Horizon Scanning Series

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

Data Storage and Security

This input paper was prepared by Vanessa Teague and Chris Culnane

Suggested Citation

Teague, V and Culnane, C (2018). Data Storage and Security. 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.

ACOLA Project On AI

Chris Culnane & Vanessa Teague
University of Melbourne
{c.culnane, vjteague}@unimelb.edu.au

August 2018

1 Introduction

The term "AI" isn't very well defined—anything that algorithms can now do, but which human brains previously did, is called "AI." A lot of the questions you have asked relate to data security and privacy issues that don't specifically relate to AI, but would apply in an AI context just as they would in any other.

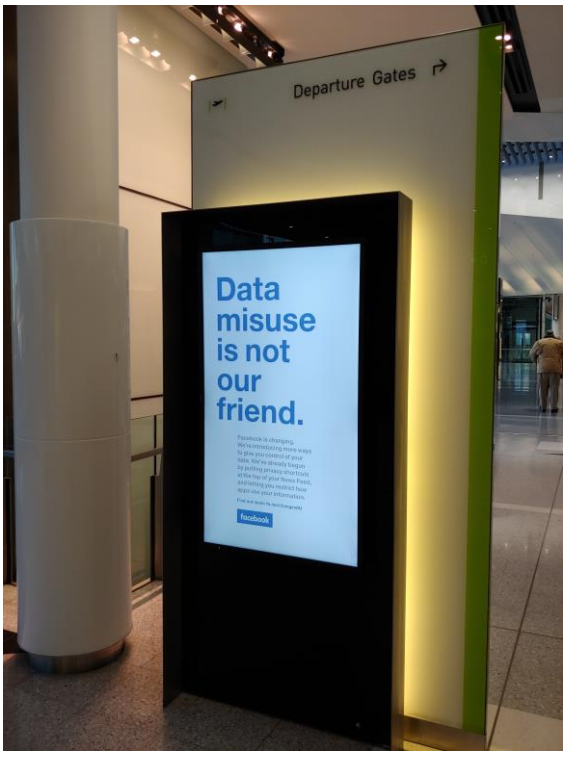

Figure 1: The consequences of losing public trust in data use

2 Questions

2.1 Are there implications for data storage and security from the use of AI?

Yes, in at least four different ways:

Secure storage of the input data. This isn't a problem of AI *per se*, but a problem caused by the massive, and often sensitive, datasets that AI often requires as its input. This issue would exist even if AI were never applied to the data, but becomes even harder to solve when an AI system is dependent on having access to all the data. See Section 2.6.

Secure storage or communication of the outputs which may convey sensitive information from the inputs. See Section 2.6.

Adversarial uses of AI For example, some AIs that do well on random or typical data can malfunction when targeted with deliberately misleading examples [LM05, HJN+11, BCM+13]. This is particularly relevant in adversarial settings such as facial recognition for law enforcement.

Problems of fairness associated with the use of algorithms rather than human decisionmaking[1] [CDPF+17, HBC16, O'N16]. This may be a bias in the algorithm, or a bias in the input data that is reflected in what the algorithm learns and subsequently applies.

2.2 Does AI data storage present implications for different cultural groups within Australia and New Zealand?

I'm sure it does, but you need to get a more expert response from some indigenous data sovereignty experts, such as Prof Maggie Walter from UTas or Prof Tahu Kukutai from U Waikato, NZ.

2.3 Are there implications for the re-identification of anonymised data by AI, through perhaps using a combination of data sets?

Yes. Rather than being a risk of AI, this is a risk of the proliferation of highly detailed data about individuals that AI uses—this risk would exist even if AI was never actually applied. The algorithms used for re-identification can be very straightforward—often just simple data linking [CRT17]. The more data available about an individual, the easier it is to re-identify a record about them.

Of course, this is one of the many things that humans can do already but AI could do very efficiently on a massive scale. AI is highly effective at finding latent patterns in data, which makes it perfect for re-identification.

[1] Of course, human decisionmaking may also be biased and unfair.

The pace of development in AI, and the increasing detail of data gathered about individuals, outstrip the progress of de-identification. This means that datasets get easier to re-identify over time—the risk increases because of a combination of algorithmic progress (in AI etc.) and the increasing availability of auxiliary data.

2.4 Could this be prevented, policy space?

Criminalizing re-identification would simply mean that Australian scientists couldn't examine whether, and notify the Australian government when, a dataset was easily re-identifiable. This would not make re-identification any harder for malicious actors.

For preventing re-identification by corporations such as credit agencies and insurance companies, it would be better policy to enforce some form of *Algorithmic Transparency.* For example, companies (or government agencies) could be required to explain to each customer what data they held about that customer, where they acquired it from, and how they used it to reach a decision about the person.

Opportunities for redress by individuals for breaches of privacy would disincentivize sharing or publication of identifiable data without consent. Penalties for data misuse (such as discrimination or extortion) are also appropriate.

Improved data protection policies are required, which will empower the consumer to be informed of how their data is being used, the conclusions being drawn from it, and a right to access, correct, and delete their data. Due to the covert nature of re-identification, is it also necessary for companies to be able to establish and demonstrate the provenance of the data they use. It should be beholden on them, and therefore indirectly on the supplier of the data, to demonstrate that data was collected with consent and is permitted to be used for the purpose intended.

2.5 Are there issues around offshore vs onshore data storage? Does it actually matter where data are stored?

Yes.

If the data is stored offshore and not end-to-end encrypted, then it has to be assumed that it is easily available to the government of whatever country it is stored in, even if the cloud storage provider offers encryption at rest.

The legal jurisdiction covering the data matters when we do not have globally agreed privacy standards. If the data is ever available in an unencrypted form on the offshore server that presents a problem for effective privacy oversight and may hinder appropriate redress for people whose information is included in the data set.

If it is stored offshore, but end-to-end encrypted, *i.e.* with keys that are held in Australia, then it has to be assumed that the *encrypted* data is available to the government of the country in which it was stored. If the encryption is sound, this may be an acceptable risk, but note that most systems for end-to-end

encrypted file storage expose at least some metadata, such as who accessed what file when. Even for end-to-end encrypted data, some countries are considering laws that would force software companies to provide their government with a secondary mechanism of access to that encrypted data. It is important not to buy encryption software from any countries with such laws.

If it is stored onshore in Australia, then of course this is no guarantee that it will be secure. Data breaches happen all the time, with attackers from within Australia and overseas. End-to-end encrypted cloud storage is one good tool for protecting the data, along with standard mechanisms for secure access and deletion.

2.6 How can we ensure secure storage of data?

Write the only copy onto magnetic tapes and hold them in front of a strong electromagnet.

A far more interesting question is *How can we ensure reasonably secure storage of data while also allowing appropriate access for analysis?*

2.6.1 Technological Solutions

This is a very active area of research, with answers in a few main directions:

- traditional access control,
- Differential Privacy and
- secure multiparty computation.

These are not mutually exclusive—they could all be applied together. For example, a secure research environment, with formally restricted access control, could use Differential Privacy to perturb the answers before showing them to the analyst, and use secure computation for analysis on datasets stored elsewhere.

Secure (multiparty) computation uses cryptography to allow two (or more) computers to evaluate a function on each of their private inputs, without revealing what those inputs are. For example, a set of pharmacists could compute the total number of sales of a particular medication, without revealing their individual sale totals. This does not guarantee that the answer protects privacy: if the computation is an election outcome, and the vote is unanimous, then this reveals exactly how everyone voted. Secure computation has numerous practical applications and has been used in practice by Google [IKN+17], who partnered with a third party to compute the total number of users who had seen an ad and subsequently bought the item in a store. Crucially, they were able to do this without revealing who the customers were, or even how many had seen the ad *or* been to the store.

Secure computation platforms are freely available online [DPSZ12, EFLL12]. Some use (partially) homomorphic encryption, which means that some computations (such as addition) can be performed while the data remains encrypted.

However, their computational speed is limited - some simple computations run quite fast, but more complex machine learning algorithms rapidly become infeasible.

Differential Privacy [DR+14] addresses the complementary problem: it limits the amount of information that can be leaked, by the answer to a query, about any particular individual. In its simplest form, it consists of randomly perturbing the algorithm's output so as to introduce uncertainty about its true value, hence hiding individual details.

In very large datasets, *local differential privacy* can still yield accurate results: each individual input is randomly perturbed first, then the algorithm is applied to the differentially-private data. Both Apple and Google [ACG+16] have run example projects using these techniques, in addition to academic research.

It is important to understand that Differential Privacy is a bound on information leakage, not a guarantee of perfect privacy. If the same data is reused across multiple differentially-private mechanisms, information about individuals can gradually be more accurately inferred.

Combining techniques from cryptography and multiparty computation with Differential Privacy is an active area of research. Many federated data analysis platforms borrow some techniques from each, though not all are designed around rigorous and provable security guarantees.

2.6.2 Policy and regulatory solutions

This is not solely a technology issue. Unless an entity is going to be held accountable for failures to protect and secure data there is little motivation for them to do so. A lack of effective regulation has led to entities showing a complete lack of concern for data security, which has led to numerous breaches. In spite of advances in security, cryptography, and information security management, the number and scale of breaches continues to increase. http://www.informationisbeautiful.net/visualizations/worlds-biggest-data-breaches-hacks/

References

[ACG+16] Martin Abadi, Andy Chu, Ian Goodfellow, H Brendan McMahan, Ilya Mironov, Kunal Talwar, and Li Zhang. Deep learning with differential privacy. In *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*, pages 308–318. ACM, 2016.

[BCM+13] Battista Biggio, Igino Corona, Davide Maiorca, Blaine Nelson, Nedim Šrndić, Pavel Laskov, Giorgio Giacinto, and Fabio Roli. Evasion attacks against machine learning at test time. In *Joint European conference on machine learning and knowledge discovery in databases*, pages 387–402. Springer, 2013.

[CDPF+17] Sam Corbett-Davies, Emma Pierson, Avi Feller, Sharad Goel, and Aziz Huq. Algorithmic decision making and the cost of fairness. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 797–806. ACM, 2017.

[CRT17] Chris Culnane, Benjamin IP Rubinstein, and Vanessa Teague. Health data in an open world. *arXiv preprint arXiv:1712.05627*, 2017. https://arxiv.org/pdf/1712.05627.pdf.

[DPSZ12] Ivan Damgård, Valerio Pastro, Nigel Smart, and Sarah Zakarias. Multiparty computation from somewhat homomorphic encryption. In *Advances in Cryptology–CRYPTO 2012*, pages 643–662. Springer, 2012. Full version at https://eprint.iacr.org/2011/535.pdf.

[DR+14] Cynthia Dwork, Aaron Roth, et al. The algorithmic foundations of differential privacy. *Foundations and Trends® in Theoretical Computer Science*, 9(3–4):211–407, 2014.

[EFLL12] Yael Ejgenberg, Moriya Farbstein, Meital Levy, and Yehuda Lindell. Scapi: The secure computation application programming interface. *IACR Cryptology EPrint Archive* `https://eprint.iacr.org/2012/629.pdf`, 2012:629, 2012. Code and docs at https://cyber.biu.ac.il/scapi/.

[HBC16] Sara Hajian, Francesco Bonchi, and Carlos Castillo. Algorithmic bias: From discrimination discovery to fairness-aware data mining. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, pages 2125–2126. ACM, 2016.

[HJN+11] Ling Huang, Anthony D Joseph, Blaine Nelson, Benjamin IP Rubinstein, and JD Tygar. Adversarial machine learning. In *Proceedings of the 4th ACM workshop on Security and artificial intelligence*, pages 43–58. ACM, 2011.

[IKN+17] Mihaela Ion, Ben Kreuter, Erhan Nergiz, Sarvar Patel, Shobhit Saxena, Karn Seth, David Shanahan, and Moti Yung. Private intersection-sum protocol with applications to attributing aggregate ad conversions. 2017. https://eprint.iacr.org/2017/738. pdf.

[LM05] Daniel Lowd and Christopher Meek. Adversarial learning. In *Proceedings of the eleventh ACM SIGKDD international conference on Knowledge discovery in data mining*, pages 641–647. ACM, 2005.

[O'N16] Cathy O'Neil. *Weapons of math destruction: How big data increases inequality and threatens democracy*. Broadway Books, 2016.