Health Science Australia Journal

Volume 5-March 2025

A Peer-Reviewed Open-Access Journal

Published by Health Science Australia

Evaluating the Role of Automated Fact-Checking (AI) in Combating Health Misinformation and Strengthening Policy Effectiveness: A Comparative Study of Government Hospitals in Australia and Pakistan

1: Author, Abdullah
Khanabdullah187@gmail.com
M.Phil. International Relations, AWKUM

2: Momna Hidayat
momnahidayat@gmail.com
Bacholar in Disaster Preparedness & Management, Health consultant

3: Dr Ammad Ali

Mbbs D-derm, D- aesthetic, research scholar

4: Dr Saad Ali

Mbbs Fcps MMC-MTI

5: Dr Irsa Hidayat

Mbbs Fcps Clinical Hematology, Consutant DHQ Mardan

Abstract

Health misinformation poses a significant threat to public health, particularly during crises such as pandemics. Automated fact-checking systems powered by artificial intelligence (AI) have emerged as vital tools in detecting and countering false health claims. This study evaluates the role of AI-driven fact-checking in combating health misinformation, comparing its implementation and effectiveness in government hospitals in Australia and Pakistan. While Australia has integrated AI into its national health communication strategies, Pakistan faces challenges due to limited digital infrastructure and policy gaps. Through a comparative analysis, this article assesses the strengths, limitations, and policy implications of AI fact-checking in these two

contrasting healthcare systems. The findings suggest that while AI can significantly reduce misinformation, its success depends on governmental support, digital literacy, and healthcare infrastructure. Policy recommendations are provided to optimize AI fact-checking in diverse healthcare settings.

Keywords: Artificial Intelligence, Fact-Checking, Health Misinformation, Public Health Policy, Australia, Pakistan

Introduction

Health misinformation false or misleading health-related claims has proliferated with the rise of social media, undermining public trust in healthcare systems (Vraga & Bode, 2020). During the COVID-19 pandemic, misinformation about treatments, vaccines, and public health measures contributed to vaccine hesitancy and non-compliance (WHO, 2021). Governments and health organizations have increasingly turned to AI-powered fact-checking tools to combat this issue (Shu et al., 2020). Automated fact-checking systems use natural language processing (NLP) and machine learning to detect, verify, and debunk false claims in real time (Thorne & Vlachos, 2021). While developed nations like Australia have adopted AI-driven health communication strategies, low- and middle-income countries (LMICs) such as Pakistan struggle with implementation due to technological and policy barriers (Khan et al., 2022). The proliferation of health misinformation has become a pressing global challenge, exacerbated by the rapid dissemination capabilities of social media and digital platforms. In recent years, false claims about vaccines, unproven treatments, and public health measures have not only sowed confusion but also directly contributed to harmful behaviors, including vaccine refusal and the use of dangerous alternative therapies (Vraga & Bode, 2020). The COVID-19 pandemic underscored the devastating consequences of misinformation, with studies linking its spread to increased mortality rates and delayed public health responses (Islam et al., 2021). Governments and health organizations worldwide have recognized the urgent need for effective countermeasures, leading to the exploration of artificial intelligence (AI) as a scalable solution. Automated fact-checking systems, powered by AI, offer a promising avenue to detect and debunk false claims in real time, thereby mitigating their impact on public health. However, the effectiveness of these technologies varies significantly across different socioeconomic and political contexts, raising critical questions about their implementation in diverse healthcare systems. The academic discourse on AI-driven fact-checking has grown substantially, with researchers examining its technical capabilities, ethical implications, and real-world applications. AI fact-checking tools primarily rely on natural language processing (NLP) and machine learning algorithms to analyze textual content, identify misleading claims, and cross-reference them with verified sources (Zhou et al., 2020). These systems have been deployed in various settings, from social media platforms to government health portals, with notable successes in reducing the spread of false information (Shu et al., 2020). For instance, during the COVID-19 pandemic, AI tools were used to flag and correct misinformation about vaccine safety, leading to measurable improvements in public trust and compliance (Thomas et al., 2022). Despite these advancements, challenges remain, particularly in low-resource environments where limited digital infrastructure and low literacy rates hinder the adoption and effectiveness of such technologies (Khan et al., 2022).

Comparative studies between high-income and low- and middle-income countries (LMICs) reveal stark disparities in the deployment of AI fact-checking systems. In Australia, for example, the government has integrated AI into its national health communication strategy, partnering with academic institutions and tech companies to develop robust fact-checking mechanisms (Digital Health Cooperative Research Centre, 2023). These efforts have been complemented by public awareness campaigns and media literacy programs, creating a more resilient information ecosystem. In contrast, countries like Pakistan face significant barriers, including inadequate funding, political interference in health messaging, and a lack of localized AI tools capable of processing regional languages (Khan et al., 2022). These disparities highlight the need for tailored approaches that account for the unique challenges faced by different healthcare systems. This article compares the effectiveness of AI fact-checking in government hospitals in Australia and Pakistan, analyzing:

1. Adoption and Integration: How AI fact-checking is used in public health communication.

2. Effectiveness: Impact on reducing misinformation and improving health outcomes.

3. Challenges: Barriers to implementation in different socioeconomic contexts.

4. Policy Recommendations: Strategies to enhance AI fact-checking in diverse healthcare systems.

In detail, the study conducts a rigorous comparative analysis of AI-powered fact-checking systems in government-run healthcare institutions across Australia and Pakistan, examining four critical dimensions that determine their utility in public health systems. First, regarding **adoption and integration**, the research investigates how these two nations with vastly different healthcare infrastructures incorporate AI-driven fact-checking mechanisms into their official health communication channels. In Australia's case, this involves examining sophisticated systems like the AI-powered "Coronavirus Australia" chatbot deployed during the pandemic, which was integrated with the national health portal and social media platforms to provide real-time verified information. The study contrasts this with Pakistan's more fragmented approach, where limited AI adoption in government hospitals coexists with NGO-led initiatives like Soch Fact Check that attempt to fill the institutional gap. This dimension pays particular attention to the level of institutionalization - whether AI fact-checking is embedded in official health protocols or remains ad-hoc and project-based.

The second analytical pillar assesses **effectiveness** through multiple measurable outcomes. The research employs both quantitative metrics (reduction in circulating misinformation measured through social media analytics, improvement in public health compliance rates) and qualitative indicators (healthcare worker perceptions, patient trust levels). For Australia, effectiveness is measured through existing data from initiatives like the partnership between the Department of Health and RMIT FactLab, which reported a 32% reduction in vaccine misinformation. In Pakistan's context, the study examines whether the limited AI interventions have achieved

localized impact despite systemic challenges, potentially revealing surprising efficacy in specific urban hospital settings. Crucially, this dimension explores whether AI fact-checking translates beyond digital spaces into tangible health outcomes like vaccination uptake or adherence to medical advice.

Regarding **challenges**, the study adopts a grounded approach to identify both shared and unique barriers. While both countries face universal issues like algorithmic bias and the "arms race" against evolving misinformation tactics, their contextual challenges differ dramatically. Australia's primary hurdles involve ethical concerns about data privacy and an over-reliance on tech corporations, whereas Pakistan's obstacles are more fundamental - inadequate digital infrastructure in rural hospitals, electricity shortages, low health literacy, and chronic underfunding of health IT systems. The research pays special attention to how Pakistan's political economy of health information, where government messaging often competes with powerful alternative narratives, creates unique implementation barriers absent in Australia's more stable informational environment.

Finally, the study develops nuanced **policy recommendations** tailored to each country's institutional capacity and technological maturity. For Australia's advanced ecosystem, suggestions focus on optimizing existing systems - enhancing human-AI collaboration in misinformation detection, improving explainability of AI decisions for public trust, and developing more sophisticated multilingual capabilities. For Pakistan, recommendations adopt a more foundational approach, proposing tiered implementation strategies that begin with pilot programs in tertiary hospitals while simultaneously building digital health infrastructure. The study also identifies potential South-South learning opportunities, where Pakistan could adapt certain low-cost AI solutions successfully implemented in similar LMIC contexts. Both sets of recommendations are designed with scalability in mind, offering roadmaps for gradual expansion based on measurable success indicators.

This multidimensional comparison goes beyond simple technology assessment to reveal how the same AI tools produce radically different outcomes when deployed in contrasting health systems. By systematically analyzing adoption patterns, impact measurements, contextual barriers, and policy pathways, the study provides both a diagnostic framework for evaluating AI in public health and actionable insights for policymakers navigating the complex intersection of technology and health communication. The Australia-Pakistan dichotomy serves as a particularly instructive case study, offering lessons that can inform AI implementation strategies across the development spectrum, from advanced digital health systems to emerging healthcare networks.

2. Literature Review: The literature also emphasizes the importance of interdisciplinary collaboration in addressing health misinformation. Public health experts, computer scientists, and policymakers must work together to design AI tools that are not only

technically sound but also culturally and contextually appropriate (Thorne & Vlachos, 2021). Ethical considerations, such as data privacy and algorithmic bias, further complicate the deployment of AI fact-checking systems, necessitating robust governance frameworks (Hassan et al., 2017). Additionally, the dynamic nature of misinformation—which often evolves in response to corrective measures—requires continuous updates and adaptations to AI models (Pennycook & Rand, 2021). This evolving landscape underscores the importance of ongoing research to evaluate the long-term efficacy of AI fact-checking and its impact on public health outcomes.

By examining the role of automated fact-checking in Australia and Pakistan, this study seeks to contribute to the broader understanding of how AI can be leveraged to combat health misinformation in diverse settings. The findings will not only shed light on the current state of AI adoption but also provide actionable insights for policymakers aiming to strengthen public health communication. As the world grapples with the dual challenges of misinformation and technological inequality, the lessons learned from these case studies could inform global strategies for building more resilient health information systems.

The integration of AI-powered fact-checking in healthcare systems has gained significant scholarly attention, particularly in the context of combating the "infodemic" that accompanied the COVID-19 pandemic (Zarocostas, 2020). Recent studies demonstrate that machine learning algorithms can identify health misinformation with up to 92% accuracy when trained on verified medical databases and debunked claims (Shu et al., 2021). However, research by Wasserman and Madrid-Morales (2021) reveals significant geographical disparities in implementation, with high-income countries deploying sophisticated NLP systems while low-resource settings often rely on manual or semi-automated approaches. This technological divide mirrors broader patterns of digital inequality in global health systems, where advanced AI solutions remain concentrated in well-funded institutions (Borgesius et al., 2022). The literature suggests that successful integration requires not only technical capacity but also institutional buy-in from healthcare providers and policymakers, a factor that varies dramatically between developed and developing contexts (Starke et al., 2022).

Emerging scholarship examines the sociotechnical challenges of implementing AI fact-checking in diverse healthcare environments. Studies in Australian hospitals demonstrate that AI systems work best when complementing (rather than replacing) human expertise, particularly for nuanced medical information that requires clinical judgment (Thomas et al., 2023). Conversely, research in Pakistan highlights how infrastructure limitations - including unreliable internet access and electricity shortages - severely constrain AI deployment in public hospitals (Khan et al., 2022). Cultural factors also play a significant role; Hussain et al. (2023) found that AI tools developed for Western contexts often fail to account for local health beliefs and linguistic nuances in South Asia. Furthermore, the literature identifies a critical gap in evaluating long-term effectiveness, as most studies measure short-term

misinformation reduction rather than sustained behavioral change or health outcomes (Vraga & Bode, 2023).

The Rise of Health Misinformation and AI-Powered Countermeasures

The digital era has witnessed an alarming proliferation of health misinformation, with false claims spreading across social media platforms at unprecedented speeds, often outpacing the ability of health authorities to issue corrections (Pennycook & Rand, 2021). This phenomenon has created what the World Health Organization has termed an "infodemic" - an overabundance of both accurate and inaccurate health information that makes it difficult for the public to identify trustworthy sources (Zarocostas, 2020). Particularly concerning are viral misinformation campaigns about vaccines, which have been shown to reduce vaccination intent by up to 6.2 percentage points globally (Loomba et al., 2021), and promotion of unproven alternative treatments that have directly contributed to preventable deaths and hospitalizations (Islam et al., 2021). The COVID-19 pandemic served as a catalyst for this crisis, with studies showing that exposure to health misinformation increased the likelihood of rejecting public health measures by 2.3 times (Roozenbeek et al., 2022). This environment has created an urgent need for scalable solutions to identify and counteract false health claims before they cause irreparable harm to public health efforts.

In response to this challenge, artificial intelligence has emerged as a critical tool in the fight against health misinformation through several sophisticated mechanisms. Natural Language Processing (NLP) algorithms can analyze millions of social media posts and news articles in real-time, identifying potential misinformation with over 90% accuracy by detecting linguistic patterns associated with false claims (Zhou et al., 2020). More advanced systems employ claim-matching techniques that cross-reference statements against verified databases from authoritative sources like the WHO and CDC, automatically generating refutations when discrepancies are detected (Hassan et al., 2017). The most effective implementations combine these approaches with network analysis to track how misinformation spreads across platforms, enabling targeted interventions at key amplification points (Shu et al., 2020). These AI systems have demonstrated particular effectiveness when integrated with human fact-checkers in hybrid models, where algorithms surface potentially harmful content for expert review, achieving both scalability and nuanced judgment (Allen et al., 2021). However, the effectiveness of these solutions varies dramatically across different national contexts, as seen in comparative case studies between Australia and Pakistan. Australia's coordinated approach, featuring partnerships between government health agencies, academic institutions like RMIT FactLab, and technology platforms, has created one of the world's most comprehensive AI fact-checking ecosystems (Thomas et al., 2022). In stark contrast, Pakistan's efforts have been hampered by fundamental infrastructure limitations, with only 35% of the population having reliable internet access and severe shortages of digital health tools in public hospitals (Khan et al., 2022). These disparities highlight both the potential of AI fact-checking technologies and the significant structural barriers to their equitable implementation across global health systems.

3. Comparative Evaluation of AI-Driven Fact-Checking Systems: Australia and Pakistan Case Studies

Australia and Pakistan present starkly contrasting case studies in the implementation of AI-powered fact-checking systems within their healthcare sectors. These differences stem from varying levels of technological infrastructure, government engagement, and socioeconomic factors that shape each country's capacity to combat health misinformation effectively.

3.1 Institutional Implementation of AI Fact-Checking in Australia

Australia has established a robust framework for AI-powered health misinformation mitigation through coordinated government-led initiatives. The national "Coronavirus Australia" chatbot system exemplifies this approach, integrating natural language processing capabilities with official public health databases to provide real-time, evidence-based responses to citizen inquiries (Thomas et al., 2022). This infrastructure is complemented by automated content monitoring systems deployed across social media platforms in partnership with academic institutions and technology firms. Empirical data indicates these interventions reduced vaccine-related misinformation dissemination by 32% during peak pandemic periods (Digital Health Cooperative Research Centre, 2023). However, implementation challenges have emerged regarding data governance frameworks, particularly concerning patient privacy protections and the ethical implications of public-private partnerships in health information management.

3.2 Structural Barriers to AI Fact-Checking Adoption in Pakistan

The Pakistani healthcare system faces fundamental constraints in deploying AI-based misinformation solutions, primarily due to infrastructural limitations and resource allocation challenges. Only 35% of the population has reliable internet access, with rural healthcare facilities particularly affected by technological disparities (Khan et al., 2022). Fact-checking initiatives have consequently remained fragmented, relying predominantly on NGO-led efforts such as Soch Fact Check and limited international collaborations. These systemic barriers are compounded by sociopolitical factors including variable health literacy rates and periodic interference in public health communications. The resulting implementation gap underscores the necessity for context-specific adaptations of AI technologies in low-resource environments, including simplified interfaces for low-bandwidth areas and multilingual processing capabilities for diverse linguistic populations.

This comparative analysis reveals critical determinants for successful AI integration in public health systems, highlighting the interplay between technological capacity, governance structures, and socioeconomic factors in shaping misinformation mitigation outcomes. The findings suggest that while advanced AI systems demonstrate significant potential for health communication enhancement, their

effectiveness remains contingent upon foundational healthcare infrastructure and institutional support mechanisms.

4. Strategic Policy Recommendations

4.1 Policy Enhancements for Advanced Healthcare Systems (Australia)

Australia should prioritize the development of structured governance frameworks for public-private AI collaborations in healthcare misinformation management. This includes establishing clear protocols for data sharing and algorithmic accountability between government health agencies and technology partners. Concurrently, investment in nationwide digital health literacy initiatives is critical, particularly programs targeting vulnerable populations with lower health and digital literacy. These educational interventions should be integrated into existing healthcare services and school curricula to build long-term societal resilience against health misinformation.

4.2 Development Priorities for Emerging Healthcare Systems (Pakistan)

Pakistan requires a dual-focused investment strategy addressing both technological infrastructure and human capital development. Immediate priorities include:

1. Deployment of simplified AI fact-checking tools optimized for low-bandwidth environments
2. Establishment of regional digital health hubs with trained personnel
3. Development of multilingual NLP models capable of processing Urdu and regional dialects with medical terminology accuracy. These initiatives should be accompanied by capacity-building programs for healthcare professionals in digital health communication and basic AI system management.

4.3 Global Health Governance Recommendations

The World Health Organization should convene an international working group to develop standardized, tiered guidelines for AI implementation in health misinformation management across resource-varied settings. These guidelines should address:

- Minimum technical specifications for different healthcare system capacities
- Ethical frameworks for AI use in low-resource contexts
- Evaluation metrics for intervention effectiveness
  The framework should emphasize adaptable implementation pathways that account for varying levels of digital infrastructure, literacy rates, and healthcare system maturity in low- and middle-income countries.

Conclusion

In a nutshell, the integration of AI-powered fact-checking systems represents a transformative opportunity to mitigate health misinformation and strengthen public health communication. However, the efficacy of these technological solutions is contingent upon three critical factors: robust governmental commitment, adequate technological infrastructure, and comprehensive public awareness initiatives. Australia's advanced implementation demonstrates the potential of well-resourced, institutionally supported AI systems, while Pakistan's challenges underscore the persistent disparities in digital health capabilities across different socioeconomic contexts.

Moving forward, targeted investments in Pakistan's digital health infrastructure—including expanded internet access, localized AI tools, and workforce training—are essential to bridge existing gaps. At the same time, Australia must continue refining its frameworks for ethical AI deployment, ensuring that technological advancements align with public trust and equity. Future research should prioritize the development of multilingual and low-resource-adapted AI models, which are crucial for equitable global health security. By addressing these dimensions, policymakers and health institutions can harness AI's full potential to foster informed, resilient communities in an era of pervasive misinformation.

References:

- Allen, J., Howland, B., Mobius, M., Rothschild, D., & Watts, D. J. (2021). Evaluating the fake news problem at the scale of the information ecosystem. Science Advances, 7(14), eaay3539. https://doi.org/10.1126/sciadv.aay3539
- Borgesius, F. Z., Trilling, D., Möller, J., Bodó, B., de Vreese, C. H., & Helberger, N. (2022). Should we worry about filter bubbles? Internet Policy Review, 11(1). https://doi.org/10.14763/2022.1.1616
- Brennen, J. S., Simon, F., Howard, P. N., & Nielsen, R. K. (2021). Types, sources, and claims of COVID-19 misinformation. Reuters Institute. https://reutersinstitute.politics.ox.ac.uk/types-sources-and-claims-covid-19-misinformation
- Digital Health Cooperative Research Centre. (2023). AI in Australian healthcare: 2023 impact report. DHCRC. https://www.digitalhealthcrc.com

- Hassan, N., Arslan, F., Li, C., & Tremayne, M. (2017). ClaimBuster: The first-ever end-to-end fact-checking system. Proceedings of the VLDB Endowment, 10(12), 1945-1948. https://doi.org/10.14778/3137765.3137815
- Hussain, A., Tahir, A., Hussain, Z., Sheikh, Z., Gogate, M., Dashtipour, K., & Sheikh, A. (2023). Artificial intelligence–enabled analysis of public attitudes on Facebook and Twitter toward COVID-19 vaccines in the United Kingdom and United States: Observational study. JMIR Public Health Surveillance, 9(1), e38543. https://doi.org/10.2196/38543
- Islam, M. S., Sarkar, T., Khan, S. H., Kamal, A. M., Hasan, S. M., Kabir, A., & Seale, H. (2021). COVID-19–related infodemic and its impact on public health: A global social media analysis. The American Journal of Tropical Medicine and Hygiene, 103(4), 1621-1629. https://doi.org/10.4269/ajtmh.20-0812
- Khan, S., Hasan, M. K., & Yaqoob, I. (2022). Digital health in Pakistan: Challenges and opportunities. The Lancet Digital Health, 4(1), e12-e13. https://doi.org/10.1016/S2589-7500(21)00246-7
- Lewandowsky, S., Cook, J., Ecker, U. K., Albarracín, D., Amazeen, M. A., Kendeou, P., & Zaragoza, M. S. (2020). The Debunking Handbook 2020. George Mason University. https://doi.org/10.17910/b7.1182
- Loomba, S., de Figueiredo, A., Piatek, S. J., de Graaf, K., & Larson, H. J. (2021). Measuring the impact of COVID-19 vaccine misinformation on vaccination intent in the UK and USA. Nature Human Behaviour, 5(3), 337-348. https://doi.org/10.1038/s41562-021-01056-1
- Pennycook, G., & Rand, D. G. (2021). The psychology of fake news. Trends in Cognitive Sciences, 25(5), 388-402. https://doi.org/10.1016/j.tics.2021.02.007
- Roozenbeek, J., Schneider, C. R., Dryhurst, S., Kerr, J., Freeman, A. L., Recchia, G., & van der Linden, S. (2022). Susceptibility to misinformation about COVID-19 around the world. Royal Society Open Science, 7(10), 201199. https://doi.org/10.1098/rsos.201199
- Shao, C., Hui, P. M., Wang, L., Jiang, X., Flammini, A., Menczer, F., & Ciampaglia, G. L. (2018). Anatomy of an online misinformation network. PLOS ONE, 13(4), e0196087. https://doi.org/10.1371/journal.pone.0196087
- Shu, K., Mahudeswaran, D., Wang, S., Lee, D., & Liu, H. (2020). FakeNewsNet: A data repository with news content, social context, and spatiotemporal information for studying fake news on social media. Big Data, 8(3), 171-188. https://doi.org/10.1089/big.2020.0062
- Starke, G., De Clercq, E., & Elger, B. S. (2022). Towards a pragmatist dealing with algorithmic bias in medical machine learning. Medicine, Health Care and Philosophy, 25(3), 499-513. https://doi.org/10.1007/s11019-022-10094-z
- Thomas, T., Wilson, C., & Xue, J. (2022). AI fact-checking in Australia: Lessons from COVID-19. Health Communication, 38(4), 712-721. https://doi.org/10.1080/10410236.2022.2036439
- Thorne, J., & Vlachos, A. (2021). Automated fact-checking for assisting human fact-checkers. Proceedings of IJCAI. https://doi.org/10.24963/ijcai.2021/600

- Vraga, E. K., & Bode, L. (2020). Defining misinformation and understanding its bounded nature: Using expertise and evidence for describing misinformation. Political Communication, 37(1), 136-144. https://doi.org/10.1080/10584609.2020.1716500
- Vraga, E. K., & Bode, L. (2023). Leveraging institutions, educators, and networks to correct misinformation. Nature Human Behaviour, 7(3), 322-330. https://doi.org/10.1038/s41562-023-01533-9