Beyond Personal Information: A Path to Protect Canadians Against Digital Harms

Christelle Tessono
Center for Information Technology Policy
Princeton University

Introduction

In late February, the Canadian government banned TikTok from government mobile devices, following a review from the Chief Information of Canada which found that the application presented an "unacceptable level of risk to privacy and security". However, critics of the ban have called this a "distraction" as these concerns are neither new or unique to TikTok. Researchers at the Citizen Lab published a report analyzing the platform and found that TikTok collects similar types of data to track users and serve targeted ads as other popular social media platforms. This made me wonder: what do platforms know about us? What strategies are they using to collect and analyze our data? But most importantly, what options do we have in Canada to protect against digital harms? In the following memo, I will argue that existing legislative frameworks in Canada cannot address the individual and collective harms raised by platforms because they focus on protecting personal identifiable information instead of all forms of what Teresa Scassa calls *human-derived data* in her piece.

What do platforms do with our data?

Platforms collect data ranging from our phone's geolocation, the content we share and like on Instagram and TikTok, our health information collected from wearable healthcare tech such as a Fitbit, shopping transactions, to our browsing behaviour to name a few examples. As discussed by scholars Linnet Taylor et al. and Graef & van der Sloot, once this information is collected, it is often de-identified and curated to build large databases containing information that reflects the behaviour and activities of users. Then, computational tools are applied to these databases and draw insights from the aggregated data collected to identify patterns, preferences, and behaviors of the groups of people whose data has been collected. As argued by Barocas & Nissenbaum, the computational process of analyzing large databases to generate new information, commonly referred to as *data mining*, "breaks the basic intuition that identity is the greater source of potential harm because it substitutes inference for using identifying information

as a bridge to get at additional facts." In other words, insights drawn from these datasets can provide additional information about an individual or a group, without any personal identifiers.

What are the implications behind the use of these technologies?

At an individual level, it is hard to identify what/when/why/how models are applied and inferences are made about us. At present, researchers, whistleblowers, and journalists are the main routes to uncover these issues, such as the *Wall Street Journal* investigation of Meta Platforms Inc, which revealed that the platform knew about Instagram's negative impact on teenage girls.

At a group level, automated forms of data analytics affect how groups of people are identified. As scholars Lanah Kammourieh et al. note, these systems can identify groups in four different ways. First, they can identify groups and infer information about them without a predefined hypothesis. Second, they can identify groups within a population that had no connection to one another prior to analysis. Third, they can identify groups through new analytical approaches and thus create groups based on previously unknown characteristics. Lastly, these practices might identify groups without analysts' knowledge, thus running the risk of harming people.

What is particularly difficult about such inferences is that these computational analytical tools may discriminate against people by sorting them into groups that do not fall under legally protected categories (e.g. race, gender, disability) and without their personal information being exposed. This makes it difficult for someone to know if they were being profiled and discriminated against. As a result, privacy and data protection legislative strategies that focus solely on protecting identifiable personal information "distracts from, and may even give rise to, problems involving groups profiled anonymously from within huge digital datasets".

What types of group harms emerge from making inferences through these databases? As noted in a report by the Citizen Lab on the collection of mobility data, although databases may contain de-identified or aggregated data, the risk of re-identification remains as it is possible to draw "inferences or correlations from the data or by overlaying it with known personal information." A 2009 study by Harvard professor Latanya Sweeney proved this by re-identifying the names of over 40% participants from a sample of anonymous participants of a DNA study.

Aside from the risk of identification, there is also the risk of surveillance of historically marginalized groups, or even political targeting as we have learned from the Cambridge Analytica scandal. Most significantly, automated decision-making systems deployed to analyze this data tend to misidentify, misclassify, and inaccurately predict outcomes.

How does Canada fare in the face of these challenges?

In terms of data protection and privacy legislation, the Canadian government has two sets of laws. First, the *Privacy Act* which governs the federal government collection, use, disclosure, retention, and disposal of personal information. Second, the *Personal Information Protection and Electronic Documents Act* (PIPEDA), which outlines how the private sector handles personal information during a commercial activity. Provinces and territories have their own laws governing private and public sector usage of personal information, though here I will only discuss federally mandated legislation. Both the Privacy Act and PIPEDA primarily focus on the protection of personal information. The Acts both define personal information as "information about an identifiable individual." which leaves a significant gap around protecting data that is not identifiable.

In June 2022, the Canadian government tabled Bill C-27: *Digital Charter Implementation Act,* which consists of 3 separate Acts. First, the *Consumer Privacy Protection Act* (CPPA) seeks to modernize PIPEDA to adapt to emerging digital technology challenges. Second, the *Personal Information and Data Protection Tribunal Act* looks to create a tribunal to impose penalties for contraventions of the CCPA. Finally, *the Artificial Intelligence and Data Act* (AIDA) seeks to create a statutory framework to "regulate international and interprovincial trade and commerce in artificial intelligence systems by establishing common requirements [...] for the design, development and use of those systems." Regarding data protection, the CPPA is different from PIPEDA as it introduces provisions on data de-identification, deletion, and children's protection. More specifically, it defines de-identification as the "means to modify personal information so that an individual cannot be directly identified from it, though a risk of the individual being identified remains." Moreover, the CPPA seeks to provide safeguards for minors by considering their personal information as sensitive. Yet, these acts leave much to be desired.

How do we move forward?

To fight against emerging digital harms, the Canadian government should modernize privacy legislation and expand protections to non-identifiable information. This would involve implementing the following amendments to Bill C-27:

1. Protections for all human-derived data, which includes personal information, de-identified and anonymized data as Teresa Scassa proposes.
2. Instate a prohibition on the re-identification of de-identified data, as recommended by the parliamentary study on the use of mobility data during the COVID-19 pandemic.
3. Empowering the Office of the Privacy Commissioner of Canada to enforce both public and private sector privacy laws, investigate breaches, draft regulation, and audit companies.
4. Defining in the CPPA what constitutes a 'legitimate commercial interest' and 'public good' in the collection, storage, use, transfer, and sale of private data, as recommended by the parliamentary study on the use of mobility data during the COVID-19 pandemic.

Furthermore, given that emerging technologies relying on AI systems heavily infringe on privacy and other numerous human rights, the proposed AIDA needs significant improvement. The government should look to establish a robust independent regulatory framework by providing the Office of the Privacy Commissioner of Canada with adequate powers to enforce the law and craft sector-specific regulation. Moreover, we need a statutory framework that addresses the core human rights risks of algorithmic systems. This would include, but not be limited to, establishing clear limitations and guidelines on the design and development of algorithmic systems that:

1. Impact the health and financial outcomes for individuals and communities.
2. Are used to access social services or humanitarian aid.
3. Are used to profile and influence peoples' behaviour.
4. Use biometric or health-related bodily information to uniquely identify and categorize people.

With Bill C-27 being presently debated at the House of Commons, the government has a unique opportunity to enact a legislative framework that not only protects Canadians against digital harms, but ensures the safe and equitable development of digital technologies.

Christelle Tessono is a tech policy researcher and Emerging Scholar at Princeton University's Center for Information Technology Policy. *She thanks the participants of the February 2023 workshop on platform governance for their generous feedback on an earlier version of this piece.*