Submission
No 36

INQUIRY INTO ARTIFICIAL INTELLIGENCE (AI) IN NEW SOUTH WALES

Organisation: NSW Independent Commission Against Corruption

Date Received: 25 October 2023

I·C·A·C

INDEPENDENT COMMISSION AGAINST CORRUPTION

NEW SOUTH WALES

SUBMISSION

TO THE

LEGISLATIVE COUNCIL INQUIRY INTO

ARTIFICIAL INTELLIGENCE IN NEW SOUTH WALES

PORTFOLIO COMMITTEE No.1 – PREMIER AND FINANCE

October 2023

Introduction

The Independent Commission Against Corruption ("the Commission") welcomes the opportunity to make a submission to the Committee's Inquiry into Artificial Intelligence ("AI") in New South Wales.

This submission is focused primarily on items 1(a), (g) and (h) of the Inquiry's terms of reference.

This submission is not intended to constitute legal advice and may not represent the Commission's concluded view on relevant matters that remain under consideration.

About the Commission

The Commission was established as an independent body in 1988 to investigate and prevent corruption in and affecting the NSW public sector.

Corrupt conduct can take many forms and is commonly thought of as the abuse of entrusted power for private gain.[1] However, under the *Independent Commission Against Corruption Act 1988* ("the ICAC Act"), corrupt conduct is defined more broadly. It may involve conduct, such as fraud and blackmail, by any person affecting the exercise of public official functions, or conduct that could impair public confidence in public administration for a range of matters.[2]

The Commission's principal functions are set out in s 13 of the ICAC Act. In summary, s 13 provides that the Commission:

- as part of an investigation process, consider whether laws, methods of work, practices or procedures should be changed to reduce the likelihood of the occurrence of corrupt conduct
- examine laws, practices and procedures of public authorities to facilitate the discovery of corrupt conduct
- instruct, advise, and assist public authorities on ways in which corrupt conduct may be eliminated, the likelihood of corrupt conduct reduced, and the integrity and good repute of public administration promoted
- enlist and foster public support in combating corrupt conduct and in promoting the integrity and good repute of public administration.

Summary

The Commission is not an expert in AI and notes that it has not yet investigated any matters where sophisticated AI has been used to plan or execute corrupt conduct. However, as set out below, there are reasons to suspect that AI will play a role in the Commission's future work.

AI has the potential to increase the scale and reach of corruption, fraud and misinformation that could undermine confidence in the integrity of public institutions. It could also make it more difficult for such conduct to be detected, investigated and prosecuted by NSW public bodies. Technological systems that mimic human behaviours, and that are also automated and capable

[1] Transparency International, "What is corruption?", https://www.transparency.org/en/what-is-corruption

[2] Refer to s 8(2) of the ICAC Act in relation to fraud and blackmail and s 8(2A) relating to conduct that impairs, or that could impair public confidence in public administration involving a range of matters including collusive tendering, dishonestly obtaining public funds for private advantage, or fraudulent application for licences.

of autonomy, are of serious concern and will likely challenge the anti-corruption capabilities of public institutions, including those of the Commission. This is because advanced technology can offer a dangerous combination of opacity, anonymity, psychological distance, speed and efficiency, and personalisation not seen before.[3]

On the other hand, AI can also assist corruption prevention efforts, but this is dependent on an agencies' capability, capacity, and priority to capture relevant data and utilise it for corruption detection purposes. The NSW Public Sector Treasury Circular TC18-02, the Australian Standard on *Fraud and Corruption Control* (AS 8001-2021) and the Audit Office's 2015 *Fraud Control Improvement Kit* all recognise that organisations should apply data analytic techniques in the detection of fraud and corruption. AI systems generating "content, forecasts, recommendations, or decisions for a given set of human-defined objectives" is the technological means for carrying out data analytics.[4]

The Commission utilises intelligent technology in its investigations to review vast data sets and to enrich data, for example for translation and transcription. The Commission is acquiring increasing amounts of data relevant to its investigations which must be analysed. The Commission's investigations continue to be complex, with significant reliance on financial analysis and computer forensics.

Potential integrity-related benefits from AI

AI provides government with opportunities to prevent, detect and investigate fraud and corruption.[5] It may also play a role in promoting integrity, creating transparency and fostering accountability.[6]

In simple terms, corrupt conduct entails the abuse of official discretion. Accordingly, to the extent that AI can successfully remove or limit the way humans wield discretion, the opportunities for corruption are also limited. That is, subjective (and thus potentially corrupt) human decision-making is replaced (or assisted) by disinterested AI. The possible applications are yet to be fully contemplated but it is not difficult to imagine AI being deployed across a number of areas that are known to involve corrupt conduct such as procurement, recruitment, payroll and regulatory decision-making.

[3] TF Blauth, OJ Gstrein & A Zwitter, "Artificial Intelligence Crime: An overview of Malicious Use and Abuse of AI"*IEEE Access* Volume 10, 2022 at p.77115; N Kobis, JF Bennefon, I Rahwan, "Bad machines corrupt good morals" *Nature Human Behaviour*, 2021, 5(6), pp.679-685.

[4] ISO/IEC 22989 defines an artificial intelligence system as an 'engineered system that generates outputs such as content, forecasts, recommendations or decisions for a given set of human-defined objectives'. The engineered system can use various techniques and approaches related to artificial intelligence to develop a model, to represent data, knowledge, processes, etc. which can be used to conduct tasks.

[5] C Santiso, "Here's how technology is changing the corruption game" *World Economic Forum*, 28 February 2019 https://www.weforum.org/agenda/2019/02/here-s-how-technology-is-changing-the-corruption-game/. Accessed 13 October 2023; P Aarvik, "Artificial intelligence – a promising anti-corruption tool in development settings?", U4 Anti-corruption Resource Centre, 2019; The World Bank, *AI in the public sector Maximising Opportunities Managing Risks, Equitable Growth, Finance and Institutions Insight*, Washington DC, 2020 https://openknowledge.worldbank.org/handle/10986/35317. Accessed 13 October 2023.

[6] F Odilla,"Avoiding minority reports: using AI responsibly in anti-corruption" *Corruption Justice and Legitimacy,* 23 October 2022, https://www.corruptionjusticeandlegitimacy.org/post/avoiding-minority-reports-using-ai-responsibly-in-anti-corruption. Accessed 13 October 2022

In addition, AI tools can potentially:

- facilitate the provision of information or services to the community, and the flow of information from the community to government[7]
- facilitate and, in some cases, automate, transparency portals of government-released information or public databases[8]
- proactively detect corruption risks allowing treatment before significant harm arises.[9]

The following overseas examples show the opportunities to use AI for anti-corruption purposes.[10]

Public procurement – Brazil's Governance Risk Assessment System

The World Bank Team in Brazil developed an AI system that identifies red flags of potential fraud in public procurement processes. The AI system analyses public procurement from 12 states, a number of cities and government departments. The system involves one of the world's largest data lakes, made up of 27 datasets with over 250 million data points and approximately US$100 billion in public procurement expenditure.

The tool has identified:

- more than 500 firms owned by public servants working at the same procuring government agency
- 450 firms with partners receiving social welfare (indicating these are possibly strawmen and not genuine tenderers)
- 857 companies that won tenders against companies with at least one partner in common (indicating possible bid rigging), 30,302 different bidders with the same phone number, 2,882 different bidders with the same address, and 154,911 different bids from the same IP address (indicative of non-genuine bids).

Claims fraud – US Centres for Medicare and Medicaid Services

The Healthcare Fraud Prevention Partnership is a public-private partnership with 79 participants. The key strategy was to move beyond the 'pay and chase' approach of recovering improper payments and focus on prevention by suspending improper and potentially fraudulent claims for further analysis. It pools data to analyse multiple claims to identify health service providers with patterns of suspect billing, for example, for more

[7] I Adam & M Fazekas, "Are emerging technologies helping win the fight against corruption? A review of the state of evidence", *Information Economics and Policy,* vol 57, December 2021 https://doi.org/10.1016/j.infoecopol.2021.100950.

[8] F Odilla,"Avoiding minority reports: using AI responsibly in anti-corruption" *Corruption Justice and Legitimacy,* 23 October 2022, https://www.corruptionjusticeandlegitimacy.org/post/avoiding-minority-reports-using-ai-responsibly-in-anti-corruption.). Accessed 13 October 2022; Open Data Charter, *Open Up Guide: Using open data to combat corruption* (2017) https://www.open-contracting.org/resources/open-guide-using-open-data-combat-corruption. Accessed 13 October 2022.

[9] A Petheram, W Pasquarelli & R Stirling, *The Next Generation of Anticorruption Tools: Big Data, Open Data & Artificial Intelligence Research Report,* Oxford Insights, May 2019, https://www.oxfordinsights.com/ai-for-anti-corruption

[10] Numerous case studies are set out in the U4 Report, *AI a promising anticorruption tool in development settings* (2019) and The World Bank, *AI in the public sector Maximising Opportunities Managing Risks* (2020).

> services than could reasonably be rendered in a single day. The Fraud Prevention System identifies and triages suspect health service providers for investigation and is a method to prevent fraud, waste and abuse.
>
> In the period 2011 to 2015, the fraud prevention system has helped to identify suspected improper claims, and thereby save nearly $1.5 billion which would otherwise have been paid.

The idea that AI can be used to sift through vast amounts of data to detect fraudulent transactions is extremely attractive (see box below "Data analytics – an area for improvement").

However, as was seen in the case of Australia's Robodebt Scheme, automation of public administration can fail because of human and technological error, bias and lack of oversight. The Commission is not in a position to commentate on the various ways in which AI can over-promise and under-deliver but it is self-evident that a degree of caution is required.

Data analytics – an area for improvement

The Commonwealth *Crimes Legislation Amendment (Powers, Offices and Other Measures) Act 2018* ("CLAPOOM") authorises the collection, use and disclosure of personal information for an 'integrity purpose'.[11] An integrity purpose includes preventing, detecting, investigating or dealing with serious misconduct by persons working for Commonwealth bodies, fraud affecting Commonwealth bodies, or offences against Chapter 7 of the *Criminal Code Act 1995* relating to the proper administration of government. There is no equivalent law in NSW.

The Explanatory Memorandum to the CLAPOOM Bill noted that the measures in the Bill would assist Commonwealth entitles to prevent, investigate and deal with fraud against the Commonwealth and would have a positive financial impact, improve the provision of government services, and improve public safety.[12] The new provisions reduce the complexity of investigating or otherwise controlling fraud against the Commonwealth to help increase recoveries and prevent fraud occurring.

Importantly, CLAPOOM also provides for the preparation of guidelines (in consultation with the Commonwealth Information Commissioner) that seek to facilitate the pursuit of integrity objectives without misusing personal information.

The Commission's experience is that relatively few NSW public agencies use data analytics as a key component of their corruption control framework. Reform along the lines of CLAPOOM would assist.

AI technology and deepfakes

It is well-known that AI technology can be used to create authentic 'deepfakes' of the appearance, voice and body language of an individual. Techniques for impersonating or

[11] CLAPOOM commenced on 25 August 2018. It inserted Part VIID – 'Collecting, using and disclosing personal information that may be relevant for integrity purposes' into the *Crimes Act 1914* (Cth).

[12] https://parlinfo.aph.gov.au/parlInfo/search/display/display.w3p;query=Id%3A%22legislation%2Fems%2Fr5838_ems_296f17bb-b3ce-4598-be58-a1462095e300%22. Accessed 13 October 2023.

deceiving others have existed throughout recorded history, so in one sense, deepfakes are just the latest iteration of long-running pattern. However, the rapid improvements in AI technology appear have created a novel threat because they can convincingly impersonate public figures. Deepfake images and videos may not leave forensic traces in edited digital media, making them difficult for humans or even machines to detect.[13] Consequently, deepfakes could be used to:

- carry out more sophisticated forms of phishing scams and cybercrime[14] [15] [16] (discussed in more detail below).
- embarrass, undermine or even blackmail individuals, including senior public officials
- spread disinformation, thus making it difficult for citizens to differentiate between truthful and untruthful information. In turn, this could adversely affect the operation of public agencies and democratic institutions.

It is also foreseeable that an autonomous AI system could itself engage in wrongdoing.[17] Technology that can mimic human behaviours and characteristics (voice, appearance, language), but are also automated and capable of autonomy, could initiate actions that, if carried out by a human, would be regarded as unethical or fraudulent. Such harms could seriously challenge the anti-corruption and integrity capabilities of public institutions, including the Commission, and the justice system. In part, this is because the Commission can only make findings of corrupt conduct about an individual. If the harms caused by AI cannot be traced back to a human, investigative agencies such as the Commission may not be in a position to make adverse findings.

Similarly, if there is doubt about whether a fraudulent document has been authored or used by a human, it will be difficult for an investigator to attribute culpability.

AI-enhanced cybercrime

Public authorities hold valuable information about people, places, processes, and systems that make them a lucrative target. Members of the Committee will be aware that many cybercrimes are carried out via social engineering schemes. Cybercrime perpetrated through social engineering has increased alarmingly within a short period of time.[18] Nearly one in five data

[13] M Caldwell, JTA Andrews, T Tanay & LD Griffin, "AI-enabled future crime" *Crime Science* vol 9 (2020) at 14; M Brundage et al*, The Malicious Use of Artificial Intelligence: Forecasting, Prevention and Mitigation,* ArXiv abs/1802.07228 (2018); O Bendel, "The synthesization of human voices", *AI & Soc*, 34, 83-89 (2019).

[14] KPMG *Generative AI models – the risks and potential rewards in business* (2023) p.11

[15] C Stupp, "Fraudsters Used AI to Mimic CEO's Voice in Unusual Cybercrime Case", *Wall Street Journal*, 30 August 2019, https://www.wsj.com/articles/fraudsters-use-ai-to-mimic-ceos-voice-in-unusual-cybercrime-case-11567157402

[16] T Brewster, "Fraudsters Cloned Company Director's Voice In $35 Million Heist, Police Find" *Forbes,* 14 October 2021, https://www.forbes.com/sites/thomasbrewster/2021/10/14/huge-bank-fraud-uses-deep-fake-voice-tech-to-steal-millions/?sh=d72c74975591

[17] Examples include learning-based cyber attacks or AI based monitoring of persons. M Caldwell, JTA Andrews, T Tanay, LD Griffin, "AI-enabled future crime", *Crime Science* 9, 14(2020) https://doi.org/10.1186/s40163-020-00123-8

[18] https://www.scamwatch.gov.au/research-and-resources/scam-statistics. Accessed 13 October 2023.

breaches in the first half of 2023 were caused by social engineering or impersonation.[19] A number of public agencies have been adversely affected.

Simply put, AI makes it easier to construct a successful social engineering scheme. As noted above, this could involve the use of deepfakes but other schemes could benefit from AI.

A social engineering scheme requires the criminal to put effort into identifying and profiling their targets and crafting an approach that is most likely to deceive those targets. AI can potentially automate this process.

Arguably, AI can also make it easier for cybercriminals to operate across national borders and rapidly transfer data and funds. To a large extent, physical borders have become inconsequential with regard to technology, as individuals using sophisticated tools can move efficiently through them instantaneously.[20]

Australian governments have already committed significant funds to combatting cybercrime and the Commission is not in a position to comment on whether this represents an adequate response to the threat. However, dealing with AI-driven cybercrime is likely to consume an increasing proportion of public sector budgets.

Exploitation of AI by public officials

The Commission has investigated public officials who have tampered with ICT systems to engage in, or cover up, corrupt conduct. Consequently, the idea that a bad actor could deliberately create or manipulate AI to pursue a corrupt purpose is not far-fetched. A sophisticated individual could take advantage of AI's vulnerabilities by poisoning data and algorithms, manipulating models, introducing security backdoors or 'BadNet' coding, or otherwise manipulate the trustworthiness of system outputs.[21]

Vulnerabilities in large language models, for example, when used for recruitment decisions is prone to exploitation: "the computer's understanding of language is statistical, not semantical".[22] Individuals can also take advantage of (or even sell) information about known vulnerabilities of a model for personal gain.[23] Furthermore, the opacity of an algorithmic system may render such conduct problematic to detect and investigate. [24] In particular, outputs produced by AI technology may be difficult to audit.

[19] https://www.oaic.gov.au/privacy/notifiable-data-breaches/notifiable-data-breaches-publications/notifiable-data-breaches-report-january-to-june-2023#number-of-individuals-worldwide-affected-by-breaches . Accessed 13 October 2023.

[20] M Smith & G Urbas, *Technology Law*, 2021 Cambridge p.19

[21] TF Blaugh, OJ Gstrein and A Zwitter, "Artificial Intelligence Crime: An overview of malicious use and abuse of AI", *IEEE Access*, vol 10, pp.77110-77122, https://doi.org/10.1109/ACCESS.2022.3191790

[22] T Walsh, *Machines Behaving Badly: The morality of AI*, La Trobe University Press & Black Inc, Melbourne VIC 2022

[23] J Li, S Ji, T Du, B Li, T Wang, "TextBugger: Generating Adversarial Text Against Real-world Applications", *Network and Distributed Systems Security Symposium 2019,* San Diego 24-27 February 2019, https://arxiv.org/pdf/1812.05271.pdf

[24] K Hayward & M Maas, "Artificial intelligence and crime: A primer for criminologists", *Crime Media Culture an International Journal* 17(2), Sage Journals, June 2021, 209-233 https://doi.org/10.1177/1741659020917434

Deference to AI

Some literature suggests that humans become overly reliant on, or deferential to, decisions made by AI-enabled technology.[25] Some potential consequences include the following:

- A human cannot or will not challenge the AI-produced analysis/decision, which might be flawed
- As reliance on AI builds over time, there will be fewer humans with the skill and experience to perform the mental tasks now carried out by AI. As a result, the AI becomes difficult to replace.
- Humans may effectively outsource the execution of ethically risky or questionable behaviour. Kobis et al state that “*AI agents offer the dangerous trifecta of opacity, anonymity and social distance that enables people to psychologically dissociate themselves from the unethical act*”.[26] This may be the case even when an individual recognises that an AI system recommends they break ethical rules. That is, the deference to advanced technology, may blind a person to action that is highly unethical.

A related problem stems from the “black box” nature of AI. The Commission understands that in practice, it may very difficult to obtain a detailed explanation of how sophisticated AI technology has made a particular decision. Crootof et al state “*But algorithmic systems often cannot supply satisfying reasons for their determinations; indeed, it is sometimes impossible even for those who design or regularly use certain algorithms to explain how they reach conclusions*” and may be “. . . *literally uninterrogable by human agents*”.[27]

AI and government transactions

A significant proportion of the matters examined by the Commission involve allegations that applications, submissions and offers made to public agencies are fraudulent or contain misstatements. This includes areas such as:

- procurement, including false invoicing, collusive tendering, order splitting and the use of fictitious suppliers
- recruitment, including CV fraud
- grant applications, including the false claims about eligibility for funding
- development applications and building certificates
- applications for business licences
- failures to disclose conflicts of interest or other information required.

As is the case with social engineering (mentioned above), submitting a false document to a public sector agency requires effort and judgement. This can be a difficult task for a human and corruption investigation often entails identifying red flags in the relevant documentation.

Unfortunately, AI has the potential to automate the production of false or misleading documents that are more likely to deceive a public official. For example, it is highly likely that an

[25] For example, see R Crootof, M Kaminski and WN Price, “Humans in the Loop”, *Vanderbilt Law Review*, V 23:2, pp. 429-510.

[26] M Leib, N Kobis, RM Rilke, M Hagens, B Irlenbusch “Corruptive force of AI-generated advice”, *arXiv Cornell University* 2021. https://www.researchgate.net/publication/349335637_The_corruptive_force_of_AI-generated_advice

[27] R Crootof, M Kaminski and WN Price, “Humans in the Loop”, *Vanderbilt Law Review*, V 23:2, p. 479.

AI technology will become better than a human at producing an authentic looking, but still false invoice.

The Commission anticipates AI could be used deliberately to produce a false document (for example, a job applicant asks AI to create a false CV). In other situations, the AI may produce a false statement without specifically being asked to do so (for example, the job applicant asks AI to polish their CV to improve their chances of getting a job).

This poses various questions relating to intent (did the individual deliberately provide false information?), foreseeability (should the individual have known that the responses generated by AI are likely to be false?), and liability (to what extent is that individual culpable?). These issues are likely to play a role as to whether dishonest practices such as collusion, price fixing or anti-competitive conduct have occurred.[28]

If harm becomes apparent, the black box nature of algorithms can increase plausible deniability, victims can become psychologically abstract and removed from the conduct, and responsibility can be deflected to the technology.[29]

Democracy and public discourse

One of the Commission's principal functions is to promote the integrity and good repute of public administration. This involves ensuring public institutions remain trustworthy, and that trust in public institutions is not undermined.

It is well known that AI can accelerate the spread of rumour, conjecture, inadvertent misinformation and deliberate disinformation.[30] Automated 'social bots', non-genuine accounts created with fake identities (for example with Generative AI tools), and the analysis of big data for targeted approaches to influence outcomes, can undermine public confidence in policy making by manipulation of the public. Large language model (LLM) generated content can be misused in democratic processes such as public policy consultations to mislead public opinion and can drive the misallocation of resources and misinform election outcomes.[31] The spread of mis- and disinformation about government can undermine the public's trust in government decision-making, its policies, administration and ultimately undermine public service.

Research indicates that for online platforms, false rumours spreads significantly farther, faster, deeper and more broadly than the truth, and the effects are more pronounced for false political news. One reason is that false news is often more novel and more likely to be shared, than

[28] K Hayward & M Maas, "Artificial intelligence and crime: A primer for criminologists", *Crime Media Culture an International Journal* 17(2), Sage Journals, June 2021, 209-233 https://doi.org/10.1177/1741659020917434

[29] T Miller, "Explanation in artificial intelligence: Insights from the social sciences". *Artif. Intell. 267,*1-36 (2019); D Gunning, M Stefik, J Choi and T Miller, S Stumpf, GZ Yang, "XAI – Explainable artificial intelligence" *Science Robotics* 4 (2019); J Dana, RA Weber, & JX Kuang, "Exploiting moral wiggle room: experiments demonstrating an illusory preference for fairness". *Economic Theory* 33, 67-80 (2007); JT Hancock & J Guillory, "Deception with technology", *The Handbook of the Psychology of Communication Technology* (ed. Sundar, S.S) 1 January 2015, 270-289; N Kobis, JF Bonnefon, I Rahwan, "Bad machines corrupt good morals", *Nature Human Behaviour*, 2021, 5(6), pp.679-685.

[30] T Graham, A Bruns, G Zhu, R Campbell, "Like a virus, The coordinated spread of coronavirus disinformation", *Australia Institute & Centre for Responsible Technology*, 1 June 2020, at p.5

[31] G Bell, J Burgess, J Thomas, S Sadiq, "Rapid response information report: Generative AI – language models (LLMs) and multimodal foundational models (MFMs)", *Australian Council of Learned Academics at* p.12

truthful information.[32] Mainstream media, politicians and public institutions and other 'influencers' are at risk of increasing the visibility of fake news by engaging with disinformation, even when done critically, in contexts where denials can be misinterpreted as a conspiracy to hide the real truth.[33]

Technology systems risks and outsourcing

The exploitation of AI can be exacerbated by vulnerabilities in supply chains. It is envisaged that NSW government entities will be consumers rather than developers of AI. This can also mean that training procedures or pre-trained models are outsourced.[34] Where a contractor is engaged to assist in designing, building or deploying an AI solution, this introduces risks that require management to ensure public benefit outcomes. The Commission has investigated ICT technology related matters where managers have deferred to the expertise of contractors and consultants without establishing their trustworthiness.

The ARC Centre of Excellence in Automated Decision Making and Society (ADM+S) identified evidence in the Royal Commission into the Robodebt Scheme which demonstrated the challenges of identifying who is responsible for ensuring the legal operation of the ADM because of multiple and separate actors. [35] In its submission, the ADM+S recommended that at the inception, design and procurement stages of developing and deploying ADM systems there should be clarity about responsibility for a system and a clearly articulated map and chain of accountability in addition to algorithmic auditing, allowing appeals not just of individual decisions but of the system itself.

One other important mechanism for algorithmic accountability is the ability for the public to access the necessary data and algorithms upon which a decision is based. Section 121 of the *Government Information (Public Access) Act 2009* (NSW) requires that all contracts with external parties ensure that the government agency have immediate right of access relating to the performance of services by the contractor. The NSW Information Commissioner has advised:

> *A feature of the shift to digital decision-making and service delivery is the increase in third-party providers and outsourcing arrangements made between government agencies and contractors. However, outsourcing should not obstruct a citizen's right to access data that concerns decisions made by government agencies as a result of this data.*[36]

Section 121(2)(c) contains an exemption to information access. A contractor is not required to provide algorithmic transparency, if disclosing that information "*could reasonably be expected to place the contractor at a substantial commercial disadvantage whether at present or in the*

[32] S Vosoughi, D Roy and S Aral, "The spread of true and false news online", *Science* 359 1146-1151 (2018)

[33] T Graham, A Bruns, G Zhu, R Campbell, "Like a virus, The coordinated spread of coronavirus disinformation", *Australia Institute & Centre for Responsible Technology*, 1 June 2020 at p.21 & p.24

[34] K Hayward & M Maas, "Artificial intelligence and crime: A primer for criminologists", *Crime Media Culture an International Journal* 17(2), Sage Journals, June 2021, 209-233, https://doi.org/10.1177/1741659020917434

[35] ARC Centre of Excellence for Automated Decision-Making and Society, *Submission to the Royal Commission into the Robodebt Scheme* 10 February 2023 at p.13, https://robodebt.royalcommission.gov.au/publications/ano999900010028

[36] Information and Privacy Commission, GIPA Case Studies, https://www.ipc.nsw.gov.au/information-access/gipa-case-studies. Accessed 13 October 2023.

future".[37] Clause 4(d) from the Table to s14 of the GIPA Act contains a public interest consideration against disclosure on the basis that disclosure could "prejudice any persons' legitimate business, commercial, professional or financial interests". Both provisions are potentially inconsistent with NSW's AI Ethical Principle of transparency for the use of AI and the National AI Ethics Principle of contestability.

The NSW AI Ethical Principles notes under the principle of transparency "*Review mechanisms will ensure citizens can question and challenge AI based outcomes…Projects should demonstrate how the public can question and seek reviews of AI based decisions, how the community can get insights into data use and methodology…*".[38] The National AI Ethics Principle of contestability states that "*When an AI system significantly impacts a person, community, or group or environment there should be a timely process to allow people to challenge the use or outcomes of the AI system…There should be sufficient access to the information available to the algorithm and inferences drawn, to make contestability effective*" [39].

The Information Commissioner has highlighted the importance of preserving access rights where an agency enters into a contract with a private sector entity and the potential for curtailment of a citizen's right to access government information. The Information Commissioner has advised that "*as government increasingly adopts digital technology, it has a duty to implement practices that safeguard legislated commitment to open government and fundamental rights of access to information*" which takes into consideration information used to develop and support digital solutions including algorithms, source code, test suites, data sets and variables.[40]

Public officials should demonstrate their uses of AI generate fair, accurate and consistent results. That endeavour will require sharing with the public explainable information about data inputs, the algorithm and how outputs are reached.[41] Providing adequate transparency may prove challenging when AI tools used by agencies are created by outside contractors. It has been noted that "*When faced with requests for transparency, companies can invoke trade secret protection to try to preserve the proprietary nature of their algorithm's operations and*

[37] The application of s121 was tested in the matter of *O'Brien v Secretary DCJ* [2022] NSWCATAD 100 which found there was no right of information access as the terms of contract between the agency and contractor was entered into prior to the GIPA Act, and did not provide the agency a right of access to that information. The Tribunal also found that disclosure could also reasonably be expected to place the contractor at a substantial commercial disadvantage in relation to the DCJ such that disclosure could lead to DCJ no longer needing to engage the specialist services of the contractor in carrying out its statutory functions (at paragraph 103)

[38] Digital NSW, "Mandatory Ethical Principles for the use of AI", https://www.digital.nsw.gov.au/policy/artificial-intelligence/artificial-intelligence-ethics-policy/mandatory-ethical-principles. Accessed 13 October 2023.

[39] Department of Industry, Science and Resources, "Australia's AI Ethics Principles" https://www.industry.gov.au/publications/australias-artificial-intelligence-ethics-framework/australias-ai-ethics-principles. Accessed 13 October 2023.

[40] Information and Privacy Commission, "IPC Case Notes", https://www.ipc.nsw.gov.au/information-access/agencies/case-notes; "Fact Sheet - Automated decision-making, digital government and preserving information access rights – for agencies", https://www.ipc.nsw.gov.au/fact-sheet-automated-decision-making-digital-government-and-preserving-information-access-rights-agencies#_ftn13. Accessed 13 October 2023.

[41] LM Ben Dor and C Coglianese, "The Procurement Path to AI Governance", *The Regulatory Review*, A Publication of the Penn Program on Regulation, 27 June 2022; also M Hickok, "Public procurement of artificial intelligence systems: new risks and future proofing", *AI & Soc* (2022).

underlying data. Failing to push back on that secrecy can pose legal as well as ethical quandaries for government".[42] One path forward is to "elevate procurement standards and contractual arrangements" to ensure principles of open government are met.[43]

AI and the Commission's investigative and intelligence functions

Almost all modern law enforcement and investigative bodies, including the Commission, utilise technology platforms for obtaining, storing and analysing electronic evidence.[44] Sophisticated solutions are also available for case management and intelligence.

As is apparent from published investigation reports and annual reports, the volume of electronic evidence the Commission examines has increased significantly. This is partly driven by the number and complexity of investigations undertaken but another key factor is the number of computers and smart devices the Commission must acquire in each investigation in order to obtain all of the relevant electronic evidence.

This means that the Commission often acquires more evidence that an individual investigator (or even a team) can possibly analyse.

The recent *Law Enforcement (Powers and Responsibilities) Amendment (Digital Evidence Access Orders) Act 2022* has facilitated access to cloud data for ingestion into the forensic imaging systems at the Commission.[45] This will further expand the volume of evidence available to the Commission in its investigations.

The Commission's investigators currently use AI functionality to review large datasets. The Commission also uses AI tools to enrich data, as a preliminary step for translation and transcription, prior to human verification or proofing.

AI can be used as a tool to enhance intelligence in numerous ways including:

- filtering, sorting and analysing large data sets
- pattern recognition
- forecasting and modelling
- sentiment analysis (in the tone of text)
- detection of anomalies in data
- data integration and multi-source analysis.

The Commission also recognises the risks associated with to utilising advanced technology (including AI) in its investigative and intelligence function. These risks include:

- accuracy of data and analysis
- inability to reference information
- devaluation of human analyst in favour of the perceived superiority if technology
- potential biases in AI algorithms

[42] LM Ben Dor and C Coglianese, "The Procurement Path to AI Governance", *The Regulatory Review*, A Publication of the Penn Program on Regulation, 27 June 2022

[43] Information and Privacy Commission, "Fact Sheet - Guide to section 121 of the GIPA Act for agencies", https://www.ipc.nsw.gov.au/fact-sheet-guide-section-121-gipa-act-agencies. Accessed 13 October 2023.

[44] These are sometimes called "eDiscovery" platforms.

[45] The Commission's powers, including the execution of a digital evidence access order, are operationalised through policies that outlines the requirements for the use of powers under the ICAC Act.

- hallucinations.[46]

In summary, investigative agencies such as the Commission cannot carry out their functions without access to advanced technology. The Commission's 2022-2025 Strategic Plan states that it aims to be "at the forefront of using best practice investigative techniques and digital technologies". Inevitably, this will entail the use of AI but with an eye on the risks described above.

[46] Data hallucination refers to the process of generating or interpreting data in a way that misrepresents reality or factual information. It occurs for various reasons, including biased data collection methods, flawed algorithms or human errors. See D Fallman, "Tackling Data Hallucination: Company Strategies and Industry Insights" *Forbes Technology Council*, 16 October 2023, https://www.forbes.com/sites/forbestechcouncil/2023/10/16/tackling-data-hallucination-company-strategies-and-industry-insights/?sh=1e3041014af1. Accessed 17 October 2023.