We asked

Expert Panel on Emergent Technology were asked to provide advice on a proposal to further develop a pilot of two geographic profiling algorithms.

In June 2023, we asked the Panel's advice on a proposal to further develop a pilot of two geographic profiling algorithms. One of the algorithms has been established as well-founded and useful around the world, and the second has been developed by a researcher at Police. The algorithms seek to provide intelligence support to investigations to assist them in prioritising their investigative resources (ie. staff).

They said

1. The panel thanks NZP colleagues for the report, and also for the chance to work through some discussion points at our meeting on 29 June 2023. We appreciate the opportunity to have had a discussion at this early stage, and also acknowledge the thoughtful responses to our questions.

2. The Panel acknowledges that the referral is important and passes the NZP necessity principle.

3. Two different types of profiling algorithms were put forward by NZP for discussion: (a) the Anchor Point algorithm; and, (b) the Activity Location algorithm (GPSmart). NZP advised that they are pilot testing this process with live cases at present and that the pilot considers prioritisation output from anchor point prediction and GPSmart output.

4. The Panel recognises that currently the pilot is evaluating whether the algorithms are appropriate and worthwhile, and that any future use would require a more formal business case to be made. We nevertheless reiterate that pilot protocols will need to be different to operational protocols and that there is a difference between evaluating a pilot and a nationwide rollout.

5. The Panel also notes that, ideally, any evaluation would be conducted independently and at the very least recommends that future, ongoing evaluation needs to be done by a different person/group to the one who used and/or developed the algorithm. As robust, detailed and appropriate evaluation criteria are key, the Panel would welcome an opportunity to review and discuss these as they are developed.

6. NZP advised there was no time limit on the pilot, but that it was about reaching saturation. NZP suggested 10 cases was deemed sufficient to have a clear steer on whether this was a valuable addition to NZP tech. The Panel acknowledge this but feel it would be helpful to have such parameters more clearly signposted.

7. The Panel understands that there is a growing empirical base for the use of GPT and it is a well-established activity within the UK and other jurisdictions. These experiences have highlighted a number of risk areas, however, including a potential for heuristic bias which had tragic results in the Rachel Nickell case. The Panel noted that NZP will utilise caveats to avoid putting undue weight on the outputs from the algorithm and it would be helpful to have these made explicit.

8. While there is growing evidence of GPT in terms of solved cases showing that the use of algorithms can reduce the search areas for police, this still does not necessarily translate into live cases. The Panel acknowledges the accuracy of the NZ findings in terms of solved cases but recommends that care is taken in formulating evaluation criteria and measurements into live cases.

9. The Panel also noted that input data relies on accurate and complete records which from what we know about police data can be problematic. Human agency plays a major role in decision

making, and in particular in potentially overriding decisions based on algorithmic outputs. We understand that there will be capacity for manual overriding of such decisions. Data verification can also be an issue and we know from UK police data that there are myriad examples of missing fields, incorrect entries (especially age and ethnicity) inputting errors. The Panel suggests that a mechanism for at least a proportion of reliability checks on at least a random proportion of the data is developed.

10. It will be helpful to also have clarity around commissioning criteria and what will trigger use of a geographic profile as these may be nuanced depending on type of case e.g. use in serial arson, rape or murder where this may be a risk to life or robbery where there is a risk to property. We therefore recommend that NZP create a diagram of investigatory flows showing where the human decision and the algorithm start and stop; and where (and with whom) commissioning responsibilities lie.

11. While we appreciate that NZP advised that GPT is based on top locations rather than long term histories of potential offenders, the Panel still has concerns that algorithms could over prioritise particular demographics. Location can be used as a proxy for ethnicity, religion, gender and there is a risk that historical oversurveillance and bias may lead to a) biased data sources leading to biased algorithm outcomes and b) human-led suspect selection potentially still being a source of bias.

12. The Panel further recognises that GPT data will come from a range of sources (e.g. places people have committed crime before, places where they have been victim/witness/present at a location for crime or another call for service, places that NZP have noted as frequenting, and places people have been arrested at, etc.). We also noted, however, that NZP acknowledge that these sources often derive from some form of interaction with NZP. The Panel therefore recommends continued caution on minimizing bias against particular groups or individuals, and that any evidence of such bias occurring will require a rigorous mitigation strategy moving forward.

13. The Panel notes the transparency from NZP in their acknowledgement that NZP acknowledged that Māori will be over-represented in the data, and we recognise that there is ongoing Māori/Pasifika liaison to advise on ways to develop process around how an input list is generated in a way that is respectful and pays heed to risk/bias/impacts for Māori. While we do not want to comment on this specifically, we would like to reiterate that it is not easy to make sense from the paper how GPT manages from a Māori perspective and especially how it can offer assurances that it will not end up with over-surveillance. It may be useful for NZP to develop and use metaphors that can help to give assurance that analysis is following tikanga. It will be crucial to bring Māori into this governance of this, especially if doing a live trial in a Māori community.

14. We note that the report states that GP-SMART "was calibrated on NZ data from 2009-2018 and there is no reason to believe the calibrated values need to be updated" (p. 24). It may be worth considering how COVID, and associated migration patterns, may have changed behaviours. We appreciate that there may not be an urgent need to consider such levels of complexity but nonetheless put this forward as something to consider in future.

15. The Panel recommends that NZP carefully consider how they report interpreted **results** and weigh up:

a. Timescales as the information required may be time critical (especially if GPT is rolled out nationally)
b. Good practice in terms of the length of time from commissioning to when report delivered
c. Form of word caveating status of interpretation
d. Creating and disseminating formal terms of reference from commissioning party
e. Maintenance of alternative hypotheses if and when appropriate to help ensure parallel lines of enquiry pursued

16. A final note is simply a reiteration that GPT are probability estimates and provide one input into an investigation which should be taken into account with other evidence.

We did

We incorporated the advice of the Panel, especially where it recommended the development of process and protocols for any eventual wider scale use. We committed to review and evaluation through the processes of the Evidence Based Policing Centre. We engaged the resources of our advisory group on Maori, Pasifika and Ethic services to support us with ongoing guidance. We are currently developing a way to describe the configuration of these algorithms and their use to the public.

We also provided further information to the Panel regarding the type of data being used to allay concerns of particular demographics and the concern of its use as a proxy for protected categories. Lastly, we committed to an evaluation cycle that will collect information on the parameters used to identify input suspects. Where offenders are apprehended for any pilot cases, the evaluation will examine where they ranked on the prioritised suspect lists and whether there is any disparity between demographic groups in the proportion of suspects that ranked higher than the offender (i.e., false positives).