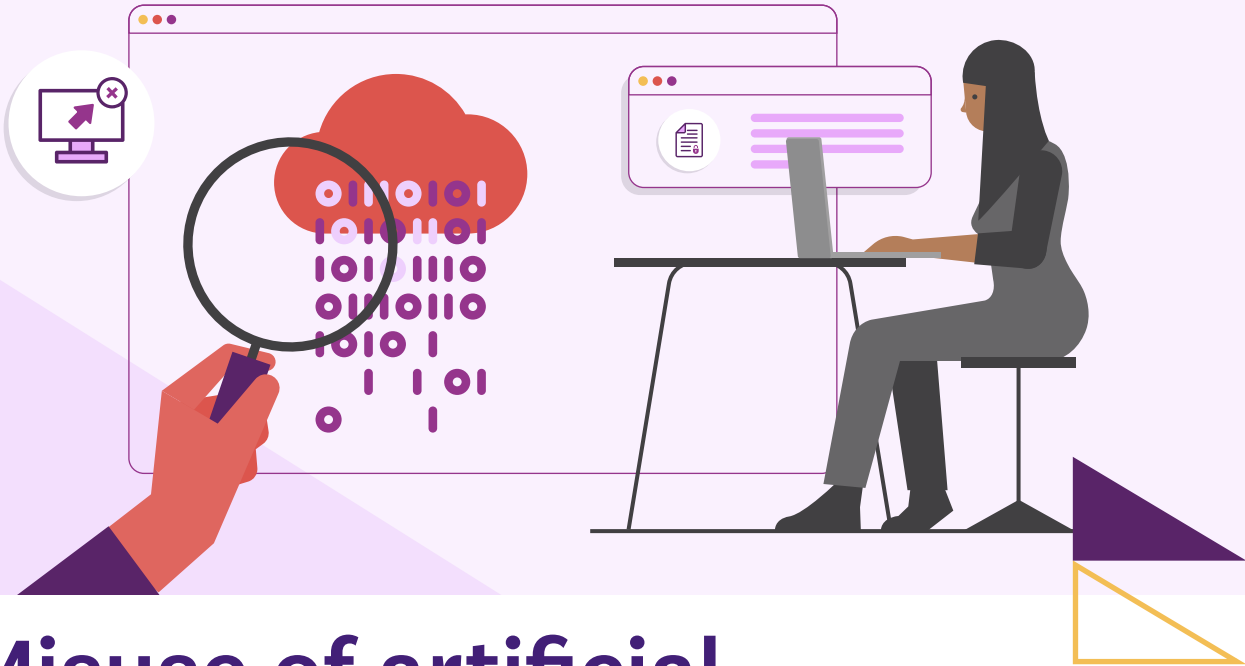




# Misuse of artificial intelligence in data analysis

## Guidance material – integrity scenario

This guidance material presents an ‘integrity scenario’ which is intended to assist in public sector training regarding integrity risks, exploring how without thorough oversight, the implementation of an artificial intelligence (AI) system in government data analysis can lead to misconduct and corruption. It highlights the risks of relying solely on complex AI models, the importance of transparency and robust oversight in the implementation and use of AI models. The scenario demonstrates how pressure to show positive results, combined with the complexity of AI models, can lead to unethical behaviour and compromise public trust.

### The situation

A Victorian government department implemented an AI system to analyse complex qualitative datasets, including survey responses and community feedback on a critical policy area. The AI’s analysis across the first year of data showed positive trends, including increased community satisfaction with service delivery and endorsement of the policy. These outcomes were presented in briefing products and external publications.

The department’s data analyst was responsible for overseeing the AI system’s operation. Their tasks included managing the survey data from the community, preparing the code for analysis (including implementing the AI models and assumptions that underpinned the analysis), verifying results and presenting findings to department leadership and the public.

The AI system worked well and saved the department considerable time and resources. The results were well received and showed increased community satisfaction in the first year, reflecting favourably on the department.

## How did the misconduct start?

Following an initial period of encouraging results, the data analyst noticed that the AI's analysis of the department's data was becoming more negative over time. The data analyst presented their initial findings to their direct manager. The manager had recently been in a meeting with senior stakeholders in their department and had presented on the importance of showcasing the AI's benefits to the community in this policy area. The manager - with this political context in mind - directed the data analyst to review and reconsider the assumptions underpinning the AI model analysis, to see if the results could become 'more favourable' for the position advocated by the department. The manager also hinted that the data analyst's chances of securing a promotion would be higher if they were able to present encouraging results and secure more funding for their team.

The data analyst initially felt uncomfortable with this request, but did not want to disappoint their manager or potentially slow the momentum of their career growth. The data analyst decided to adjust some of the code that underpinned the AI model's analysis. The analyst made some minor tweaks until the findings were presented in a more optimistic light. These changes were not documented, and when the findings were published again for department leadership and the public, the change in methodology was not made clear.

## How did it escalate?

Encouraged by the praise and the apparent success of the AI system, the data analyst continued to make small adjustments to the methodology when predictions did not align with expectations or political goals. The analyst also began to selectively filter out data that would not lead to desired outcomes, justifying it as 'smoothing out anomalies' in the AI's processing.

The department lacked clear protocols for managing and adjusting AI inputs and outputs. Few people in the department - beyond the data analyst and their manager - understood how the AI model worked. The absence of ethical guidelines and insufficient training on the AI's model led to a reduced critical evaluation of its predictions, or adequate oversight of its use.

The issue was exposed when a media report was published outlining the community's lack of support for a policy position, which went against the findings published on the website by the department. Some senior executives in the department were suspicious of the findings and ordered an investigation - by an independent provider - into the survey data and feedback that was provided as input to the AI model.

The investigation uncovered the data analyst's and the manager's long-standing practice of manipulating the results. It revealed that the AI system, if fed accurate data, had been correctly predicting less favourable trends, but these warnings were obscured by the analyst's adjustments.

While the data analyst had not profited personally from these actions and did not secure the promotion they had hoped for, their manipulation of the process had misled the public and was a breach of the professional standards of the public service to provide frank and fearless advice. As a result of the investigation, the analyst and manager received formal warnings and were reassigned to different roles.





## Lessons learned

- Ensure there is transparency in how AI (and other models) are used and how to interpret results.
- Allow for third party verification of AI models and other data models within departments.
- Implement robust data verification processes, including regular audits of AI inputs and outputs.
- Maintain human oversight of AI-driven analyses, especially in areas critical to public interest.
- Provide comprehensive AI ethics and literacy training to all staff involved in AI-driven processes.
- Establish clear guidelines for data management and adjustment, with requirements for documentation and peer review.
- Create safe channels for reporting concerns about AI systems or data integrity without fear of reprisal.
- Regularly rotate responsibilities in areas where AI is used to prevent the emergence of unchecked data manipulation patterns.
- Ensure transparency in AI-driven processes, ensuring data sources and analytical methods are understandable to stakeholders and the public.

This product was prepared based on findings from research and stakeholder consultations from IBAC's 2024 Public Sector Strategic Assessment. It is representative and created for educational purposes only. Any resemblance to real persons, organisations, or incidents is purely coincidental.

## If you experience or suspect public sector corruption, report it to IBAC



Fill out the secure online form to report at [www.ibac.vic.gov.au](https://www.ibac.vic.gov.au)



If you have difficulty accessing the online form, call us on **1800 422 284** for further assistance.



If you need help with translation, call Translating and Interpreting Service on **13 14 50** or visit [www.ibac.vic.gov.au/mylanguage](https://www.ibac.vic.gov.au/mylanguage)