



# AI use in Managing Public Interest Disclosures

Risks and Mitigation Considerations for  
Designated Officers



**Public Interest  
Commissioner**  
*of Alberta*

# Educational Purpose

**This presentation is intended for designated officers.**

It is provided for educational purposes to:

- › Demonstrate potential AI use cases
- › Highlight associated risks and considerations
- › Support informed decision-making about AI use in the disclosure process





# Overview

- › Data Sovereignty
- › Confidentiality
- › Information Integrity and Reliability
- › Procedural Fairness
- › Legal/Privacy Risks
- › Employee and Public Trust
- › Mitigation Measures



**Public Interest  
Commissioner**

# Data Sovereignty

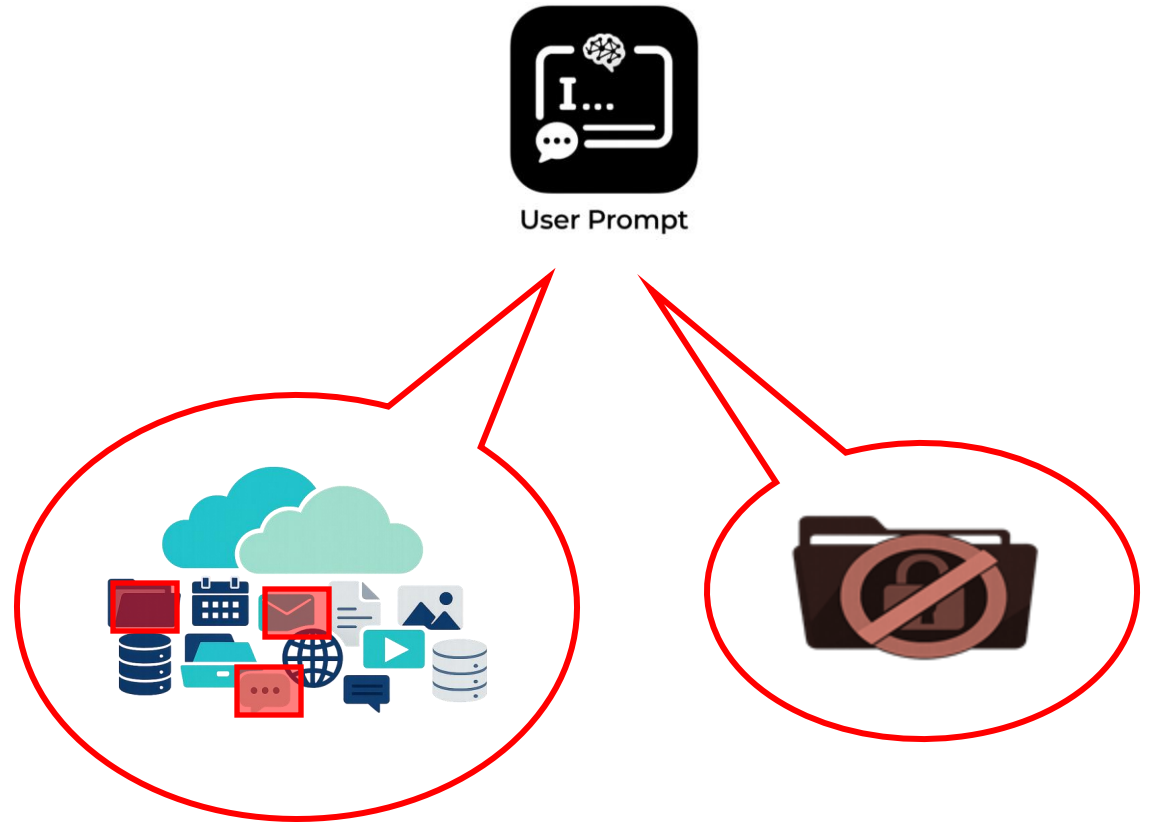


- › Investigations under the *Public Interest Disclosure (Whistleblower Protection) Act* (the Act) involve highly sensitive information that, if exposed, can cause harm to an organization and the individuals involved.
- › Although Cloud-based AI service providers may offer data residency in Canada, AI tools may transmit and process data internationally. The data then becomes subject to the laws in that region.
- › Policy for the responsible use of AI is essential to staff using unauthorized AI tools who risk inadvertently exposing confidential investigative and sensitive materials externally. Free and consumer (i.e.; non-enterprise /government tier systems) use and learn from data that is uploaded.

# Confidentiality

## Access to Confidential Public Interest Disclosure Records

- › A confidentiality risk arises if public interest disclosure records are stored in locations where permissions and access controls are not set. Use of AI tools could give unauthorized persons access to confidential public interest disclosure information.



# Confidentiality

## Data Aggregation

- › AI systems access and combine many documents at once, infer relationships, timelines, and identities, then produce synthesized narratives the user may never have constructed.
- › There is a risk that unauthorized inference may identify parties, including a whistleblower.



## Information Integrity and Reliability



The reliability and integrity of the information obtained and used when managing public interest disclosures is critical as decisions, findings, and outcomes must be **accurate, fair, and defensible**.

Investigations often involve serious matters of public interest and may have serious impacts on individuals' rights, reputations, and employment. If information is incomplete, distorted, misattributed, or unreliable, it can lead to flawed conclusions, procedural unfairness, loss of trust in the process, and legal or regulatory challenges.

Maintaining high information integrity ensures that findings are based on verifiable evidence, withstand scrutiny, and uphold the credibility of the public interest disclosure regime.

# Information Integrity and Reliability



## › Hallucinations and Fabrication Risk

AI tools may generate information that is very plausible sounding but false, including invented facts, relationships, and interpretations

## › Traceability of Evidence

AI may produce synthesized outputs without clear attribution to the source

## › Bias in AI Systems

AI relying on biased data can influence perceptions and undermine neutrality

## › Outdated Information

Data aggregation can result in use of outdated information

## › Exclusion of Key Information

Relying on AI to determine relevance can result in key evidence being missed or excluded

# Procedural Fairness

Procedural fairness is a foundational requirement in public interest disclosure investigations.

Investigations must be fair, impartial, evidence-based, and defensible.

AI tools operate in a 'black box' where the logic and reasoning is unclear.



Public Interest  
Commissioner



# Procedural Fairness

## Incomplete Evidence

**Risk:** AI tools may summarize, cluster, or surface information based on statistical relevance rather than legal relevance.

## Bias and Assumptions

**Risk:** AI systems are trained on large datasets that may have embedded cultural, organizational, or linguistic biases.

## Improper Decision-Making

**Risk:** If outputs influence conclusions, AI moves from an assistant to a decision-maker.

## Inability to Challenge AI Outputs

**Risk:** Parties may not be able to meaningfully review and respond to AI-generated content, creating an unfair process.

## Record-Keeping and Reviewability

**Risk:** AI prompts and outputs may not be retained or reproducible, making it difficult to withstand oversight or judicial scrutiny.

# Legal/Privacy Risks



## *Protection of Privacy Act (POPA)*

- › Public bodies must explicitly notify users if collected personal information will be input into an “automated system to generate content or make decisions, recommendations or predictions” (*Protection of Privacy Act (Alberta)*, s. 6(2)(d)).
- › Notice must be given at time of collection.
- › Every reasonable effort must be taken to ensure that the information is accurate and complete.
- › Information must be retained for at least one year after using it.
- › A Privacy Impact Assessment may also be required if there is a significant change in an administrative practice that will involve the collection, use, and disclosure of personal information where the loss could result in a real risk of significant harm.



**Public Interest  
Commissioner**

## Legal/Privacy Risks



## Judicial Review

- › Decisions of designated officers may be reviewed by the courts.
- › Case law related to AI use during administrative investigations is limited.
- › The courts have not decided that AI may be a substitute decision-maker.
- › Decisions must be justified, transparent, and intelligible – decision-makers must show how and why conclusions were reached.

# Legal/Privacy Risks



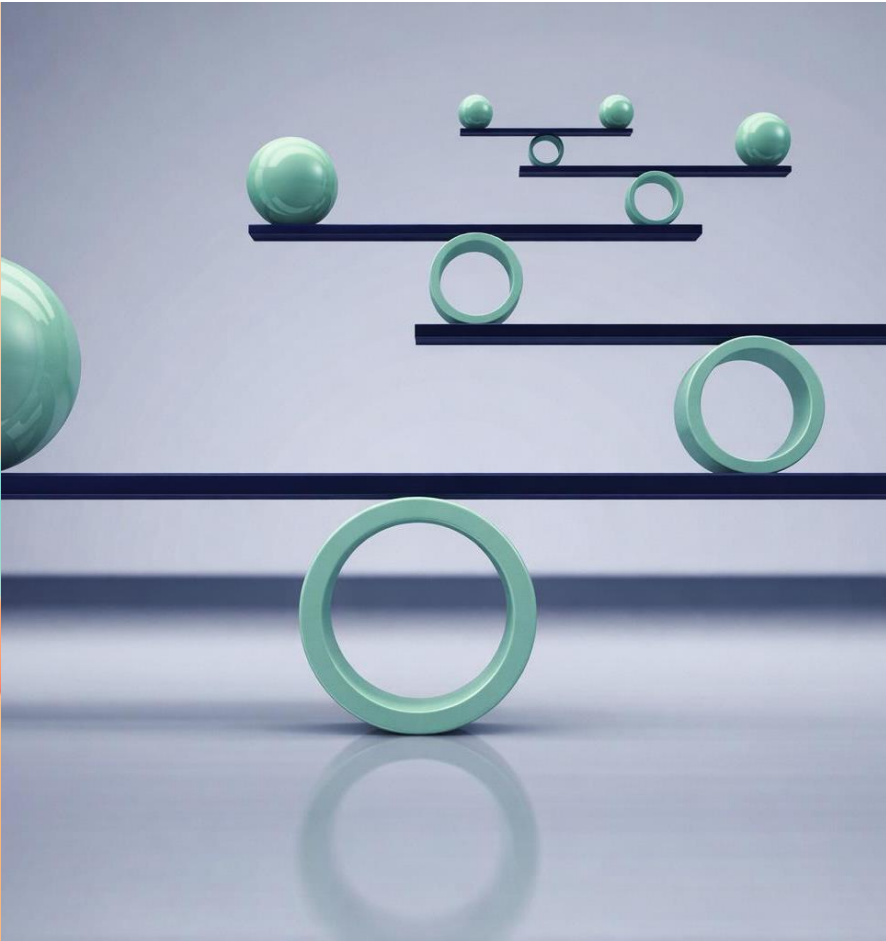
- › AI tools should never be a substitute for a lawyer.
- › AI tools fundamentally lack the accountability, ethics, and strategic judgment required for the practice of law.
- › AI tools can provide faulty legal guidance that can result in significant legal risk.

# Employee and Public Trust

Investigations must be seen as being independent, free from bias, accurate and fair

- › AI systems are trained on organizational data and historical records using common language patterns. This risks investigative outcomes being seen as influenced by institutional biases.
- › Over-reliance on AI tools may be perceived as sensitive allegations being reduced to automated processes. This may result in a loss of trust that investigations are the result of the independent judgment of investigators.
- › AI operates in black boxes, where logic and sources are not easily explainable, making it difficult for stakeholders to consider the investigation outcomes accurate or the process fair.

# Mitigation Measures



- › **AI Responsible Use Frameworks:** Understand and follow your organization's policies and frameworks for responsible AI use.
- › **Data Sovereignty:** Use only enterprise-approved AI tools with confirmed Canadian data residency and contractual data controls.
- › **Confidentiality:** Do not input identifiable whistleblower or investigation information into web-based AI tools. Get expert advice on segregating public interest disclosure records from your environment and restrict AI access.
- › **Information Integrity and Reliability:** Treat AI outputs as unverified; independently confirm all facts against sources and original evidence.
- › **Procedural Fairness:** Do not rely on AI for credibility assessments, findings, or recommendations; ensure decisions are explainable and reviewable.
- › **Legal and Privacy Risks:** Ensure compliance with POPIA requirements prior to AI use as part of your processes. Use experts for legal advice.
- › **Employee and Public Trust:** Be transparent about AI use and reinforce that human judgment governs all investigative decisions.

## Contact us

1-855-641-8659  
[info@pic.alberta.ca](mailto:info@pic.alberta.ca)



**Public Interest  
Commissioner**