMs no Combate à Desinformação: A Influência do Tamanho do Modelo na Detecção de Fake News no Brasil

Pedro T. Schettini, Guilherme Possari, Bruno S. Faiçal, Rodolfo M. Barros

[1]Departamento de Computação - Universidade Estudal de Londrina (UEL)

Abstract. *Disinformation and the spread of fake news have significant negative impacts across various domains, particularly on democracy, as they can be used to manipulate public opinion during electoral processes. Although the scientific literature presents studies aimed at detecting fake news, there are still considerable gaps regarding the Portuguese (Brazilian) language. In this context, this study investigates the hypothesis that the number of parameters in LLMs (Large Language Models) influences the detection of fake news. The results[1] suggest that the hypothesis is valid. Additionally, other influencing factors were observed, as well as a performance limit inherent to the analyzed characteristics.*

Resumo. *A desinformação e propagação de fake news (notícias falsas) causam grandes impactos negativos em diversas áreas, entre elas a democracia, porque pode ser usada para manipular a opinião pública no processo eleitoral. A literatura científica possui estudos que investigam alternativas para detectar fake news, mas ainda existem diversas lacunas para o idioma português (do Brasil). Nesse sentido, este estudo analisa a hipótese de que a quantidade de parâmetros nos modelos de LLM (do inglês, Large Language Models) impacta na identificação de fake news. Os resultados sugerem que a hipótese é válida. Adicionalmente, outros fatores podem impactar, e que existe um limite de desempenho inerente às características analisadas.*

1. Introdução

As *fake news* (notícias falsas) são notícias escritas com o intuito de enganar ou manipular pessoas em busca de alterar opiniões ou decisões sobre algum fato ou ação a ser realizada [Roumeliotis et al. 2025]. Diante do cenário brasileiro atual, as *fake news* persistem como uma ferramenta de influência em um Brasil com panorama polarizado, impactando de forma negativa ao degradar o processo democrático na escolha dos governantes [Gôlo et al. 2023].

Dado o recorrente e crescente impacto negativo das *fake news*, as propostas encontradas na literatura passam pela escolha de uma ferramenta que consiga adaptar-se diante da natureza maleável das estratégias de manipulação modernas [Roumeliotis et al. 2025]. Por conseguinte, os grandes modelos de linguagem (LLMs), destacados por sua elevada capacidade de processamento de dados e pluralidade funcional, demonstram-se como expoentes na possibilidade de diferenciar entre documentos falsos ou verdadeiros. Tal abordagem tem sido investigada e documentada por diversos estudos recentes em variados idiomas [Qu et al. 2024, Hu et al. 2024, Pelrine 2023], incluindo o idioma português [Gôlo et al. 2023].

[1]It is important to note that this study is still in progress, and the results presented in this document are of interest to the scientific community as they may contribute to new or ongoing research.

Este artigo realiza uma investigação sobre o desempenho dos LLMs em detectar *fake news* frente seu número de parâmetros. Nesse sentido, a metodologia aplicada é baseada no trabalho realizado por [Gôlo et al. 2023], utilizando o mesmo dataset composto por notícias rotuladas em 'Real' ou 'Fake', assim possibilitando a comparação entre os resultados.

Este artigo está organizado da seguinte forma: a seção 2 os experimentos e modelos utilizados para análise, enquanto a seção 3 apresenta os principais resultados alcançados. Por fim, na seção 4 são sintetizadas as principais contribuições e os próximos passos deste estudo.

2. Metodologia

A metodologia aplicada é baseada na abordagem realizada por [Gôlo et al. 2023], tendo o conjunto de notícias, *prompt* de interação com os LLMs e parte das métricas semelhantes. Contudo, são adicionados as medidas de acurácia e tempo de resposta médio de cada LLM. Essa abordagem possibilita que os resultados de ambos os estudos possam ser comparados. Os modelos de LLM utilizados são detalhados na Tabela 1.

Tabela 1. Detalhe dos LLMs utilizados neste estudo

Modelos	Tamanho (GB)	Parâmetros (Bilhões)
deepseek-r1:7b	4,7	7,62
gemma3:12b	8,1	12,20
gemma3:4b	3,3	4,30
gemma3:1b	0,8	1,00
gemma3n:e4b	7,5	6,87
gemma3n:e2b	5,6	4,46
llama3.1:latest	4,9	8,03
llama3.2:3b	2,0	3,21
llama3.2:1b	1,3	1,24

O conjunto de dados utilizado possui 300 notícias no idioma português com conteúdo ligado ao período de eleição presidencial do Brasil ou com conteúdo relevante ligado a política. A distribuição das notícias nas classes 'Real' ou 'Fake' é respectivamente de 152 e 148, originadas de diversas fontes digitais de notícias [Gôlo et al. 2023].

O prompt utilizado para requisitar que os LLMs realizem a classificação das notícias em 'Fake' ou 'Real' é definido como: *'You are a fack-checker. Answer whether the following news is fake or real. Your answer should be only the word fake or real. Follow the news:* ***news****. Remember, your answer should be only the word fake or real'*. O termo 'news' em destaque no *prompt* é substituído pelo conteúdo da notícia que se deseja classificar pelo LLM. Os experimentos foram executado em um MacBook Pro com chip Apple Silicon M4 de 10 núcleos (4 desempenho e 6 eficiência) e 16GB de memória RAM unificada LPDDR5.

As métricas utilizadas para avaliar os modelos foram precisão (*precision* - **p**), que mede a confiabilidade de determinar se as notícias são reais ou não; revocação (*recall* - **r**), que determina a cobertura que o modelo teve sobre os exemplos de cada classe; F1-Score (f_1), representado pela média harmônica entre a precisão e a revocação; f_1-macro, definido

pela média do f_1 das duas classes; e a métrica de acurácia, que indica a porcentagem total de acertos sobre todas as previsões feitas.

3. Resultados e Discussão

Diante dos resultados obtidos durante os experimentos, o modelo Gemma3 com 12 bilhões de parâmetros apresentou maior acurácia e melhor desempenho nas demais métricas, com exceção da revocação na classe *real*. Contudo, mesmo na métrica que não obteve o melhor desempenho, a diferença é de 0,01. A Tabela 2 detalha o desempenho alcançado por cada modelo avaliado no experimento de maneira específica em cada classe e de forma global através da métrica de acurácia.

Tabela 2. Comparação de medidas de desempenho

Models	fake			real			f_1-macro	Acurácia
	p	r	f_1	p	r	f_1		
deepseek-r1:7b	0,74	0,89	0,81	0,86	0,70	0,78	0,79	0,79
gemma3:1b	0,51	0,85	0,63	0,57	0,19	0,29	0,46	0,51
gemma3:4b	0,87	0,72	0,79	0,76	0,89	0,82	0,80	0,81
gemma3:12b	**0,98**	**0,94**	**0,96**	**0,94**	0,98	**0,96**	**0,96**	**0,96**
gemma3n:e2b	0,91	0,81	0,86	0,83	0,92	0,88	0,87	0,87
gemma3n:e4b	0,96	0,77	0,85	0,81	0,97	0,88	0,87	0,87
llama3.1:latest	0,97	0,44	0,60	0,64	**0,99**	0,78	0,69	0,72
llama3.2:1b	0,81	0,24	0,37	0,56	0,95	0,71	0,36	0,60
llama3.2:3b	0,90	0,24	0,37	0,57	0,97	0,72	0,55	0,61

O desempenho dos modelos Gemma3 demonstram um ganho significativo de acurácia perante a elevação da quantidade de parâmetros. De 1 bilhão para 4 bilhões, houve um ganho de uma taxa de 0,30. O mesmo é observado com o aumento de 0,15 entre os modelos de 4B para 12B. Já o Gemma3n, visto que não há nenhum ganho entre modelos, destaca-se por uma taxa constante e alta de acurácia independente do número de parâmetros. A família Llama3.2 mantém a semelhança de baixo ganho real de acurácia com aumento do tamanho do modelo, assim como o Gemma3n, porém com taxa baixa. O Llama3.1 obteve um aumento de acurácia, mas configurada como baixa também.

O Deepseek-r1 mantém um comportamento equilibrado entre as classes, mesmo apresentando o pior tempo de execução entre todos os modelos (veja a Figura 1).

Levando em consideração a Figura 1, a análise geral feita é a dominância dos modelos Gemma em relação aos demais, especialmente o de 12 bilhões. Apesar desse modelo apresentar o segundo maior tempo de execução (com aproximadamente 18 segundos no tempo médio para a classificação de cada notícia), é o modelo que alcançou o melhor desempenho na capacidade de identificar se uma notícia é verdadeira ou falsa com base no seu conteúdo. É importante ressaltar que, embora modelos maiores obtenham melhores resultados, seu uso cotidiano pode se beneficiar de heurísticas para evitar processamento redundante, como filtrar notícias já processadas.

Por fim, é importante ressaltar que que o modelo gemma3:12b alcançou desempenho semelhante ao alcançado pelo Gemma 2 (27b) apresentado por [Gôlo et al. 2023] na capacidade de identificar se o conteúdo de uma notícia é real ou falsa.

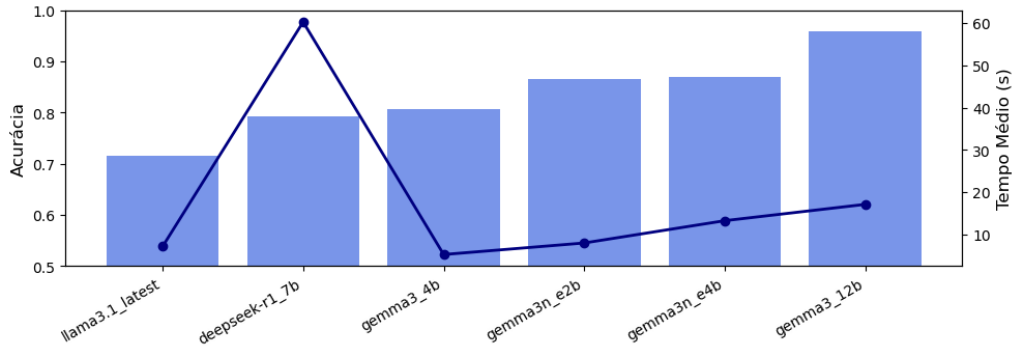

Figura 1. Comparação da Acurácia e Tempo Médio de Resposta para os modelos com acurácia maior que 0,7

4. Conclusão

Este trabalho apresentou os resultados comparativos da análise avaliativa de LLMs enquanto classificadores de notícias, entre reais ou falsas, buscando encontrar a relação entre o tamanho do modelo e o desempenho.

Diante dos resultados obtidos, foi possível observar que a família dos modelos Gemma3 teve o ganho mais expressivo de acurácia perante o aumento de parâmetros, com destaque para o Gemma3 de 12 bilhões de parâmetros. Em relação às outras implementações dos LLMs, observa-se um aumento na taxa de acurácia conforme há um aumento do tamanho com proporção menor ao apresentado pelos modelos Gemma3. Contudo, os modelos do Llama3.2 e Llama3.1 se destacaram pelo baixo desempenho alcançado apesar da quantidade de parâmetros que possui (quando comparado com outros LLMs utilizados na análise). O tempo de execução do Deepseek-r1 também se destacou de forma negativa por superar em aproximadamente $3x$ o tempo do segundo LLM com maior tempo médio de resposta (Gemma3 de 12 bilhões) e ter um desempenho inferior a diversos outros modelos utilizados, tais como Gemma3 e Gemma3n.

Esse comportamento sugere novas perguntas de pesquisa que direcionam os próximos experimentos e análises, tais como: (i) pode ser identificado limite na capacidade de LLMs (independente do número de parâmetros) em determinar a veracidade de notícias utilizando exclusivamente o seu conteúdo?; (ii) modelos com menos parâmetros de versões mais recentes de uma determinada implementação pode superar modelos maiores mais antigos?; e (iii) é possível identificar padrões no conteúdo das notícias que foram classificadas de forma errada que seja capaz de torná-la imperceptível aos LLMs?

Referências

Gôlo, M. P. S., Mori, A. L. V., Oliveira, W. G., Barbosa, J. R., Graciano-Neto, V. V., de Lima, E. A., and Marcacini, R. M. (2023). On the use of Large Language Models to Detect Brazilian Politics Fake News. In Proceedings of the Brazilian Symposium on Artificial Intelligence (SBIA) 2023, pages 1–12. Sociedade Brasileira de Computação (SBC).

Hu, B., Sheng, Q., Cao, J., Shi, Y., Li, Y., Wang, D., and Qi, P. (2024). Bad actor, good advisor: Exploring the role of large language models in fake news detection. In Proceedings of the AAAI Conference on Artificial Intelligence, volume 38.

Pelrine, A. e. a. (2023). Towards reliable misinformation mitigation: Generalization, uncertainty, and gpt-4. Arxiv.

Qu, Z., Meng, Y., Muhammad, G., and Tiwari, P. (2024). Qmfnd: A quantum multimodal fusion-based fake news detection model for social media. Information Fusion.

Roumeliotis, K. I., Tselikas, N. D., and Nasiopoulos, D. K. (2025). Fake news detection and classification: A comparative study of convolutional neural networks, large language models, and natural language processing models. Future Internet, 17(1).