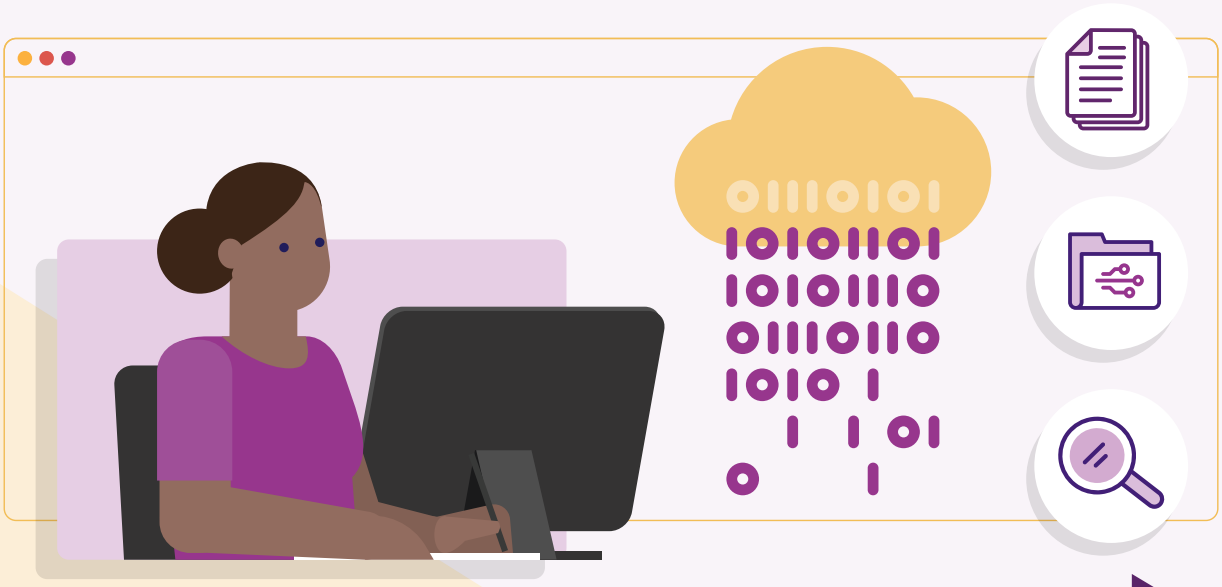

Integrity risks associated with artificial intelligence

Guidance material

This guidance material examines the integrity risks that are associated with the use of artificial intelligence (AI). AI is a fast-emerging technology that enables computer systems to perform tasks normally requiring human input, including information generation, visual perception, speech recognition, decision-making and translation between languages.

AI systems, if misused, can present integrity risks in their design and application within the public service. This includes risks associated with the use of AI by public servants, and external individuals or organisations using AI to mislead public servants. Public sector organisations should provide clear guidelines to staff to ensure AI is used with integrity in mind.

What are the risks?

AI systems can be intentionally designed for corrupt purposes

AI systems present opportunities to deliberately mislead, misrepresent or manipulate information. Examples of integrity risks associated with AI systems include:

- Committing identity fraud, such as using video or audio manipulation to pretend to be senior public servants or politicians, which can lead to risks of sensitive information being leaked.
- Generating audio or video content that is deliberately fake or misleading, which can spread misinformation and create distrust in public institutions.
- Manipulating AI systems that screen resumes for employment, to deliberately favour certain groups or individuals, leading to improper employment practices.

- Using AI systems to generate text, answer questions or customise content in a way that mimics human writing – this can influence reports and content, based on inherent bias that may exist in the AI system, and the data or information it utilises to form its writing.
- Automating administrative tasks, which may seem to increase productivity but can also be used fraudulently (e.g., to generate false invoices or timesheets).
- Consider where and how the information you input into the AI system is being stored, and how it may be used or shared. Inputting sensitive information or data may be a breach of privacy or security.

AI systems can be manipulated with false input data

AI systems are designed for specific tasks, and it can be difficult to see how they make decisions or understand the processes behind their results. This lack of transparency can lead to significant risks, especially if the data fed into these systems is manipulated. When this happens, the outputs can become unreliable or biased. Examples of integrity risks include:

- Altering training or input data into AI systems that affect government decision-making. This can lead to poor allocation of resources, biased hiring practices, unfair tender outcomes, inaccurate policy analysis, and other unintended consequences.
- Businesses applying for grants with false AI-generated documentation, which can enable fraudulent access to government funding.
- Manipulating algorithms used for law enforcement or social support systems, which may compromise the fairness of legal outcomes or eligibility determinations.

AI systems can be applied or misused in a way that creates integrity risks

Many AI systems can be applied or misused in ways that risk integrity within the public service. Examples of integrity risks include:

- Inputting sensitive personal information into AI systems, which can damage public trust in the ability of government organisations to maintain privacy.
- Inputting sensitive commercial or cabinet-in-confidence information into generative AI systems, which can damage the integrity of government decision making process.
- Using AI to reduce workload, which can escalate into poor quality or incorrect advice, eroding trust in the advice given by the public sector.
- Using AI to save time, which can create risks of timesheet fraud if public servants are not honest about how they are using their time.

What can public sector organisations do?

When working with or designing AI systems, public sector organisations should:

- Consider blocking AI systems that are not appropriate for use.
- Introduce pop-up messages to alert staff to the risk of inputting information or data into AI systems.
- Provide clear guidelines to public servants on how to use AI and identify tools approved for specified purposes.
- Provide training to public servants on the ethical use of AI, risks of the misuse of AI in their organisations and relevant guidelines from the Office of the Victorian Information Commissioner.
- Ensure sensitive information is protected and only used in AI systems if confidentiality and security can be maintained.
- Create guidelines for how public servants disclose the use of AI in their work.

This product was prepared based on findings from desktop research and stakeholder consultations from IBAC's 2024 Public Sector Strategic Assessment. All information contained in this document should not be considered as evidence for, or accusations of, corruption.

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**

If you have difficulty accessing the online form, call us on **1300 735 135** 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**

May 2025

EDUCATION AND PREVENTION

Strengthening Victoria's Integrity

ibac