Why Canada's Artificial Intelligence and Data Act Needs "Mental Data"

Dylan J. White & Joshua August Skorburg

Published online: 25 Apr 2023.

Submit your article to this journal ⬈

Article views: 90

View related articles ⬈

View Crossmark data ⬈

it comes to many of the types of mental data that exist outside of our heads. Nonetheless, Palermos would have us protect, *beyond the power of a warrant*, data that people neither expect to be kept private nor which society expects to be kept private. It is a crime to possess at least some data, e.g. child abuse materials, because the possession of that data is an "integral" part of committing another crime, e.g. child abuse (Osborne v. Ohio 1990). Most would agree that mental data, like regular data, ought to be protected until a warrant is presented. And, sure, the data inside our embodied minds is protected as a technological matter right now. But, were it technologically possible for the government to read our minds, I imagine many people would accept that that they should also be able to legally obtain evidence of child abuse, or other similarly heinous crimes from your mind. That claim, certainly, is contentious. But, I hope now with all the above

discussion that the falsity of Palermos's claim, that all data *ever* included in the "extended mind" should be absolutely protected, is not.

FUNDING

REFERENCES

Clark, A. 2003. *Natural-born Cyborgs: Why minds and technologies are made to merge.* New York, NY: Oxford University Press.

Katz v. United States, 389 U.S. 347. 1967.

Osborne v. Ohio, 495 U.S. 103. 1990.

Palermos, S. O. 2023. Data, metadata, mental data? Privacy and the extended mind. *AJOB Neuroscience* 14 (2):84–96. doi:10.1080/21507740.2022.2148772.

Taylor & Francis
Taylor & Francis Group

Check for updates

OPEN PEER COMMENTARIES

Why Canada's Artificial Intelligence and Data Act Needs "Mental Data"

Dylan J. White (ID) and Joshua August Skorburg (ID)

University of Guelph

By introducing the concept of "mental data," Palermos (2023) highlights an underappreciated aspect of data ethics that policymakers would do well to heed. Sweeping artificial intelligence (AI) legislation is on the horizon. To name just a few: In April 2021, the European Union (EU) proposed the AI Act that encompasses all sectors (save the military), and all manner of AI. In June 2022, Canada tabled Bill C-27 which, alongside new proposals for consumer privacy, contains the Artificial Intelligence and Data Act (AIDA). Most recently, the United States established a Blueprint for an AI Bill of Rights.

None of the proposed legislation, however, considers the harms to individuals or groups that might result from access to, disclosure of, or manipulation of, mental data. We concur with Palermos (2023) that this is a pressing problem that needs to be addressed.

In this commentary we focus on Canada's AIDA, though our arguments generalize to other regulatory proposals for AI.

Canada's AIDA sets out to regulate "high-impact" AI systems, and designates "high-impact" based on the harm they could cause. This risk-based approach entails distinct regulatory requirements based on the level of assessed risk. Its stated purpose is "to prohibit certain conduct in relation to artificial intelligence systems that may result in serious harm to individuals or harm to their interests" (*Artificial Intelligence and Data Act, 4(b)*). In Section 5(1), harm is defined as:

a. Physical or psychological harm to an individual;
b. Damage to an individual's property; or
c. Economic loss to an individual

CONTACT Dylan J. White ✉ dwhite11@uoguelph.ca 🖳 University of Guelph, Guelph, N1G 2W1, Canada.

Physical harms, damage to property, and economic loss may be more-or-less quantifiable, but what are we to make of psychological harm? Recent work on the ethics of the attention economy (Castro and Pham 2020) has identified psychological harms such as nudging, manipulation, and attention commodification in many applications powered by AI (especially social media). Yet, these harms are notoriously hard to pin-down and often rest on unsophisticated empirical accounts of human psychology (White under review). By proposing a new subset of data categorization called "mental data," Palermos (2023) complicates this already fraught landscape by proposing to include "the non-neuronal parts of extended minds" as additional sites of potential psychological harm. This complication is necessary, however, because if Palermos (2023) is on the right track, then psychological harm at the hands of AI is more widespread than the proposed legislation can accommodate.

Is Palermos on the right track? We think so. Skorburg (2019) and many others have argued that smartphones and other similar technologies already constitute the kind of bidirectional coupling necessary for an artifact to count as *mind extension*. This insight uncovers a previously undertheorized way in which agents might be harmed: by having their mental data accessed, disclosed, or manipulated, and so their autonomy compromised.

Our concern is that Canada's AIDA and other proposals are ill-equipped to recognize and address the harms connected to mental data. For example, mistaking what should be properly categorized as mental data for metadata may result in AI applications being mislabeled as low-impact. Consider a dieting app. The collection and use of metadata about everyday eating occasions such as time, location, social context, and temporal proximity to exercise may not be considered high-risk enough to label the system high-impact. However, as Biel et al. (2021) show, this metadata can be used to accurately discriminate eating habits (meals vs. snacks, binging, fasting, etc.) and associated psychological states. AI-enabled dieting apps such as Nutrino use such metadata to make personalized recommendations, potentially becoming bidirectionally integrated with cognitive processes (e.g., deciding, planning, remembering, evaluating) associated with shopping, cooking, eating, exercising, etc. In other words, the process of self-regulation can *extend to include the app*. Given this level of potential cognitive integration, the relevant data should be counted as mental, not meta-, data.

How does reconceptualizing (some) metadata as mental data affect the AIDA's high-impact calculus?

As Palermos (2023) remarks, "mental data are essentially no different to the data stored in the neural web of our brains" (90) Because we agree that mental privacy is essential for foundational values like mental autonomy and psychological individuality (Palermos 2023), we see the access to, disclosure of, and manipulation of mental data enabled by the bidirectional coupling with these apps worthy of a high-impact designation because of the associated risks. This is not to say that *mind extension* is bad, but rather is a double-edged sword, capable of both enhancing and undermining autonomy.

For example, dieting apps may enable better self-regulation and promote healthy eating habits, but for some, they may also result in unhealthy outcomes. The level of consumption moderation and associated recommendations on dieting apps can enhance rigidity and anxiety about caloric intake and disordered eating (Simpson and Mazzeo 2017). Articulated this way, an app that creates and uses mental data to monitor and manipulate eating habits should be appropriately labeled high-risk and subject to the corresponding regulatory requirements. As it stands, the AIDA is unequipped to deal with this manner of potential psychological harm. When these apps (and other AI systems) are appreciated as *mind extensions* that create and use *mental data*, the nature, prevalence, and seriousness of these harms becomes much clearer.

At this point, one might object that philosophical arguments about extended minds and mental data are too messy and contentious to be properly integrated into policy. What criteria should be used to establish when the kind of bidirectional coupling necessary for mental data is present in a technological interaction? How do we quantify the psychological harms associated with accessing mental data? These are tricky questions that need to be addressed to successfully integrate a concept of mental data into AI policy. That said, we do not need to provide definitive answers to prevent potential harms. There is already precedent for using philosophically complex concepts for effective public policy.

Take sentience, a philosophically contested concept if there ever was one. Sentience (or related concepts derived from it), is being deployed widely for animal welfare and rights policy, including the proposed Universal Declaration on Animal Welfare (UDAW) and the United Kingdom's (UK) Animal Welfare (Sentience) Act. Dawkins (2022) argues that because sentience is a philosophically and scientifically contested term, it should be separated from policy. This view has its merits, but including sentience in animal

welfare policy considerations, despite its contested nature, has clear upsides as well. In 2021, Jonathan Birch and colleagues developed a report for the UK government based on their findings that all cephalopod mollusks and decapod crustaceans should be considered sentient. The report identified the clear evidence gaps, but drew the conclusion that, based on the scientific findings, these animals could be reasonably assumed to be sentient. As a result, the UK's Animal Welfare (Sentience) Act included both taxa. Thus, denying sentience as relevant to policy because of its contested nature could have led to the exclusion of many animals from certain protections.

What are the implications for the concept of mental data and the AIDA? First, we need to be rigorous, both conceptually and empirically, to clearly define what constitutes *mind extension* and *mental data*. Theoretical and methodological frameworks for empirically testing extended cognitive systems have been recently proposed (Favela et al. 2021). Further, uncertainties and evidence gaps should be readily acknowledged. As with sentience, we do not need an airtight, once-and-for-all settled definition of *mental data* for it to be put to good policy use.

The AIDA needs to be able to accommodate the kinds of psychological harms that extended minds make possible. Because the AI functionality of many of the tools that we use on a day-to-day basis such as smartphones allows for the kind of bidirectional coupling outlined above, these psychological harms are already taking place and are likely to become much more common in the foreseeable future.

How should AIDA do this? Taking a broadly functionalist stance would be a good start. As Carter and Palermos (2016) note, there is existing legal precedent for doing so. A 1994 United Kingdom decision saw bodily injury defined as "injury to any of those parts of his body *responsible for his mental and other faculties*" (Lord Hobhouse, *Regina v. M. Chan Fook*).

One need only substitute "psychological harm" for "bodily injury" and add "parts of the environment" to "parts of the body" to begin to address these shortcomings of AIDA. Coupled with rigorous empirical and theoretical methods for determining when something constitutes *mental data*, decisions can be reasonably made regarding when compromising someone's smartphone or other personal technology should count as a psychological harm.

In doing so, the scope of AIDA and its ability to capture meaningful harms at the hands of AI systems would become that much more robust and effective. Crucially, more careful work is required to integrate a concept like *mental data* into AIDA, but the examples and argument outlined here provides a general framework for beginning to do this essential work.

FUNDING

ORCID

Dylan J. White ⓘ http://orcid.org/0000-0001-9235-2991
Joshua August Skorburg ⓘ http://orcid.org/0000-0002-3779-5076

REFERENCES

Biel, J.-I., N. Martin, D. Labbe, and D. Gatica-Perez. 2021. Bites'n'Bits: Inferring eating behavior from contextual mobile data. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2 (4): 1–33. doi:10.1145/3161161.

Bill C-27, An Act to enact the Consumer Privacy Protection Act, the Personal Information and Data Protection Tribunal Act and the Artificial Intelligence and Data Act and to make consequential and related amendments to other Acts, 1st Sessions, 44th Parliament, 70-71 Elizabeth II, 2021-2022. https://www.parl.ca/DocumentViewer/en/44-1/bill/C-27/first- reading

Birch, J., C. Burn, A. Schnell, H. Browning, and A. Crump. 2021. *Review of the evidence of sentience in cephalopod molluscs and decapod crustaceans*. London, UK: The London School of Economics and Political Science. https://www.lse.ac.uk/business/consulting/assets/documents/Sentience-in-Cephalopod-Molluscs-and-Decapod-Crustaceans-Final-Report-November-2021.pdf

Carter, J. A., and S. O. Palermos. 2016. Is having your computer compromised a personal assault? The ethics of extended cognition. *Journal of the American Philosophical Association* 2 (4):542–60. doi:10.1017/apa.2016.28.

Castro, C., and A. K. Pham. 2020. Is the attention economy noxious? *Philosophers' Imprint* 20 (17):1–13.

Dawkins, M. S. 2022. *The science of animal welfare: Understanding what animals want*. Oxford: Oxford University Press.

Favela, L. H., M. J. Amon, L. Lobo, and A. Chemero. 2021. Empirical evidence for extended cognitive systems. *Cognitive Science* 45 (11): e13060. doi:10.1111/cogs.13060.

Palermos, S. O. 2023. Data, metadata, mental data? Privacy and the extended mind. *AJOB Neuroscience* 14 (2):84–96. doi:10.1080/21507740.2022.2148772.

Simpson, C. C., and S. E. Mazzeo. 2017. Calorie counting and fitness tracking technology: Associations with eating disorder symptomatology. *Eating Behaviors* 26:89–92. doi:10.1016/j.eatbeh.2017.02.002.

Skorburg, J. A. 2019. Where are virtues? *Philosophical Studies* 176 (9):2331–49.

White, D. J. Under review. Paying attention to attention: Psychological realism and the attention economy. Manuscript under review.