International Journal of Public Administration in the Digital Age
Volume 13 • Issue 1 • January-December 2026

From Retaliation to Protection:
Reimagining Whistleblowing in the South African Public Sector Through AI-Driven Solutions

Christian Kayembe
https://orcid.org/0000-0002-1014-0037
University of Johannesburg, South Africa

Mbasa Mvenene
https://orcid.org/0000-0002-1898-5819
Walter Sisulu University, South Africa
Received: October 3rd, 2025 | **Accepted:** January 7th, 2026

ABSTRACT

This study examines how Artificial Intelligence (AI) can enhance whistleblower protection in South Africa's public sector, where corruption continues to undermine governance and service delivery. Using a desktop qualitative approach, the study reviewed peer-reviewed literature, policy reports, and credible online sources through documentary and conceptual analysis. The findings highlight that whistleblowers remain critical in exposing corruption, malpractice, and maladministration, yet they face high levels of retaliation and insecurity. AI tools such as encrypted chatbots and blockchain platforms present practical solutions to safeguard anonymity, strengthen accountability, and improve trust in reporting systems. The study concludes that leveraging AI in public procurement processes could both protect whistleblowers and curb corruption. These insights contribute to ongoing debates on digital innovations in the Fourth Industrial Revolution and their implications for public policy.

KEYWORDS

Artificial Intelligence, Corruption, Public Sector, Whistleblowing, South Africa

INTRODUCTION

South Africa's public sector continues to be plagued by entrenched corruption, undermining governance, service delivery, and public trust. Despite numerous anti-corruption efforts, including legislative reforms, oversight bodies, and civil society initiatives, systemic corruption persists, often fueled by a culture of impunity and institutional complicity (Calitz, 2023; Mkhize, 2023; Onyango, 2021). One of the most potent tools for promoting accountability in such contexts is whistleblowing. However, in South Africa, whistleblowers face significant threats to their safety and livelihoods, which has created a chilling effect on disclosures (Nortje, 2023; Radulovic, 2023).

Whistleblowing, defined as the disclosure of misconduct, fraud, or illegality within an organization to internal or external actors (Miceli et al., 2008), is a critical mechanism in democratic accountability. The South African Human Rights Commission (2022) describes it as an indispensable element in the fight against corruption, particularly in post-apartheid South Africa's aspiration to become a truly ethical Rainbow Nation. Yet, between 2000 and 2021, over 1970 politically or corruption-motivated assassinations were reported, many targeting whistleblowers (Global Initiative Against Transnational

DOI: 10.4018/IJPADA.399055

Organised Crime, 2022). High-profile cases such as the assassinations of Babita Deokaran in 2021 and Cloete and Thomas Murray in 2023 underscore the extreme risks faced by individuals who expose public sector malfeasance (Radulovic, 2024; Special Investigating Unit, 2023).

These incidents are not isolated. Deokaran, an acting Chief Director in the Gauteng Department of Health, was targeted for exposing R850 million in irregular COVID-19 procurement contracts. Similarly, whistleblowers like Jimmy Mohlala (died in 2009), Moss Phakoe (died in 2009), Sindiso Magaqa (died in 2017), and Cloete Murray (died in 2023) were all killed after exposing corruption in local government and national procurement processes (Corruption Watch, 2013; Matamba, 2023; Yende, 2019). These murders reveal a systemic failure to protect whistleblowers and enforce accountability.

South Africa's legislative framework for whistleblower protection, primarily the Protected Disclosures Act (PDA) 26 of 2000 and its 2017 Amendment, offers formal protections but suffers from serious limitations in implementation and scope (Calitz, 2023; Nwoke, 2019). Key deficiencies include the lack of anonymity guarantees, inadequate institutional support, the absence of incentives, and weak enforcement mechanisms. As Onyango (2021:9) argues, whistleblower protection in South Africa remains "underperformed in practice and underdeveloped in theory," with most of the robust scholarly models developed in Global North contexts lacking applicability to South African realities.

Corruption Watch (2013) noted that a primary barrier to reporting corruption is the fear of reprisal. This fear is well-founded: whistleblowers often suffer career sabotage, public vilification, financial ruin, psychological distress, and physical harm. This retaliatory ecosystem not only silences whistleblowers but dis-incentivizes future disclosures, weakening transparency in governance (Berendt & Schiffner, 2022; Zouvia, 2020). The traditional, largely paper-based whistleblowing channels in South Africa are not fit for purpose in an environment where confidentiality and protection are life-and-death issues. Existing mechanisms fail to guarantee anonymity, non-traceability, or secure data transmission, critical factors in preventing exposure and retaliation (Elhassan Gadour, 2024; Odilla, 2023).

In response to these systemic weaknesses, this study explores how artificial intelligence (AI) can be harnessed to enhance the safety, protection, and anonymity of whistleblowers in South African public sector procurement. Literature increasingly points to the transformative potential of AI technologies, including encrypted chatbots, machine learning (ML) algorithms, blockchain-powered disclosure platforms, and natural language processing (NLP) tools, to detect fraud, secure disclosures, and monitor retaliation risks in real time (Brand, 2020; Jayakrishnan & Murali, 2019; Singh et al., 2023). Moreover, digital platforms can guide whistleblowers in making legally sound disclosures while maintaining strict confidentiality.

Technologies such as anonymous online drop boxes and secure digital reporting tools, pioneered in jurisdictions like Germany, the United States, and Sweden, offer compelling models for South Africa (Berendt & Schiffner, 2022). These systems utilize privacy-enhancing technologies and end-to-end encryption to ensure whistleblower anonymity, which is often the most effective shield against retaliation. However, their successful implementation depends on legal and institutional reforms that embed such digital tools within the country's whistleblowing frameworks.

While South Africa ratified the United Nations Convention Against Corruption (UNCAC) in 2004 and committed to protecting whistleblowers under Article 32, substantive measures remain largely aspirational. The convention mandates that states provide whistleblowers with physical protection and access to safe disclosure channels, including the use of remote communication technologies. Yet, more than two decades later, the South African government has yet to adopt AI-based systems capable of delivering on this commitment (Just Share, 2022; United Nations, 2004).

This study seeks to examine how AI technologies can be leveraged to enhance whistleblower protection in South Africa's public sector, with a focus on procurement processes. The central research question guiding this investigation is: How can AI be leveraged to enhance whistleblower safety