

Intersections Between Government Data and AI Strategies: A Case Study of Technology Policies in Canada's Federal Service

AUTHORS SECTION

Mahetaji, Kaushar

University of Toronto, Canada | kaushar.mahetaji@mail.utoronto.ca

Zogheib, Ciara

University of Toronto, Canada | ciara.zogheib@mail.utoronto.ca

Spencer, Ryan

Canada | ryanwalterspencer@gmail.com

ABSTRACT

AI and data are mutually influential, with AI outputs shaped by training data and data often generated, processed, and categorized by AI. The use of both AI and data by government organizations is guided by policy documents; existing research has explored data policies *or* AI policies but has rarely put both in conversation, despite their linked subject matter. We adopt a mixed-methods approach to analyze the data and AI strategies of the Government of Canada, investigating whether the data-AI relationship is reflected in policy documents. Our findings demonstrate a disconnect between Canadian data and AI policies, illustrate potential implications of this disconnect, and contribute to ASIS&T 2025 conversations about the necessity of information science for the responsible, ethical use of data and AI in government settings.

KEYWORDS

Artificial Intelligence; Data Governance; AI Governance; Policy Development; Content Analysis

INTRODUCTION

Governments produce vast amounts of data. The ways in which these data are managed are guided by data policies and strategies. Frequently cited data policies include the *California Consumer Privacy Act*, put into force by the State of California in 2020, and the European Union's 2018 *General Data Protection Regulation (GDPR)* and 2022 *Data Governance Act* (State of California, 2024; European Union, 2018; European Commission, 2024). Likewise, governments publish AI policies to guide AI development and implementation, such as the European Union's 2024 *AI Act*, deemed "the world's first comprehensive AI law" (European Parliament, 2025) and the 2019 *Model AI Governance Framework* in Singapore (PDPC Singapore, 2020). While the data and AI governance literature tends to concentrate on the American case and the stricter, standard-setting policies implemented in Europe, governing AI and data constitutes a global effort. Here, we underline the significance of other regional contexts, specifically Canada..

Considering Canada's position as "the first country in the world to create a national strategy for AI" (Government of Canada, 2025a), the Canadian context is noteworthy. Canada has established numerous international collaborations for regulating data and AI (see Government of Canada (2020) for an example) and has introduced data and AI policies comparable to those developed in the United States and Europe. In this paper, we focus on the two most thorough and centralized Canadian policies at the federal level: the *2023–2026 Data Strategy for the Federal Public Service* (Government of Canada, 2023) and the *AI Strategy for the Federal Public Service 2025–2027* (Government of Canada, 2025b). With these new strategy documents, the Government of Canada continues a global trend: with the technology sector evolving rapidly, governments are heavily prioritizing data and AI in their policymaking. What is less apparent is whether these policies are in dialogue.

The two policy areas—data and AI—are inextricably linked, both discursively and technically. For example, the outputs of AI models are shaped by training data and subject to the biases contained in these data, and increasingly, AI is being used to process, generate, and categorize data (Leavy et al., 2021; Scheuerman et al. 2021). Existing research (Attard-Frost et al., 2023) has explored data policies and AI policies but has rarely put both in conversation; that is, whether the well-established mutual influence between data and AI is reflected in policy has yet to be unpacked.

To respond to this open question, this paper investigates connections between data and AI strategy documents in the Canadian government context. We use our findings to prompt discussion about the government usage of AI, and the vital role that information scholars and practitioners must play in ensuring that this usage is responsible and ethical.

METHODS AND THEORETICAL BASIS

This initial work adopts a mixed-methods approach to policy document analysis to better interrogate the relationship between data and AI governance at the federal level of government in Canada. We narrow our data sources to two central, publicly available policies: the *2023–2026 Data Strategy for the Federal Public Service*, released in 2023,

and the *AI Strategy for the Federal Public Service 2025-2027*, introduced two years later. This paper refers to these documents as “the Data Strategy” and “the AI Strategy” respectively from this point forward.

Quantitative Document Analysis

The methodology applied for quantitative document analysis combines descriptive statistics (such as text frequency), Term Frequency-Inverse Document Frequency (TF-IDF) cosine similarity (to assess word-level similarity), and topic modelling (to assess sentence- and topic-level similarity), to gauge the alignment between the *Data Strategy* and *AI Strategy* documents. We follow the precedents of Wang & Dong (2020) and Qaiser et al. (2018). Both documents are scraped and preprocessed by lowercasing the corpora and removing erroneous whitespace and English stopwords (e.g. ‘and’) prior to analysis.

Qualitative Document Analysis

For a systematic qualitative review of the *Data Strategy* and the *AI Strategy*, we adapt thematic analysis from Braun and Clarke (2006). Our thematic analysis first makes note of direct mentions of the terms “data” and “artificial intelligence” (or “AI”) and then records references to other related government policies and press releases. We move on to identify recurring motifs, beliefs, expectations, and agendas that speak to the development, implementation, and management of data and AI. This final step uncovers how the federal government employs policy to build shared visions of the current and future uses and impacts of data- and AI-based technologies—or the “sociotechnical imaginaries” of data and AI (Jasanoff & Kim, 2009, p. 120).

To inform our critical analysis of the connections between data and AI, we borrow from science and technology studies to apply the framework of “sociotechnical imaginaries,” surfacing the “collectively imagined forms of social life and social order reflected in the design and fulfillment of nation-specific scientific and/or technological projects” (Jasanoff & Kim, 2009, p. 120). Emerging studies of data and AI governance emphasize the methodological value in mapping sociotechnical imaginaries for revising existing policies and envisioning alternate realities. For instance, in a comparative study of AI strategy policy documents from China, the United States, France, and Germany, Bareis and Katzenbach (2022) pinpoint common framings of AI as part of linear progress, as inevitable, and as necessary for economic growth and competition. Feitsma and Whitehead too engage in comparative work, delineating future imaginaries in AI policy documents from the United Kingdom and the Netherlands. The authors show how the two states diverge, with the United Kingdom positioning AI as a solution, and the Netherlands being more focused on regulating AI technologies. Simultaneously, the authors highlight how both governments are similarly reactive with their policies, insinuating that government power over AI may be limited as AI is increasingly shaped by a “technological elite.” Other studies of data and AI imaginaries built through policy also explore the social, cultural, and political narratives governments construct to drive AI adoption, and the associated consequences for the ethical and effective integration of AI across societal sectors (Hälterlein, 2024; Hoff, 2023; Köstler & Ossewaarde, 2022; Mager & Katzenbach, 2021; Paltieli, 2022). Our empirical review of Canada’s *Data Strategy* and *AI Strategy* contributes to this burgeoning work on governmental imaginaries.

Positionality

Two of the authors have prior professional experience as data scientists at the Government of Canada. No internal documents are consulted or referenced here, and neither author contributed to either of the policy documents being analyzed in any capacity.

FINDINGS

Common Guiding Principles and Priorities

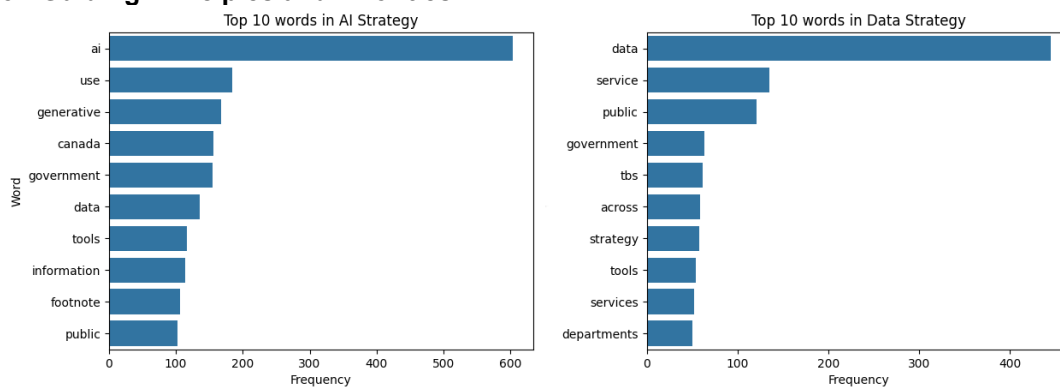


Figure 1. Word frequency plots for the AI Strategy (left) and Data Strategy (right); plots show word-level similarity.

We find that the *Data Strategy* and the *AI Strategy* share common policy and legal bases, both citing Canada’s *Digital Ambition*, the Government of Canada *Digital Standards*, the *Policy on Service and Digital*, and assorted policies relating to *Open Government* and the *Privacy Act*. Both include privacy, public service delivery, and Indigenous data sovereignty as priorities, and focus on the development of organizations (such as the AI Centre of Expertise) and roles (such as the Chief Data Officer) as a primary means of working towards those priorities (see Figure 1).

In short: we can observe that the *Data Strategy* and the *AI Strategy* have high-level and structural similarities, suggesting some level of synchronicity, but delving into the substantive content of both documents in greater detail challenges this assumption.

Data and AI—Unidirectional Influence?

The relationship between data and AI is bidirectional—that is, data can influence AI outcomes, but AI can also influence data. The strategy documents reviewed here do not reflect this reality.

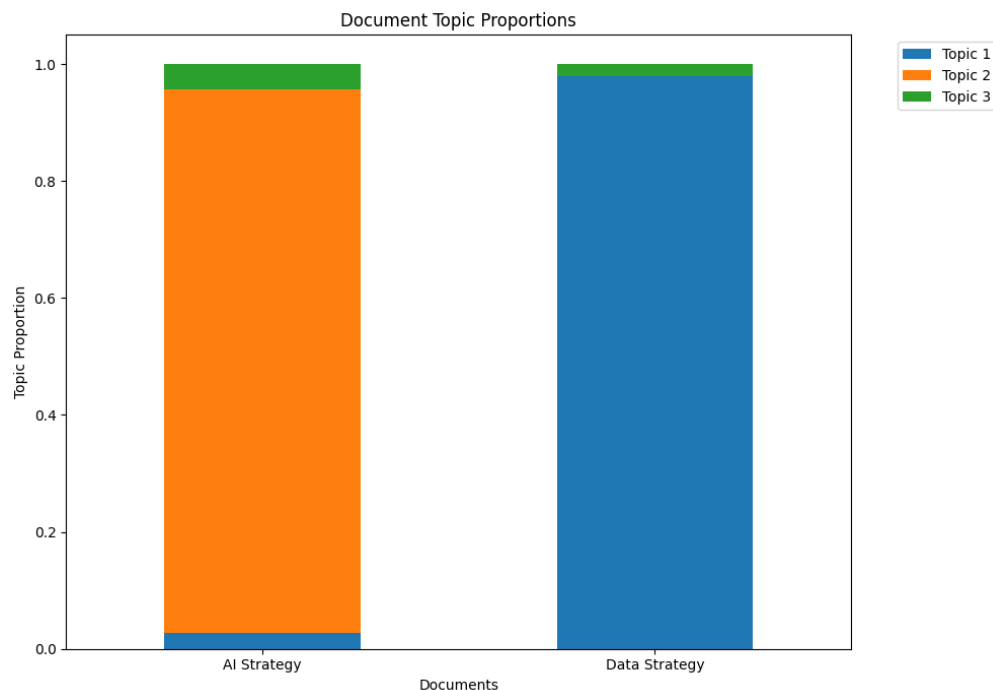


Figure 2. Proportions of topics in both documents; when semantic context is taken into account, the *AI* and *Data Strategies* are very dissimilar. Note: Topic 3 contains outlier topics that are neither data- nor AI- related.

TF-IDF vectorization of both documents with cosine similarity yielded a score of 0.81, suggesting lexical similarity on a superficial level (that is, both documents use similar words). However, as topic modelling reveals, this similarity does not necessarily translate to thematic alignment: comparison of topic vectors (that is, words as used in sentence context, not just individual words) resulted in a cosine similarity score of 0.03, indicating minimal overlap in semantic content and topics (see Figure 2). For example, when the *Data Strategy* discusses goals, it emphasizes the creation of guidelines and standards to ensure data quality; in the *AI Strategy*, on the other hand, goals are focused more on final outcomes and impacts than on systems themselves, or their governance and quality. That is to say: the *Data* and *AI Strategies* had less in common than might have been expected given the close relationship between data and AI, a relationship that (we can infer) should reasonably also exist between related strategy documents. Finer-grained analysis reveals part of the reason for this disconnect: the *AI Strategy* mentions “data” 136 times and references the *Data Strategy* as a “critical enabler for the success of the AI Strategy”. Conversely, in the *Data Strategy*, “AI” (or “artificial intelligence”, or other related terms) is not mentioned at all. The lack of direct reference to the *AI Strategy* in the *Data Strategy* is explained by the publication dates of both documents (2025 and 2023, respectively), but the complete absence of *any* plans for AI in the *Data Strategy* is potentially concerning.

With the advent of publicly available large language models and other generative AI tools, governments must prepare for the eventuality of interacting with AI-generated and synthetic data. For the Government of Canada, this eventuality is an inevitability: the *AI Strategy* describes a “lighthouse project” where “AI-driven language tools” will be used to generate translations of internal data and documents. These are not neutral decisions. Data and information scholars have documented risks of AI-generated data, including threats to data integrity, the

perpetuation of biases, and reduced transparency and accountability when synthetic data provenance is not clearly documented (Bail, 2024; Blau et al., 2024; Davison et al., 2024; Li et al., 2025). A comprehensive, current data strategy must explicitly consider best practices for producing, governing, and using AI-generated and synthetic data.

Unsteady Foundations

In addition to the disconnect in content between the *Data* and *AI Strategies*, we raise a potential inconsistency between both documents on an implementation level.

The *Data Strategy* identifies several ongoing, as-yet-unresolved issues with the Government of Canada's current data governance, infrastructure, and collection practices; per the *Strategy*, "there is still work to be done" to improve the Government's data holdings ("Missions"). The *AI Strategy* discusses the need to "support data readiness" as part of its priority areas, but other than that, does not engage with the data issues raised in the *Data Strategy*, and in fact takes the usage of these data for AI projects for granted. The "think AI" approach advocated by the *AI Strategy* ("Priority 2") challenges departments to identify AI use cases and "consider AI options before defaulting to traditional IT or HR approaches"—in other words, the *AI Strategy* document takes the stance that the Government will be utilizing its data holdings to implement AI solutions, even if (as the *Data Strategy* makes explicit), there are lingering concerns about data quality, ethics, and robustness.

We do not argue the necessity of governments developing AI strategies and research infrastructure. These are necessary to protect citizens and efficiently serve the public good (Hjaltalin & Sigurdarson, 2024). We do raise the question: is it responsible for governments to plan for and implement AI projects, knowing that training and testing data are flawed or biased? Information scholars at ASIS&T and other venues have described how the use of inadequate data can fail vulnerable populations and reduce institutional accountability (Arora et al., 2023; Dickey et al., 2024; Wickett et al., 2024). Doing so risks outcomes such as chatbots that offer contradictory information (Garcia, 2024), diagnostic tools that reproduce historical racial biases (Heldreth et al., 2024), and recruitment strategies that disproportionately affect marginalized communities (Chen, 2023). We emphasize that organizational eagerness to capitalize on emerging AI and data-driven technologies must not overlook the many social, ethical, cultural, political, and economic consequences of these technologies; and caution policymakers and developers of data and AI policies from utilizing the rhetoric of inevitability and mass technological disruption common to AI imaginaries, articulated by Bareis and Katzenbach (2022), as reasoning to rush into AI implementation before a solid foundation of data and information management, evaluation, and equity is operational.

CONCLUSION

Government conversations on digital governance took a discursive turn toward AI with the 2022 release of OpenAI's generative AI-based chatbot ChatGPT. While AI strategies appear to be new, the impacts and challenges at the heart of AI governance talks are not. Rather, they closely resemble the consequences and challenges linked to the governance of data. The strong connections between data and AI, and the relevance of these connections for policy development, are rarely explored in the data and AI literature. This paper responds with an exploration of the Government of Canada's *Data* and *AI Strategies*, bringing to the fore governmental imaginaries of data and AI, including intersections and areas of disconnect in data and AI governance.

In future studies, we will expand this work to factor in the differences in legislative authority by jurisdiction, as enumerated in Canada's *Constitution Act (1867)*. Taking a comparative stance, we will explore how data and AI strategies vary by jurisdiction, supplementing the analysis of data- and AI-related policies described in this paper for the federal level (Canada) with policy documents from the provincial level (Quebec) and the municipal level (Montreal). We focus on Montreal, Quebec because of its unique position as a tech hotspot for data and AI innovation; Montreal is celebrated as "the second global AI hub after Silicon Valley" (Turkina & Oreshkin, 2022, p. 673), bringing together the largest number of AI experts in the world and housing over 5000 technology firms and 500 companies specific to AI (Investissement Québec 2024). This cross-jurisdictional study will be significant as data and AI governance involve the collaborative effort of all levels of government.

Our case study of Canada's *Data* and *AI Strategies* ultimately recognizes the importance of regionally-informed discourse about responsible AI governance, and of positioning AI governance and imagined futures within larger conversations of data and digital governance.

GENERATIVE AI USE

We confirm that we did not use generative AI tools/services to author this submission.

AUTHOR ATTRIBUTION

Ciara Zogheib: Conceptualization, Formal Analysis, Writing – original draft, Writing – review and editing; Kaushar Mahetaji: Conceptualization, Formal Analysis, Writing – original draft, Writing – review and editing; Ryan Spencer: Conceptualization, Formal Analysis, Writing – review and editing

REFERENCES

- Arora, A., Barrett, M., Lee, E., Oborn, E., & Prince, K. (2023). Risk and the future of AI: Algorithmic bias, data colonialism, and marginalization. *Information and Organization*, 33(3), 1–7. <https://doi.org/10.1016/j.infoandorg.2023.100478>
- Attard-Frost, B., Brandusescu, A., & Lyons, K. (2024). The governance of artificial intelligence in Canada: Findings and opportunities from a review of 84 AI governance initiatives. *Government Information Quarterly*, 41(2), 1–24. <https://doi.org/10.1016/j.giq.2024.101929>
- Bail, C. A. (2024). Can Generative AI improve social science? *Proceedings of the National Academy of Sciences*, 121(21), 1–10. <https://doi.org/10.1073/pnas.2314021121>
- Bareis, J., & Katzenbach, C. (2022). Talking AI into Being: The Narratives and Imaginaries of National AI Strategies and Their Performative Politics. *Science, Technology, & Human Values*, 47(5), 855–881. <https://doi.org/10.1177/01622439211030007>
- Blau, W., Cerf, V. G., Enriquez, J., Francisco, J. S., Gasser, U., Gray, M. L., Greaves, M., Grosz, B. J., Jamieson, K. H., Haug, G. H., Hennessy, J. L., Horvitz, E., Kaiser, D. I., London, A. J., Lovell-Badge, R., McNutt, M. K., Minow, M., Mitchell, T. M., Ness, S., ... Witherell, M. (2024). Protecting scientific integrity in an age of generative AI. *Proceedings of the National Academy of Sciences*, 121(22), 1–3. <https://doi.org/10.1073/pnas.2407886121>
- Braun, V., & and Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77–101. <https://doi.org/10.1191/1478088706qp063oa>
- Chen, Z. (2023). Ethics and discrimination in artificial intelligence-enabled recruitment practices. *Humanities and Social Sciences Communications*, 10(1), 1–12. <https://doi.org/10.1057/s41599-023-02079-x>
- Davison, R. M., Chughtai, H., Nielsen, P., Marabelli, M., Iannacci, F., Van Offenbeek, M., Tarafdar, M., Trenz, M., Techatassanasoontorn, A. A., Díaz Andrade, A., & Panteli, N. (2024). The ethics of using generative AI for qualitative data analysis. *Information Systems Journal*, 34(5), 1433–1439. <https://doi.org/10.1111/isj.12504>
- Dickey, T. J., Dobreski, B., Hlava, M., Lund, B., Needleman, M., & Zeng, M. L. (2024). Information Standards and Guidelines on AI: Ethical Concerns, Trustworthiness, Quality Assessment, and Human Oversight. *Proceedings of the Association for Information Science and Technology*, 61(1), 745–748. <https://doi.org/10.1002/pra2.1092>
- European Commission. (2024, October 10). *European Data Governance Act*. Shaping Europe’s Digital Future. <https://digital-strategy.ec.europa.eu/en/policies/data-governance-act>
- European Parliament. (2025, February 19). *EU AI Act: First regulation on artificial intelligence*. Digital. <https://www.europarl.europa.eu/topics/en/article/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence>
- European Union. (2018). *General Data Protection Regulation (GDPR)*. GDPR.EU. <https://gdpr.eu/tag/gdpr/>
- Feitsma, J., & and Whitehead, M. (2025). Flattened futures: (Strangely) governing uncertain AI futures in the UK and the Netherlands. *Policy Studies*, 1–19. <https://doi.org/10.1080/01442872.2025.2454484>
- Garcia, M. (2024, February 19). What Air Canada Lost In ‘Remarkable’ Lying AI Chatbot Case. *Forbes*. <https://www.forbes.com/sites/marisagarcia/2024/02/19/what-air-canada-lost-in-remarkable-lying-ai-chatbot-case/>
- Government of Canada. (2020, June 14). *Joint Statement from founding members of the Global Partnership on Artificial Intelligence*. Innovation, Science and Economic Development Canada. <https://www.canada.ca/en/innovation-science-economic-development/news/2020/06/joint-statement-from-founding-members-of-the-global-partnership-on-artificial-intelligence.html>
- Government of Canada. (2023, April 19). *2023–2026 Data Strategy for the Federal Public Service*. Government of Canada. <https://www.canada.ca/en/treasury-board-secretariat/corporate/reports/2023-2026-data-strategy.html>
- Government of Canada. (2025a, January 31). *The Artificial Intelligence and Data Act (AIDA) – Companion document*. Innovation, Science and Economic Development Canada; Innovation, Science and Economic Development Canada. <https://ised-isde.canada.ca/site/innovation-better-canada/en/artificial-intelligence-and-data-act-aida-companion-document>
- Government of Canada. (2025b, March 4). *AI Strategy for the Federal Public Service 2025–2027*. Government of Canada. <https://www.canada.ca/en/government/system/digital-government/digital-government-innovations/responsible-use-ai/gc-ai-strategy-priority-areas.html>
- Hälterlein, J. (2024). Imagining and governing artificial intelligence: The ordoliberal way—an analysis of the national strategy ‘AI made in Germany.’ *AI & SOCIETY*, 1–12. <https://doi.org/10.1007/s00146-024-01940-0>

- Heldreth, C. M., Monk, E. P., Clark, A. T., Schumann, C., Eyee, X., & Ricco, S. (2024). Which Skin Tone Measures Are the Most Inclusive? An Investigation of Skin Tone Measures for Artificial Intelligence. *ACM J. Responsib. Comput.*, 1(1), 1–21. <https://doi.org/10.1145/3632120>
- Hjaltalin, I. T., & Sigurdarson, H. T. (2024). The strategic use of AI in the public sector: A public values analysis of national AI strategies. *Government Information Quarterly*, 41(1), 1–16. <https://doi.org/10.1016/j.giq.2024.101914>
- Hoff, J.L. (2023). Unavoidable futures? How governments articulate sociotechnical imaginaries of AI and healthcare services. *Futures*, 148, 1–13. <https://doi.org/10.1016/j.futures.2023.103131>
- Investissement Québec. (2024). *Montréal's Artificial Intelligence Hub*. Investissement Québec International. <https://iq-siteweb.azurewebsites.net/international/en/secteurs-activite-economique/technologies-information-communications/Montreal-s-Artificial-Intelligence-Hub.html>
- Jasanoff, S., & Kim, S. H. (2009). Containing the Atom: Sociotechnical Imaginaries and Nuclear Power in the United States and South Korea. *Minerva*, 47(2), 119–146. <https://doi.org/10.1007/s11024-009-9124-4>
- Köstler, L., & Ossewaarde, R. (2022). The making of AI society: AI futures frames in German political and media discourses. *AI & SOCIETY*, 37(1), 249–263. <https://doi.org/10.1007/s00146-021-01161-9>
- Leavy, S., Siapera, E., & O'Sullivan, B. (2021). Ethical Data Curation for AI: An Approach based on Feminist Epistemology and Critical Theories of Race. *Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society*, 695–703. <https://doi.org/10.1145/3461702.3462598>
- Li, N., Broner, S., Kim, Y., Mizuo, K., Sauder, E., To, C., Wang, A., Gila, O., & Shindler, M. (2025). Investigating the Capabilities of Generative AI in Solving Data Structures, Algorithms, and Computability Problems. *Proceedings of the 56th ACM Technical Symposium on Computer Science Education V.1*, 659–665. <https://doi.org/10.1145/3641554.3701946>
- Mager, A., & Katzenbach, C. (2021). Future imaginaries in the making and governing of digital technology: Multiple, contested, commodified. *New Media & Society*, 23(2), 223–236. <https://doi.org/10.1177/1461444820929321>
- OpenAI. (2022, November 30). *Introducing ChatGPT*. OpenAI Products. <https://openai.com/index/chatgpt/>
- Paltieli, G. (2022). The political imaginary of National AI Strategies. *AI & SOCIETY*, 37(4), 1613–1624. <https://doi.org/10.1007/s00146-021-01258-1>
- PDPC Singapore. (2020). *Singapore's Approach to AI Governance*. Personal Data Protection Commission Singapore. <https://www.pdpc.gov.sg/help-and-resources/2020/01/model-ai-governance-framework>
- Qaiser, Shahzad & Ali, Ramsha. (2018). Text Mining: Use of TF-IDF to Examine the Relevance of Words to Documents. *International Journal of Computer Applications*. 181. 10.5120/ijca2018917395.
- Scheuerman, M. K., Hanna, A., & Denton, E. (2021). Do Datasets Have Politics? Disciplinary Values in Computer Vision Dataset Development. *Proceedings of the ACM on Human-Computer Interaction*, 5(CSCW2), 1–37. <https://doi.org/10.1145/3476058>
- State of California. (2024, March 13). *California Consumer Privacy Act (CCPA)*. State of California — Department of Justice — Office of the Attorney General. <https://oag.ca.gov/privacy/ccpa>
- Turkina, E., & Oreshkin, B. (2022). Evolving industrial districts and changing innovation patterns: The case of Montreal. *Competitiveness Review*, 32(5), 667–685. <https://doi.org/10.1108/CR-11-2021-0165>
- Wang, J., & Dong, Y. (2020). Measurement of Text Similarity: A Survey. *Information*, 11(9), 421. <https://doi.org/10.3390/info11090421>
- Wickett, K. M., Lamba, M., & Newman, J. (2024). Putting People First in Data Quality: Feminist Data Ethics for Open Government Datasets. *Proceedings of the Association for Information Science and Technology*, 61(1), 1135–1137. <https://doi.org/10.1002/pra2.1209>