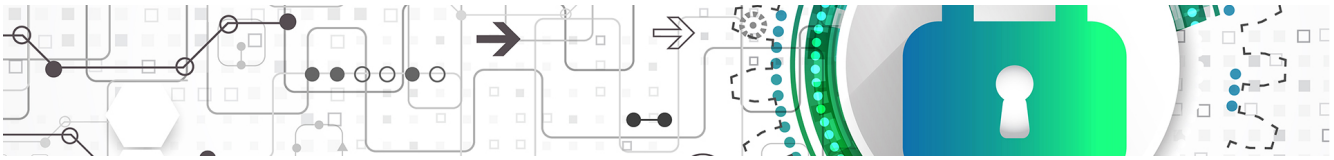

AI & Privacy: PrActIcal AdvIce for OrgAnIzAtIons

November 18, 2019

By Amanda Branch and Jasmine Godfrey

As people offer increasing amounts of personal information, they are becoming more aware of the importance of this information, and its potential uses and misuses. For example, it is not uncommon for young adults applying to university and jobs to change their name on Facebook so admissions committees and recruiters can't track down embarrassing photos of spring break shenanigans. However, in the future, protecting one's data may become more complicated than simply changing one's name on social media—for example, adults may need to trace back to find and delete recordings from their childhood AI-enabled toys.

Organizations look for ways to collect data and to piece it together to analyze, predict and influence consumer behaviour. Organizations may be collecting and using information in a way that is legal, however, we've heard story after story about collections or uses that cross over to "creepy" territory.

In an infamous anecdote from 2012, Target analyzed consumer purchases to create a "pregnancy prediction" score. Much to the surprise of her father, a teenage girl's shopping behaviour resulted in the accurate prediction of her pregnancy, which was inadvertently revealed to her father when Target sent the teenager addressed coupons for maternity clothing and nursery furniture. Interestingly, these small bits of seemingly unrelated data—such as buying larger quantities of unscented lotion and cotton balls—added up to the accurate prediction of an outcome.

The first article in our AI & Privacy series addressed the challenge of obtaining appropriate consent from users and issues related to the retention of data. This article will consider the approach that organizations are taking to manage these issues and conclude with some practical tips for organizations using AI.

Retention of data collected through smart devices

Consumers are increasingly aware of, and sensitive to, the retention of their information, particularly through smart and connected devices; however, despite an organization's best efforts, some consumers will likely remain unaware of what information is being collected and how it is being used, and for a myriad of reasons that may be out of an organization's control. Of course, a lack of transparency on the part of organizations compounds the problem, as do decisions with respect to default settings that are less privacy-oriented.

For example, a Bloomberg report revealed that Amazon's default setting is to retain individuals' conversations with Alexa. Amazon says that it uses these conversations to improve Alexa's "understanding of human speech". These voice snippets are tied to device serial numbers and the owner's first name, and are analyzed by Amazon employees. However, it is possible to turn off the sharing of this information through the Alexa app.

More recently, after consumer outcry in response to reports that humans were listening to audio Siri recordings as part of its quality evaluation process, Apple changed its default settings, announcing that it will no longer retain audio recordings of Siri interactions. Users will be able to opt in to help Apple improve Siri, and those who do can opt out whenever they want. Apple also said that only its own employees, not contractors, will be allowed to listen to audio samples of the Siri interactions, and that any recordings which are deemed to be an inadvertent trigger of Siri will be deleted. These transcriptions are associated with a random identifier, not one's Apple ID, and retained for up to six months. If someone does not want transcriptions of their Siri audio recordings to be retained, one can disable Siri and Dictation in Settings. Apple will still use computer-generated transcripts to help Siri improve.

Inevitably, a tension exists. The more complete the dataset, the better the AI learns. From an AI perspective, data is all

about the “Vs”–volume, variety, veracity, validity and verification–which favours default settings that are inclusive. The “Vs” are important to data hygiene and for data to be contextualized, which may help guard against bias. Further, a default requiring that consumers “opt in” may result in a dataset that excludes consumers that would otherwise be willing to be included, but did not go to the effort to opt-in. Finally, it is important to remember that privacy is about *identifiable* personal information, and there may be middle ground to achieve a more complete dataset, while mitigating the risk of the data being identifiable.

Practical advice for organizations

It is important to remember that, despite these concerns associated with AI’s data retention, AI brings with it many advantages, and not just to the companies that benefit from it, but also to consumers themselves. The positive side of AI is that it has been able to improve products and services by learning what consumers want, but its downside relates to data retention stemming from AI’s inability to “forget” data or information. Consider some practical tips for organizations using AI:

- Practice Privacy by Design by keeping privacy at the forefront throughout the creation and development process. This includes consideration of default settings, along with options to de-identify (pseudonymize or anonymize);
- Provide adequate employee training in the design, function and implementation of the algorithm or automated decision making process to be able to review, explain and oversee its operations;
- Designate an individual(s) who is accountable for the organization’s compliance with ethical guidelines and privacy legislation;
- Consider the impact of an automated decision on the individual, including assessing whether the use may be in violation of the Office of the Privacy Commissioner’s guidelines on inappropriate data practices (read our article here);
- Employ good data governance mechanisms, such as mapping data flows and ensuring the quality and integrity of the data.

The state of the industry presents an exciting opportunity for organizations to take the lead. As with any new class of product or service, it will take time to establish rules, regulations, best practices, and more, but it needs to happen, or consumers will lose trust in AI-enabled devices.