Professor Colin Gavaghan, Faculty of Law, University of Otago, Dunedin, New Zealand — Written evidence (NTL0047)

1.   Thank you for the opportunity to contribute this supplementary information to my oral evidence. It is not my intention to add unnecessarily to what I am sure will be copious quantities of information you will receive on many aspects of this. Instead, I will focus on the New Zealand context, and on the matter to which the Committee specifically directed me, transparency.

2.   Recognition of the importance of transparency, and the challenge of opaque "black box" systems,[1] are ubiquitous features of discussions of new technologies, and particularly of AI, algorithms and automated decision making (ADM) systems. This is as true in New Zealand as elsewhere. Our Algorithm Charter, to which all government agencies have signed, contains a commitment to "Maintain transparency by clearly explaining how decisions are informed by algorithms."[2]

3.   New Zealand law also contains a statutory 'right to reasons' in respect of official decisions:

> where a department or Minister of the Crown makes a decision or recommendation in respect of any person in his or its personal capacity, that person has the right to be given a written statement of… (c) the reasons for the decision or recommendation.[3]

While this may be relevant with regard to transparency of algorithmic decisions, it should be noted that the extent of this right has yet to be tested in that context.

4.   If this statutory and policy commitment is to be more than nominal, however, the concept requires to be unpacked a bit more. In our work, we have identified several different senses in which the term could be used and various layers and levels at which it is relevant.

5.   Perhaps the most basic level of transparency relates to knowledge of *which new technologies are being used, by whom, and for what purposes*. The New Zealand experience suggests that there is a low level of public awareness of these matters. As one youth respondent to Digital Council's focus groups said last year:

---

[1] Frank Pasquale. *The black box society: The secret algorithms that control money and information*. (Harvard University Press, 2014)
[2] Algorithm Charter for Aotearoa New Zealand (July 2020)
[3] Official Information Act 1982, Section 23 (1)

> "I wasn't aware at all of this kind of thing. I didn't even know that there's all these government departments doing this kind of thing."[4]

6.   Certain recent steps have attempted to raise the level of public awareness of such systems.

- In 2018, StatsNZ, New Zealand's official data agency, conducted and published a stocktake of Government use of algorithms;[5]
- Earlier this year, New Zealand Police commissioned consultants TaylorFry to perform a stocktake of algorithms used by NZ Police.[6]

7.   While valuable, it should be noted that each of these stocktakes amounted to a snapshot of what was being used. Absent a process for regular updating, these will rapidly become obsolete. A **regularly maintained and publicly accessible register** of algorithms and ADM systems used by government agencies would go a long way to addressing this problem.

8.   As well is knowing what is being used, and by which agency, the Digital Council's research revealed a strong desire to know more about the *purposes* for which technologies were being used. Is an algorithm designed to identify young people at risk of future offending going to inform decisions to offer upstream support, or to justify closer surveillance? A register could address this by listing the *uses* to which the technology is put, and the context in which it is deployed, as well as its form, though an insight into the values and assumptions behind that use may be more challenging.

9.   Transparency can also relate to the capacity to check, audit, query or appeal against algorithmic decisions. This can refer to *accountability* or *answerability*: an agency's or person's responsiveness to requests for information or willingness to offer justification for actions taken or contemplated. Citizens may want to know who is responsible for a system's use, and whom to contact in order to query an algorithmic decision.

10.   New Zealand's Algorithm Charter goes some way to addressing this form of transparency. Signatories commit respectively to:

- nominating a point of contact for public inquiries about algorithms, and
- providing a channel for challenging or appealing of decisions informed by algorithms.

---

[4] Digital Council for Aotearoa New Zealand. *Towards trustworthy and trusted automated decision-making in Aotearoa*, 2020. https://digitalcouncil.govt.nz/advice/reports/towards-trustworthy-and-trusted-automated- decision-making-in-aotearoa/

[5] StatsNZ. *Algorithm Assessment Report* (October 2018) https://www.data.govt.nz/assets/Uploads/Algorithm- Assessment-Report- Oct-2018.pdf

[6] TaylorFry. *NZ Police - Safe and ethical use of algorithms*. (June 2021)

11. Transparency can also relate to the *inspectability or auditability* of institutions, practices and instruments. How does this or that tool actually work? How has it been trained and tested? Algorithms can be "inspected" in two ways.

- we can enquire of their provenance; how were they developed, by whom, and for what purpose/s. This might be called *process transparency*.
- we can ask how an algorithm works, what data has it been trained on, and by what logic does it proceed. This might be called *technical transparency*.

12. For all forms of transparency, it is important to keep in mind the fundamental question: transparent/explainable *to whom*? Is our focus on rendering the system transparent to computer experts within the relevant agency; to the end user (eg the police officer on the street); to a court called upon to accept or review evidence from those systems; to the parties subject to the algorithmic decision; or to the population at large?

13. Transparency and explainability are inevitably relative to some domain of expertise. The form of transparency sought by a court or a lawyer will likely differ from the form sought by a software engineer. Coders and lawyers will want to know—and be able to understand—different things about an automated system. The level at which ordinary people subject to an automated decision might wish to understand that decision may differ again. If transparency is to be meaningful, information must be made available at a level; that is comprehensible to the relevant audience.[7]

14. Obstacles to transparency can be *technical*. Some algorithmic systems are just difficult to explain to non-experts, and with more advanced forms of AI, perhaps even to experts. Other obstacles are *legal*. Intellectual property rights might prevent the disclosure of proprietary code, or preclude access to training data, so that even if it were possible to understand how an algorithm operates, a full reckoning may not be possible for economic, legal or political reasons. This was the situation faced by the Wisconsin Supreme Court in the *Loomis* case,[8] where it held that:

> Northpointe, Inc., the developer of COMPAS, considers COMPAS a proprietary instrument and a trade secret. Accordingly, it does not disclose how the risk scores are determined or how the factors are weighed.

15. This last issue is in theory easy to avoid by adopting policies around development and procurement that avoid this problem; tools should not be procured unless their terms or use guarantee acceptable levels of transparency. As noted in my oral evidence, New Zealand Government

[7] Danaher, Gavaghan, Knott, Liddicoat, Maclaurin and Zerilli. *A Citizen's Guide to Artificial Intelligence* (MIT Press, 2021), Chapter 2.
[8] *Wisconsin v Loomis* 881 N.W.2d 749 (Wis. 2016)

appears so far to have a good record on this.

16. It should be noted, though, that such an approach may sometimes involve certain trade-offs between transparency on the one hand, and accuracy and efficiency on the one other – as where the most accurate tool on the market is offered by a developer that requires high levels of opacity.

17. Mechanisms for ensuring transparency and explainability are being deployed in a somewhat piecemeal manner across the New Zealand government sector. Two of the more promising initiatives thus far are:

   - The Ministry for Social Development have adopted a Privacy, Human Rights and Ethics (PHRaE) framework,[9] against which any proposal to use personal information (including, but not limited to those using new technologies) must be assessed.

   - New Zealand Police have established an Expert Panel for Emergent Technologies, "to provide advice and oversight from an ethical and policy perspective of Emergent Technologies."[10] They will shortly be adopting a New Technology Framework, that embeds commitments to accountability and transparency, but importantly always provides a pathway to ensuring these commitments are adhered to.

   This combination of in-house evaluation and referral to an independent expert body seems a promising approach to ensuring that commitment to high level principles are operationalized in a meaningful way. These processes do not, however, answer the "transparent to whom?" question, and further work on that is required here in New Zealand as elsewhere.

18. A final observation of a more general nature: in discussing bias during my oral evidence, I expressed the view that the relevant comparator for an algorithmic system is not some notionally perfect human decision-maker. This is true also in the context of transparency. While humans can be called to give reasons for their decisions and actions, it is well recognised that there is no perfect way of gaining access to these reasons. We are all prone to ex post rationalisations, confabulations and assorted cognitive biases, that will often deny us insight even into the reasons underlying our own decisions. Some commentators believe that algorithmic decision making tools could be designed in such a way as to be more transparent than humans I this regard. Insofar as this is technically realizable, it is a goal that should be pursued.

30 September 2021

---

[9] https://www.msd.govt.nz/documents/about-msd-and-our-work/work-programmes/initiatives/phrae/phrae- on-a-page.pdf

[10] https://www.police.govt.nz/about-us/programmes-and-initiatives/police-use-emergent-technologies/qa- emergent-technology