Horizon Scanning Series

The Effective and Ethical Development of Artificial Intelligence: An Opportunity to Improve Our Wellbeing

Trust and Accessibility

This input paper was prepared by Mark Andrejevic

Suggested Citation

Andrejevic, M (2018). Trust and Accessibility. Input paper for the Horizon Scanning Project "The Effective and Ethical Development of Artificial Intelligence: An Opportunity to Improve Our Wellbeing" on behalf of the Australian Council of Learned Academies, www.acola.org.

Submission to ACOLA's project working group on the opportunities and challenges presented by the deployment of AI in Australia and New Zealand

Mark Andrejevic
Professor
School of Media, Film, and Journalism
Monash University
mark.andrejevic@monash.edu

I approach the question of control over the collection and use of personal information based in part on nationwide survey research I conducted in Australia on this question (for discussion of the results, see Andrejevic, 2014). This is one of the few studies that has collected empirical evidence on the topic in Australia and the findings were clear: when it comes to large scale data collection there is strong support (over 90% of respondents in a representative sample) for greater control over personal information and also for more information about how this information is being used. The key issue here is not possession, but control. It is not entirely clear what an individual can do with a database of their own information, beyond poring over it to determine precisely what type of information is being collected about them -- which is interesting, but doesn't get to the key question that seemed to interest most respondents (based on qualitative interviews that followed up on the survey results): how they are affected by the use of this data. It may be possible to see that one's location is being monitored or that one's clicks are tracked, but this does little to explain how information about location or online behavior can be used to predict or influence behavior. Transparency, in this context does not mean simply letting people know that their information is being harvested, it means providing them with a clear idea about *how* it is being used -- a key point with respect to the development of data driven AI systems. By the same token, one's personal data profile, in isolation, does not provide information about how it interacts with the data of millions of other people in ways that reveal patterns unavailable to those who only have access to their own information (Turow et al., 2008). So yes, individuals should have access to their information, but this access can be largely meaningless in terms of accountability without additional information about how it is being put to use. People do not want access to their information for its own sake (it is not valuable to them, per se), but rather to gain some insight into how it might be used. Because of the emergent character of AI decision-making processes, it is not possible to specify in advance the impact that particular forms of data may have on a particular life-impacting decision.

The 2014 survey also indicated strong support (over 90%) for the ability to request that one's personal data be deleted from a particular database. People see their information as something they should be able to have some degree of control over, even

when it is collected in a transactional context. In practices this right depends upon forms of knowledge that are difficult to obtain in the case of third party data collection. It also depends upon a largely outdated conception of personal information (Andrejevic & Burdon, 2015). It is now possible to identify users in meaningful ways (ways that can be tracked back to name, address, and other specific personal information) from the traces they leave on digital platforms (people can readily be identified, for example from their internet search history)(Acar et al., 2014). Individual users can be tracked across the Internet in ways that amount to personal identification even if their names and addresses remain unknown to the trackers. The distinction between personally identifiable information and non-personally identifiable information has become blurred in significant ways. The 'right to be forgotten' may retain some meeting in the case of a search engine like Google, but how does one request to have the record of one's clickstream or browsing history removed, how does one even determine which companies have a copy of it?

When it comes to government and law enforcement access, the question breaks down into inter-related dimensions of data collection. If information about a particular individual is requested, existing restrictions on the collection and use of personal data can be used as a foundation for determining access. However, increasingly, targeted monitoring is replaced by group or classification monitoring: the request to access all information about those who fit a particular behavioral profile. In many cases, this profile may not even contain what is conventionally considered to be personally identifiable information. However, it is often a trivial matter to deduce personal information from non-personally identifiable information (for example, location information might identify one's home address that can then by used to identify particular individuals living in that home via marketing or other databases). This fact poses serious issues for regulation of access because standard protections for personal information rely on the model of targeted information collection. In these cases it might be more appropriate to monitor use than access - that is, to determine which decisions can be made based on data mining and which ones are ruled out. Or, by the same token, a regulatory decision could be taken regarding which types of information are fair game for automated forms of decision-making and which are ruled out. For example, a decision might be taken to rule out the use of genetic information in hiring decisions. Some of these decisions might fall within existing regulatory regimes, to the extent that some classes of information would amount to decision-making based on categories that are protected from discrimination (for example, certain genetic markers might have a high level of correlation with ethnic background and their use in decision-making processes could constitute discrimination on this basis).

There is no such thing as properly informed consent in the world of machine learning and large-scale data mining. This poses a fundamental regulatory, political, and ethical set of issues. By definition, automated systems generate 'emergent' outcomes -- that is they discern patterns and correlations that cannot be deduced in advance (which is the entire point of enlisting such systems). So, for example, a job screening system might determine that the Web browser used to submit a job application correlates more strongly with subsequent job performance than the content of the application. The finding is useful because unanticipated, but cannot be used to inform applicants in advance before the finding is generated. Once the finding is generated, informing applicants renders it useless. Once again, the structural issue here suggests that regulation of use may be more meaningful than the attempt to provide informed consent

(which would read something like: 'all data collected fro this application will be used in conjunction with existing data sets by automated systems to predict future job performance.' This does not lead to meaningful informed consent.

The logic of automated decision-making lends itself to the use of data for unanticipated/non-envisaged purposes. There are large potential benefits to allowing this use -- for example, it might be determined that certain lifestyle patterns can be used to anticipate and intervene pre-emptively in the treatment of some illnesses. Finding these new connections would require speculative data mining. Once again it will likely become necessary to regulate use (by data class or decision class or both -- that is, to say that some forms of data cannot be used speculatively or that some decisions cannot rely solely on AI generated recommendations).

Recent research indicates that it is difficult to effectively anonymise user data. Alternatively put, it is usually possible to reverse engineer or de-anonymise data that has been de-identified (see, for example: Narayanan et al. 2008; Ohm, 2009; Srivatsa et al. 2012). This is likely to become increasingly true as data trails become more comprehensive and multi-dimensional, allowing for new forms of cross-referencing. Protecting data from de-anonymization requires, in addition, reliable protection from data breaches, which remain an ongoing problem for both commercial and governmental data holders. The advent of the "internet of things" and "ubiquitous" computing will lead to burgeoning databases and new vulnerabilities. It seems unlikely, in the near future, that data security will improve dramatically over the situation we have witnessed in recent years. In the near term we can anticipate that more and new kinds of data will be collected for the purposes of machine learning and automated decision-making, generating new stockpiles of data to be targeted for theft.

There is a speculative imperative on the part of data collectors and data miners to hold on to data indefinitely. This was expressed best by the former Chief Technical Officer for the CIA who put it this way: "The value of any piece of information is only known when you can connect it with something else that arrives at a future point in time…Since you can't connect dots you don't have, it drives us into a mode of, we fundamentally try to collect everything and hang on to it forever" (Sledge, 2013). As more data comes online it can be merged with existing data sets to recognize new correlations. Although it is possible to require the deletion of data, the cost of storage continues to decline and the possible future value of an existing dataset is increased by the prospect of ongoing data collection that might reveal otherwise indiscernible patterns.

It is going to be increasingly difficult to regulate data collection because of the proliferation of devices and contexts in which information is gathered and stored. As "smart" spaces and devices come online, they provide new affordances and dimensions for information collection. Smart refrigerators will track eating and food shopping patterns; smart lights will gather detailed information on energy use, sleep patterns, and so on; smart speakers will gather data about the rhythms of domestic life and, eventually, the details of our conversations. The increasingly intensive and extensive forms of digital monitoring is driven in large part by the 'tyranny of convenience': if a device allows us to save time, energy, or money, history indicates that users will readily submit to increasingly comprehensive forms of monitoring. As new dimensions of information collection come online, it will be difficult for regulatory regimes to catch up: should information about an individual's mood, anxiety levels, or emotional expressions by protected? What about their biological responses captured by personal fitness

devices like Fitbit? The key challenge for regulators will be to develop general guidelines that can be used to organize the development of new forms of monitoring as they come online. We might decide, for example, that all biometric information should be off-limits to advertisers because it enables forms of manipulation that go beyond what is acceptable in a democratic society. This is unlikely to happen, but it indicates the type of decision that a society might make in order to set basic guidelines for controlling the implementation of new forms of automated decision making.

The only way to effectively regulate the overseas transfer of the personal data of Australians will be to develop smart networks that can restrict the flow of information via packet sniffing. In the near future, similar systems will be used to restrict the flow of copyrighted material. A digital watermark will be detected by servers, which will refuse transmission of unauthorized data. This will mean requiring that data gathered about Australians be digitally watermarked, which will in turn require a well developed regulatory apparatus governing the growing array of networked, 'smart' devices and infrastructures.

The upshot of these remarks is that the regulatory apparatus will likely have a difficult time keeping pace with the extent and scope of data collection associated with the development of AI, and that there are large challenges to the ethical use of data posed by this technology. Core principles of informed consent, de-identification, and the deletion of data no longer used for its primary purpose are all undermined by the operating imperatives of data-driven, automated decision-making. We will need to develop new regulatory principles and structures to address this epochal shift. These will likely rely on principles for ruling out certain classes of data and categories of decision-making when it comes to relying solely on automated systems. Such principles cannot be made on a case-by-case or individual basis but need to be determined at the societal level.

Works cited:

Acar, Gunes, Christian Eubank, Steven Englehardt, Marc Juarez, Arvind Narayanan, and Claudia Diaz. "The web never forgets: Persistent tracking mechanisms in the wild." In Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security, pp. 674-689. ACM, 2014.

Andrejevic, Mark. "Big data, big questions: the big data divide." International Journal of Communication 8 (2014), pp. 1673-1689.

Andrejevic, Mark, and Mark Burdon. "Defining the sensor society." *Television & New Media* 16, no. 1 (2015): 19-36.

Narayanan, Arvind, and Vitaly Shmatikov. "Robust de-anonymization of large sparse datasets." In *Security and Privacy, 2008. SP 2008. IEEE Symposium on*, pp. 111-125. IEEE, 2008.

Ohm, Paul. "Broken promises of privacy: Responding to the surprising failure of anonymization." *Ucla L. Rev.* 57 (2009): 1701.

Sledge, Matt. 2013. "CIA's Gus Hunt on Big Data: We 'Try to Collect Everything and Hang on to It Forever.'" Huffington Post, March 20. http://www.huffingtonpost.com/2013/03/20/cia-gushunt-big-data_n_2917842.html.

Srivatsa, Mudhakar, and Mike Hicks. "Deanonymizing mobility traces: Using social network as a side-channel." In *Proceedings of the 2012 ACM conference on Computer and communications security*, pp. 628-637. ACM, 2012.

Turow, Joseph, Michael Hennessy, and Amy Bleakley. "Consumers' understanding of privacy rules in the marketplace." *Journal of consumer affairs* 42, no. 3 (2008): 411-424.