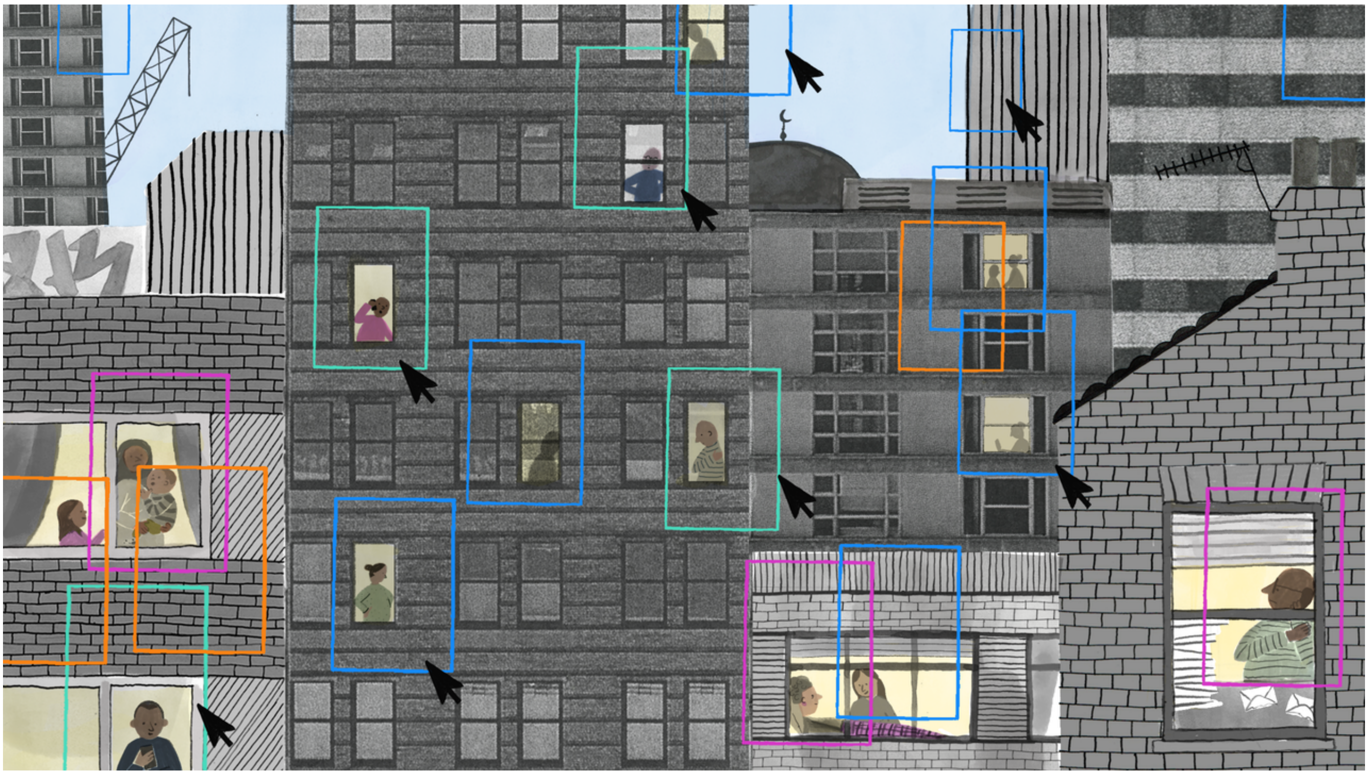
PUBLIC HEALTH
COMMUNICATION CENTRE
AOTEAROA



Has the time come for a register of AI systems used by government agencies?

27 June 2024

Ali Knott, Karaitiana Taiuru, Robyn Whittaker, John Kerr, Michael Baker

Summary

Artificial Intelligence (AI) is rapidly advancing, especially in "generative" AI tools like ChatGPT with new systems emerging frequently. Several recent reports in Aotearoa New Zealand (NZ) offer a roadmap for AI implementation, highlighting the citizens' right to know where AI is being used and how well it performs.

Evaluating AI systems is crucial, particularly in healthcare, where accuracy directly impacts health outcomes. AI use is expanding in public health service delivery. It has wider public health importance because it is changing the economic and social environment in which we live. Public transparency about AI use and performance is essential for building trust, assessing bias, and determining the level of human oversight needed. We encourage the NZ Government to regularly publish a register of all AI algorithms used across the public sector, including evaluation of their performance and their potential impact on equity.

Artificial Intelligence (AI) is progressing by leaps and bounds. The most tangible progress is currently seen in ‘generative’ AI tools: for instance, systems like ChatGPT, which can engage in humanlike dialogues with a user, or systems like MidJourney, which can generate images. Products using generative AI are being pushed hard at us, as companies compete for very lucrative new markets. Many people and many organisations are encountering AI for the first time: that includes health professionals, and health organisations.

AI as a public health issue in Aotearoa New Zealand

AI is already a major emerging issue for the healthcare sector. In NZ, a few initiatives provide useful context. Most immediately, the Prime Minister’s Chief Science Advisor recently released a report on AI and healthcare,[1] which makes many specific recommendations and outlines a set of key principles for the use of AI in health. These align with similar principles put forward by international organisations such as the OECD.

In NZ, one district health service established a framework for governance over the development and implementation of AI tools across its organisation.[2] This framework prioritises the voice of Māori and the needs of healthcare service users and their families. This has now been adopted by a national Te Whatu Ora|HealthNZ AI Expert Advisory Group.

AI also has multiple emerging uses for the delivery of improved public health services that protect and improve the health of populations in such areas as disease surveillance.[3,4] AI has the potential to facilitate improved delivery of interventions to populations that are most in need, sometimes called ‘precision public health’, analogous to the ‘precision medicine’ of clinical care.[5]

But less obviously, AI is also changing the social and economic environment in which we live through the way in which AI technologies are deployed and governed by NZ’s public sector.

Consideration of the impact of AI in government began with a 2018 stocktake of ‘algorithms’ used in government departments[6] and a 2019 academic report into uses of AI by NZ Government departments.[7] These reviews informed an Algorithm Charter for

Aotearoa, which provided one of the world's first all-of-government guides on AI use. This broader charter, which is the subject of ongoing reviews,[8] provides a helpful framework for discussions about use of AI in healthcare in NZ.

The importance of evaluating AI systems

An important focus for all applications of AI is evaluation. To state it baldly: how well does a given AI system work? How often is it right? How does it compare with a person performing the same task? Does it work equally well for all socio-demographic groups? That is, is it fair and equitable? How does it fit in, or alter, the current workflow? What are the likely impacts of implementing this AI?

It is widely recognised that Māori experience significant health inequities and bias in health service provision compared to the non-Māori population.[9,10] These poorer outcomes are likely to be mediated through multiple pathways such as discrimination when assessing housing.[11] Bias is a complex multidimensional factor to consider when designing and evaluating AI for healthcare. While it is an added short-term cost, all new health AI teams should be diverse and should account for the introduction of bias at each stage of the AI lifecycle, and then in testing and deployment, while also continually ensuring the Hauora Report[12] and the Te Tiriti AI Principles are also applied.

By not doing so, we risk following international trends, where AI tools are based on majority populations and ignore Indigenous and People of Colour and their circumstances.[13] This tendency can be linked to homogeneous design teams generalising their own need and characteristics and therefore creating or perpetuating inequitable health outcomes.

Answering questions of performance and bias is important for public sector AI systems, where principles of open government demand particular transparency.

In the case of healthcare, evaluation is obviously vital for AI tools as health outcomes are at stake. Health professionals who consider a given tool must know how well it works. They don't necessarily need a tool that does everything perfectly: their key demand is for something that *helps them* perform a given task. For professionals, evaluations of AI tools are largely useful in identifying where they can help (and where they cannot).

The public also has a right to know how well healthcare AI systems perform—and indeed patients in Aotearoa expect transparency around the use of AI in healthcare.[14] Performance information should provide an understanding of what level of human oversight is required for a given tool, and what expectations about human involvement there should consequently be.

This same right applies all through the public domain. We have a right to know how well all NZ Government-operated AI systems work, so we can frame the right expectations about how human staff should be involved for each system. For instance, ACC has a model that predicts when a claim is 'easy' and can be immediately accepted. It's useful for us to know this model works almost perfectly, so we can understand why it operates routinely without human involvement. The Department of Corrections has a model that predicts a defendant's risk of reoffending. We would like to know how accurate this model is too, so we know how its performance must be complemented by expert human assessments.

How should AI systems be evaluated?

Evaluation is central to all modern AI systems. It’s useful to distinguish (crudely) between two types of AI systems:

- ‘predictive models’ trained to perform a specific task, and
- ‘generative models’ like ChatGPT which can be applied to a range of tasks for which they are not specifically trained (eg, answering questions or translating).

See Appendix for more detail.

We understand how to evaluate predictive models. However, the question of how to evaluate generative AI systems is still very much a matter for active research. Stanford’s AI Index for 2024 notes that robust and standardised evaluations for the safety and trustworthiness of these models are ‘seriously lacking’. If we are still working out how to evaluate generative AI systems, we must be very cautious in deploying them—especially in safety-critical areas such as healthcare.

How should evaluations be presented to the public?

A central idea in the recent AI-in-healthcare report is that medical AI systems operating in NZ should be publicly evaluated.[1] If AI systems are to be widely used in healthcare, they must be trusted by the public: the report argues that public evaluations are important in establishing public trust.

This idea connects with a proposal in the broader NZ report on AI in government, that there should be a regularly published *register of AI algorithms* used in government departments, including reporting on their performance.[7]

As AI technologies continue to advance, and practical AI tools start to be deployed in large numbers, in healthcare and elsewhere in the public sector, the time has come to implement a register and robust evaluations.

What this Briefing adds

- AI capabilities and usage are increasing, including in healthcare and the wider public sector, making AI a public health issue.
- The public has a right to know how AI systems are used in public services and how well they perform, including with regard to equity and fairness.

Implications for policy and practice

- We need to develop clear frameworks for evaluating the performance of generative AI systems.
- The NZ Government should publish a register of all AI algorithms used by the public sector, with a focus on those used for guiding delivery of healthcare, public health, and wider government services.
- The register should also regularly report evaluations of the performance of AI algorithms and their likely impact on equity.

Author details

Prof Ali Knott, Professor in Artificial Intelligence, School of Engineering and Computer Science, Victoria University of Wellington

Dr Karaitiana Taiuru, Director, Taiuru & Associates; AI, Data and Emerging Tech Ethicist

Dr Robyn Whittaker, Adjunct Professor, National Institute for Health Innovation, University of Auckland and Public Health Physician, Te Whatu Ora

Dr John Kerr, Science Lead, Public Health Communication Centre, and Department of Public Health, University of Otago Wellington

Prof Michael Baker, Director, Public Health Communication Centre, and Department of Public Health, University of Otago Wellington

Appendix: Distinguishing predictive and generative AI

AI systems are *trained* to perform some tasks, on a set of ‘training examples’. Training involves optimising the system to perform as well as possible on these examples: assessing this performance during training involves quantitative evaluation.

In traditional **predictive** tools, the task for which they are trained is *the same* as the task for which they will be practically deployed. For instance, a radiology system may be trained to identify cancerous skin lesions from photographs—and then deployed to perform exactly this same task. Of course, the photographs it sees during deployment are different from those it saw during training—but there are ways of checking how well it works on ‘unseen’ images, and therefore of scientifically evaluating how well it works during real-world deployment.

In the recent **generative** AI tools, such as ChatGPT, the task the system is trained on is largely *different* from the task it will be deployed to perform. Language models like ChatGPT are trained to predict missing words from a text. (In particular ‘the next word’ in a piece of text interrupted at an arbitrary point). But after training, these systems are *used* to perform tasks that are defined at much higher levels, such as answering questions, or summarising, or translating. They are not purpose-built to do well at these specific tasks. They may be ‘aligned’ or ‘fine-tuned’ for such tasks at the end of training, but the bulk of their training is not so focussed.

There are well-established methods for evaluating predictive models. In many cases, we can simply withhold some of the training examples, and document the percentage of cases where the trained system performs correctly on these held-out examples. Evaluation of generative models is more complex, and in many cases still a matter for ongoing research (see Deci.ai’s ‘Ultimate Guide to LLM Evaluation’ and Confident AI’s ‘Definitive Guide to LLM Benchmarking’ for some recent overviews).

Image credit: Emily Rand & LOTI / Better Images of AI / AI City / Licenced by CC-BY 4.0

References

1. Office of the Prime Minister’s Chief Science Advisor. Capturing the benefits of AI in healthcare for Aotearoa New Zealand. Auckland; 2023. https://www.pmcsa.ac.nz/artificial-intelligence-2/ai-in-healthcare/
2. Whittaker R, Dobson R, Jin CK, Style R, Jayathissa P, Hiini K, Ross K, Kawamura K, Muir P, Waitematā AI Governance Group. An example of governance for AI in health services from Aotearoa New Zealand. *NPJ Digital Medicine*. 2023 Sep 1;6(1):164. https://doi.org/^{10}.1038/s41746-023-00882
3. Fisher S, Rosella LC. Priorities for successful use of artificial intelligence by public health organizations: a literature review. *BMC Public Health*. 2022 Nov 22;22(1):2146. https://doi.org/10.1186/s12889-022-14422-z
4. Brownstein JS, Rader B, Astley CM, Tian H. Advances in artificial intelligence for infectious-disease surveillance. *New England Journal of Medicine*. 2023 Apr 27;388(17):1597-607. https://doi.org/^{10}.1056/NEJMra2119215
5. Khoury MJ, Iademarco MF, Riley WT. Precision public health for the era of precision medicine. *American Journal of Preventive Medicine*. 2016 Mar;50(3):398. https://doi.org/10.1016/j.amepre.2015.08.031
6. Stats NZ. Algorithm Assessment Report; 2018. https://data.govt.nz/docs/algorithm-assessment-report/
7. New Zealand Law Foundation. Government use of artificial intelligence in New Zealand. Wellington; 2019. https://www.otago.ac.nz/__data/assets/pdf_file/0027/312588/https-wwwotagoacnz-caipp-otago711816pdf-711816.pdf
8. Taylor Fry. Algorithm Charter for Aotearoa New Zealand Year 1 Review. 2021. https://www.data.govt.nz/assets/data-ethics/algorithm/Algorithm-Charter-Year-1-Review-FINAL.pdf
9. Ministry of Health. Tatau Kahukura: Māori Health Chart Book 2015 (3rd edition). Wellington; 2015. https://www.health.govt.nz/system/files/documents/publications/tatau-kahukura-maori-health-chart-book-3rd-edition-oct15.pdf
10. Talamaivao N, Harris R, Cormack D, Paine SJ, King P. Racism and health in Aotearoa New Zealand: a systematic review of quantitative studies. The New Zealand Medical Journal. 2020 Sep 4;133(1521):55-. https://pubmed.ncbi.nlm.nih.gov/32994637/

11. Harris R, Cormack D, Tobias M, Yeh LC, Talamaivao N, Minster J, Timutimu R. The pervasive effects of racism: experiences of racial discrimination in New Zealand over time and associations with multiple health domains. Social Science & Medicine. 2012 Feb 1;74(3):408-15. https://doi.org/10.1016/j.socscimed.2011.11.004
12. Waitangi Tribunal. Hauora: Report on Stage One of the Health Services and Outcomes Kaupapa Inquiry. 2023. https://forms.justice.govt.nz/search/Documents/WT/wt_DOC_195476216/Hauora%202023%20W.pdf
13. Shanklin R, Samorani M, Harris S, Santoro MA. Ethical redress of racial inequities in AI: lessons from decoupling machine learning from optimization in medical appointment scheduling. *Philosophy & Technology*. 2022 Dec;35(4):96. https://doi.org/10.1007%2Fs13347-022-00590-8
14. Dobson R, Whittaker R. What Do Health Service Users Think About the Use of Their Data for AI Development?. In MEDINFO 2023—The Future Is Accessible 2024 (pp. 1156-1160). IOS Press. https://doi.org/10.3233/SHTI231146

Public Health Expert Briefing (ISSN 2816-1203)

Source URL:
https://www.phcc.org.nz/briefing/has-time-come-register-ai-systems-used-government-agencies