MINISTRY OF SOCIAL DEVELOPMENT
TE MANATŪ WHAKAHIATO ORA

17 June 2024

Tēnā koe

Official Information Act request

Thank you for your email of 19 April 2024 to the Ministry of Social Development (the Ministry)'s Media Team, requesting any Ministry advice regarding the use of artificial intelligence (AI) in the workplace.

On 20 May 2024, you were advised that in accordance with section 15(1) and 15A of the Act, the Ministry's decision would be provided by 18 June 2024. You were advised the Ministry needed more time to consult with other parties on the release of information in scope of your request.

You will recall that in response to several questions to the Ministry's Media team asking if and how artificial intelligence is used in the organisation, on 24 April 2024, Hannah Morgan, General Manger Information of the Ministry, provided you with a detailed summary of the Ministry's use of AI for reports, policy, and advice.

You made a further request for:

- *I'd like a copy of any advice given to the agency this year regarding the use of AI in these contexts.*

I have considered your request under the Official Information Act 1982 (the Act). Please find my decision on your request set out below.

The Ministry produced an internal memo to its Organisational Health Committee on 22 November 2023 that is in scope of your request. Attached please find:

- **Appendix:** Organisational Health Committee memo – *Interim position on Ministry use of generative artificial intelligence* dated 22 November 2023.

In releasing the memo to you, please note that some sections are withheld under section 9(2)(h) of the Act in order to maintain legal professional privilege. The greater public interest is in ensuring that government agencies can continue to obtain confidential legal advice.

Some information in the memo is withheld under section 9(2)(g)(i) of the Act to protect the effective conduct of public affairs through the free and frank expression of opinions. I believe the greater public interest is in the ability of individuals to express opinions in the course of their duty.

Note that appendix 2 has been withheld under this section, which was the draft guidance document for staff. The final version was provided to you in the earlier media response.

Note that the information withheld at paragraph 54(a) of the document relates to the Ministry of Education, and its published guidance on generative AI is available here: https://www.education.govt.nz/school/digital-technology/generative-ai/

I will be publishing this decision letter, with your personal details deleted, on the Ministry’s website in due course.

If you wish to discuss this response with us, please feel free to contact OIA_Requests@msd.govt.nz.

If you are not satisfied with my decision on your request regarding internal advice about the use of AI, you have the right to seek an investigation and review by the Ombudsman. Information about how to make a complaint is available at www.ombudsman.parliament.nz or 0800 802 602.

Ngā mihi nui

Magnus O’Neill
General Manager
Ministerial and Executive Services

MINISTRY OF SOCIAL DEVELOPMENT
TE MANATŪ WHAKAHIATO ORA

Memo

To: Organisational Health Committee

From: Hannah Morgan, General Manager Information

Tracy Voice, Group General Manager Improvement Systems and Technology

Prepared by: Alexander Cheeseman, Principal Information Advisor

Kieran O'Callaghan, Principal Information Advisor

Sebastian Lynch, Lead Information Advisor

Date: 22 November 2023

Security level: IN CONFIDENCE

This memo contains legal advice and is legally privileged. It should not be disclosed on an information request, without further legal advice.

Interim position on Ministry use of generative artificial intelligence

Purpose

1. The purpose of this memo is to describe potential opportunities for Ministry adoption of generative artificial intelligence (AI) tools like ChatGPT, the risks we have identified with Ministry use of those tools, and how managing those risks sits within the Ministry's existing Information Policy framework.
2. This memo also seeks approval for an interim position on the potential use of generative AI within the Ministry, considering the recent government System Leads' interim guidance on this topic and an anticipated increasing need to leverage these tools to realise the gains of the Te Pae Tawhiti transformation programme.

MSD's commitment to Māori

3. Mana Manaaki – any use of generative AI tools involves risks that must be carefully managed, and that these risks will impact any use of these tools with Māori data or te ao Māori concepts, or to make decisions that may impact Māori.
4. Kia Takatū Tātou – through freeing up staff capacity and by enabling better use of information we already hold effective use of generative AI tools could enable us to design and deliver more effective services for Māori.

Recommendations

5. We recommend that you:

a) **note** that generative AI tools are increasingly available, both as stand-alone products and as new features of products that the Ministry already uses, and that there are potential use cases for these tools for Te Pae Tawhiti service changes in the programme's first horizon

b) **note** that the Office of the Privacy Commissioner[1, 2] and the Government data, digital, procurement, privacy, and cyber security System Leads[3] have produced interim guidance[4] on the use of generative AI. It is expected that this guidance, and by extension our position, will be subject to change over time

c) **note** that the Ministry's existing Information, Data and Analytics Strategy and information policies, along with the Algorithm Charter and Data Protection and Use Policy provide a high-level framework for positioning generative AI, but this will need to be further supported by standards, patterns, and operational guidance. Generative AI capabilities will be realised through the guidance of MSD's Technology Strategy and (platform) roadmaps, leveraging various technology capabilities

d) **note** generative AI shares commonalities with Automated Decision-Making, with guardrails essential for both technologies recognising that they are inherently interconnected

e) **note** that any use of generative AI involves potential trustworthiness, bias, and memorisation risks (Appendix 1) that will need to be managed to meet our obligations for responsible use. It is expected that over time these risks will become more manageable as technology matures

f) **approve** the recommended interim position:

i. where there are identified business needs to use generative AI over other mechanisms, a risk-based approach should be taken to assess use, in line with System Leads' guidance and the Ministry's existing approach to the use of algorithms. This risk-based approach will be supported by the development of additional controls to mitigate each risk

ii. generative AI tools can be considered for use where there are identified business needs, but that our pace of adoption will be kept to within our ability to ensure its responsible use. This includes consideration of the policy and legislative environment we operate within.

[1] https://www.privacy.org.nz/publications/guidance-resources/generative-artificial-intelligence/
[2] https://www.privacy.org.nz/publications/guidance-resources/ai/
[3] https://www.digital.govt.nz/standards-and-guidance/technology-and-architecture/interim-generative-ai-guidance-for-the-public-service/
[4] Archives New Zealand has also released guidance restating existing obligations for management of system documentation. It has not informed this paper.

g) **approve** the attached staff guidance (Appendix 2) and two low-risk use cases:

   i. Personal use of generative AI

   ii. Using generative AI for ideation only.

h) **endorse** the Emerging Technology Advisory Group (ETAG) supporting development of exploratory pathways for future use cases, utilising Te Pae Tawhiti and to understand potential impacts from generative AI tools, including third-party use, with a view to presenting to Tai Nuku for approval.

i) **endorse** ETAG developing a process for efficiently assessing Generative AI use cases against value and risk, with a view to presenting to Tai Nuku for approval.

Background

Context

AI vs Generative AI

6. AI refers to the broader field of computer systems emulating human intelligence, encompassing various tasks like problem-solving, language processing, and image recognition. Generative AI is a subset of AI dedicated to content creation to generate creative content, such as images, text, or music. The key distinction is that AI covers a wide range of tasks, while generative AI specialises in content generation.

7. Generative AI models are built by consuming very large quantities of data, harvested from the internet and other data sources, and processed to build the tool. This data collection can involve both public and private sources, and for some models (including ChatGPT) these sources are not disclosed.

8. Generative AI presents several risks, primarily in its ability to create artificial content that can be used for malicious purposes or relied upon as factual or accurate. This technology can generate content that resembles human-created data to a degree that makes it challenging to distinguish from accurate information. Consequently, there is a heightened risk of misinformation, privacy issues, and potential harm to individuals and society.[5]

Opportunities

9. Generative AI tools such as ChatGPT are increasingly available and are both engaging to use and positioned by their creators as important productivity tools. The risks and potential benefits available from their uses are not yet well understood by Ministry staff. We are starting to receive requests to approve these tools for business use, including tasks such as assistance drafting memos, updating Confluence pages, and assisting with articulation of business issues.

10. In response to this increased attention, a working group has been established with representatives from IST, Insights, Legal Services, Policy, Te Pae Tawhiti Programme, and the Information Group, with the initial goal of bringing together potential interested parties across the Ministry. This group has since

[5] See Appendix 1 for further information.

been formally established as the Emerging Technology Advisory Group (ETAG), supporting Tai Nuku.

11. s9(2)(g)(i) [redacted]

- [redacted]
- [redacted]
- [redacted]
- [redacted]
- [redacted]
- [redacted]
- [redacted]
- [redacted] s9(2)(g)(i)

12. s9(2)(g)(i) [redacted]

- [redacted]
- [redacted]
- [redacted]
- [redacted]
- [redacted]
- [redacted] s9(2)(g)(i)

13. These use cases have the potential to boost productivity and streamline interactions but must be assessed appropriately to ensure responsible use.

14. s9(2)(g)(i) [redacted]

[6] This use case has been proposed as a test case, using GitHub Co-pilot as the tool.

s9(2)(g)(i) [redacted]

[redacted] [redacted] s9(2)(g)(i).

16. As this technology matures, it is likely that further potential uses may become more viable,[7] including client-facing uses such as chatbot assistance or context driven help messages.

17. These potential uses are strongly weighted towards enabling staff to focus their time and attention on helping clients and more generally on delivering value, in line with the overall direction of Te Pae Tawhiti and the Future Services Model.

18. It is important to note that *all* the use cases identified so far carry some level of risk that would need to be assessed and controlled before they could be safely used by the Ministry.

19. As AI continues to evolve and integrate with technology, it's conceivable that it may become inseparable from the core functionality of tools, making it challenging to toggle on or off, as vendors prioritise sustaining the primary mode of operation. For example, Microsoft is already promoting the availability of generative AI tools in the premium version of Teams via their Co-pilot functionality, which is entering early access from November 2023.[8]

20. In the absence of an agreed position on the use of generative AI, the Ministry will not be able to safely realise any possible productivity gains from its use, and staff may inadvertently place us or themselves at risk through personal or work-related use of these tools.

All-of-government advice

21. The Office of the Privacy Commissioner and the data, digital, procurement, privacy, and cyber security System Leads have issued guidance on the use of generative AI tools. The guidance aligns with the Algorithm Charter's expectations on machine learning and includes recommendations against using AI with classified information. Following discussions with DIA, the language used in their guidance reflects a broad audience which includes agencies with lower maturity and capability.

22. The key messages following our discussions with DIA were to proceed with caution, establish clear policies and guidance, assess tools, and balance any use with ethical and legal implications. Our proposed interim position reflects this advice and guidance.

Alignment with existing policy frameworks and potential controls

23. MSD's existing Information Policy Framework is appropriate to assess and manage how information is used that can be applied to generative AI. These framework components could be effectively supplemented by additional

[7] These uses will still need to be assessed on a case-by-case basis.

[8] Other examples include Google incorporating generative AI into its search engines, and Salesforce piloting their own offering across all their products.

guardrails, including control patterns for related use cases, role-based training and guidance, staff education, and as required extending existing standards or developing new ones to ensure necessary coverage.[9]

24. Depending upon the use case, potential additional controls may include access controls to limit use, promoting transparency and accountability to document model usage and work-type, applying data minimisation principles by default, and undertaking regular assurance or auditing activities.

25. Consideration may also be given to establishing feedback mechanisms that allow users to report concerns or issues related to the use of generative AI models, which are promptly addressed.

s9(2)(h) [redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

Risks

31. Any use of generative AI needs to account for three sources of information risk – trustworthiness, bias,[11] and memorisation and reproduction.[12] These risks are inherent to generative AI and can at best only be mitigated. These

[9] For example, we could introduce a specific "AI policy" to make our positioning more explicit.

s9(2)(h) [redacted]

[11] Bias is a risk also present in many non-generative uses of AI; however the black-box nature of generative AI poses additional challenges for identifying and mitigating bias.

[12] This is not an exhaustive list of risks. There is no standard listing of generative AI risks available, although see footnote 5 for a detailed analysis of risks involved in model development.

risks are described below and presented in standard information risk format with scenarios and current and proposed future controls in Appendix 1.

32. Generative AI risk should be assessed with particular care in instances where generative AI is used with Māori data, and consultation with Māori included as a consideration in this assessment. Similar care must also be applied with any use which could cause harm to individuals, whānau, or communities.

33. These risks are in addition to any associated with the specific systems and processes that the generative AI tool is integrated into, including those relating to information security and privacy as assessed through our existing certification and accreditation process.

34. The Ministry's ability to manage these risks is expected to improve over time as the technology matures, external regulation frameworks are developed, and we build expertise in designing effective controls.

Trustworthiness

35. Generative AI models may exhibit 'hallucination', where the outputs may not always be reliable or accurate. Understanding the limitations in trustworthiness can help users make informed decisions about its use. Because hallucination is a product of how generative AI models work, they can be incorrect even when their source data is consistent and without error. This includes statements made by the AI about why or how it produced specific outputs.

36. This introduces challenges to verification processes, particularly where traceability matters (including where the Ministry may require evidence of its decision-making processes, such as through the review and appeal process, litigation before the courts or when explaining decisions to the Ombudsman.)

Bias

37. Generative AI models reflect the bias in both their source data and in the moderating processes used to attempt to mitigate this prior to deployment. This bias is present in both direct outputs from AI, but also in the interpretation it applies to any inputs provided. As an example, if used to summarise a set of input data, both the framing of the output and way that different characteristics of the input data are assessed for importance will reflect this bias.

38. Any lack of visibility over the source data and moderation processes for generative AI make the specifics of this bias difficult to predict.

Memorisation and reproduction

39. Memorisation and reproduction are related risks about data used in model development being included in model outputs. Memorisation is where uncommon data points, like specific names and addresses, are recreated as outputs. Reproduction is instead the near-exact recreation of material that is under usage restrictions such as copyright, or that is outdated, incorrect, or otherwise inappropriate to use.

40. Where input data persists between users, that data may also be surfaced through these processes.

Proposed Interim Position

41. We recommend an interim position to adopt:

    a) a risk-based approach to generative AI, aligning with the Ministry's approach to algorithms, developing additional controls to mitigate each risk if necessary to ensure the use of generative AI is justified

    b) a general posture that acknowledges the potential of generative AI tools for business needs, but maintaining a responsible pace of adoption, considering the policy and legislative environment.

42. Given the number of potential use cases, tools available, and the varying level of risk between uses, it is not possible to address all scenarios in advance. Applying a risk-based approach allows the Ministry to appropriately manage business use safely and to adapt to this emerging technology in a controlled manner.

43. Further, we recommend the approval of the staff guidance set out in Appendix 2, and two low risk use cases of:

    a) Personal use of generative AI

    b) Using generative AI for ideation only.

44. Personal use of generative AI is unlikely to pose significant risk to the Ministry, if the risks are well understood by staff and no Ministry information is used. The Ministry's Code of Conduct and associated policies (in particular, Social Media and Acceptable Use of Technology) adequately cover expectations on how staff should use these tools, bringing them generally in line with Ministry requirements for general internet use.

45. Ideation, outlining, and other forms of thinking prompts will also generally be low risk, as they can be created without use of Ministry information as input data and the impacts of incorrect, biased, or reproduced outputs are minimal. This use case does not extend to the inclusion of generative AI outputs as part of a decision-making process, as robust controls would need to be in place to manage the associated risk, and those need to be tailored to the specific process.

46. Staff education is an important control for low risk uses, and an example of guidance is attached as Appendix 2. This guidance if released would need to be reviewed regularly to ensure it remained current.

All other uses approved on a case-by-case basis

47. For all other proposed uses of generative AI, approval should be managed on a case-by-case basis using our existing approach to certification and accreditation and requiring business risk acceptance at the appropriate delegation as described in the Ministry's Risk Framework. These use cases must also be informed by active business drivers, such as those emerging from Te Pae Tawhiti, and must not be able to be met effectively through other mechanisms.

48. Assessment can be managed within the Ministry's existing policy frameworks, with additional guardrails developed as required to improve consistency and better manage overall risk (as indicated in Appendix 1's future controls). It is likely that some potential uses identified by the business will not be within the Ministry's risk culture and will be declined through this process.

49. In line with OPC expectations, senior leadership approval should be required for each proposed use case[13] before they are implemented. An advisory group such as ETAG should be used to support senior leadership in these approval decisions through provision of expert advice, along with existing governance forums.

50. This position should be reviewed regularly to ensure alignment with any all-of government directives and, noting that as technology matures and generative AI becomes more common, some potential uses may shift with regards to the Ministry's risk appetite.

Comparison with other agencies and OPC and System Leaders guidance

51. There is no standard position on the use of generative AI across agencies. System Leads guidance does not advocate for a particular position (other than caution), and no agency has widespread adoption of this technology.

52. Across all agencies we contacted[14], their current position was either in development or only recently approved.

53. A small number of agencies have taken a position that no use of these tools is currently appropriate while they work through more detailed positions.

54. Other agency positions we are aware of include:

    a) s9(2)(g)(i) [redacted]

    b) s9(2)(g)(i) [redacted]

    c) Inland Revenue has created an AI oversight group with senior leaders across business units, a working group to advise and create guidance artefacts, and an interest group for staff members to discuss and propose AI use cases. A generative AI policy has been drafted which sets out expectations, definitions, and acceptable use cases

    d) ACC has procured Co-Pilot (an AI 'companion') addons for their Microsoft E5 license prior to undertaking a risk assessment, receiving Chief Executive approval to conduct a pilot while completing the comprehensive risk assessment process later.

55. Our proposed interim position is consistent with OPC and System Leads guidance, noting their concerns about use classified information, and is broadly comparable with those of other agencies.

Consultation

56. Emerging Technology Advisory Group.

57. Automated Decision-Making Working Group.

[13] OPC guidance requires this for each tool, but use cases is more appropriate for the Ministry's risk culture. ChatGPT is a tool, but we would want every use case for ChatGPT to be assessed and approved separately, as they will potentially vary significantly in risk profile.

[14] Te Puni Kokiri; Ministry of Business, Innovation, and Employment; Statistics New Zealand; Inland Revenue; Ministry of Foreign Affairs and Trade; Oranga Tamariki; and (not an agency) the Government Chief Privacy Officer

58. Tai Nuku Design Committee consultation is scheduled for its November 22 meeting.

59. We will continue to work closely with System Leads and Gartner in the formulation of essential controls and safeguards to reinforce the explicit implementation of our key information policies.

Next Steps

60. If the recommendations are accepted:

    a) the ETAG will develop a Doogle presence and associated messaging to communicate this position to Ministry staff, including to existing policies, guidance on acceptable uses, risks involved, and how to use these tools safely for those purposes

    b) the ETAG will formally incorporate assessing proposed uses of generative AI into its terms of reference as one of its core activities

    c) any initial use cases will be assessed to determine if they can be effectively used within the Ministry's risk culture. Recommendations will go to Tai Nuku for approval

    d) the Information Group, Improvement Systems and Technology, and other groups across the Ministry will work to further develop guardrails, monitoring, and assurance as required.

Attachments

Appendix 1 – generative AI risks

Appendix 2 – draft guidance

s9(2)(g)(i)

[redacted]

s9(2)(g)(i)

[redacted]

s9(2)(g)(i)

[redacted]

s9(2)(g)(i)

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]		[redacted]
[redacted]	[redacted]		[redacted]
[redacted]			

[redacted]

[redacted]

s9(2)(g)(i)

s9(2)(g)(i)

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]	[redacted]
[redacted]	
[redacted]	
[redacted]	
[redacted]	
[redacted]	
[redacted]	

s9(2)(g)(i)

[redacted]

[redacted]

s9(2)(g)(i)

- [redacted]
- [redacted]
- [redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

s9(2)(g)(i)

[redacted]

[redacted]

[redacted]

[redacted]

s9(2)(g)(i)

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

- [redacted]
- [redacted]
- [redacted]
- [redacted]
- [redacted]
- [redacted]

[redacted]