

Not Knowing What Hit You and the Disruptive Potential of the Disgruntled: Risks of geopolitical instabilities caused by artificial intelligence and emerging technologies

ISSUE

What are the risks of geopolitical instabilities caused by artificial intelligence (AI) and emerging technologies (ET)?

BACKGROUND CONTEXT

Global Affairs Canada's (GAC) top priorities include "the preservation and strengthening of an international order based on rules", which is interpreted as meaning, among other things "advancing inclusive and progressive international trade arrangements, reducing poverty, advancing gender equality"¹. Another one of GAC's top priorities is advancing a Canada's feminist foreign policy, which is understood as "promoting gender equality, human rights, inclusion and respect for diversity and inclusive governance, both in Canada and internationally. A third top priority for GAC is pursuing a progressive trade agenda "to support all segments of society in taking advantage of the economic opportunities flowing from trade and investment". Economic and social transformations induced by the development of AI & ET affect each of those priorities. They produce a large number of disgruntled people who suffer from the changes, but are in the dark as to their complex causes. By playing on people's cognitive biases, social medias participate in geopolitical destabilisation by helping to mobilise and to organize opposition to established institutions and fostering and crystallizing dissension. While this briefing note will highlight some of those economic and social transformations and potential causes, it will mainly focus on the potential consequences they may trigger and the mechanisms that will be at play in their unfolding.

Risks of accentuating the negative impacts of globalization

In the last 150 years, we have witnessed an intensification of globalization, creating a worldwide market and consumer class. In this internationally integrated economy, most daily use goods are the product of complex international transactions. Raw resources (be they material or digital) are taken from all around the world, transformed elsewhere across the globe and then shipped home. Trade arrangements facilitating the movement of goods and relatively cheap transporting costs have made possible massive economies of scale. Such economies drive down the prices of consumer goods and make them more accessible to an increasing larger number of people.

In such a globalized economic system, major decisions related to resource allocation are taken far from the individual producers and consumers and have become opaque to them. Therefore, when production shifts from one place to another, disgruntled employees (and the local economy depending on them) are in the dark; the decisions are taken too far away from them and the web of causes and justifications is often too complex for them to fully understand. Popular backlash at the consequences of globalization has been felt more acutely when its

benefits have not been perceived as being distributed fairly and when insufficient protections have been given to those who happen to lose from job displacements. International deployment of AI & ET has the potential to accentuate this backlash and increase social unrest if not managed correctly.

Amongst other things, AI & ET will allow for automatization of tasks that were once believed impossible to automatize by reason of the cognitive skills or dexterity that they require for their achievement. This will increase the capital to labor ratio in diverse spheres of the economy. While previous industrial revolutions worked through “de-skilling” work into easier tasks to be accomplished by middle to low-skilled workers, AI & ET is “up-skilling” work by emptying out the middle ground between high-skill jobs and low paying jobs².

AI & ET is expected to increase the decoupling that we have been witnessing in the last few decades between the increased productivity, on the one hand, and the number of jobs and the value of low and median wages, on the other³. The benefits from the technological advances flowing from AI & ET may not be distributed evenly among the population. Those who are able to make the capital investments in the new technologies will reap important financial benefits. But the labour forces that will be displaced by those same technologies will require significant investments in (re)training workers to be (re)inserted in the productive economy. Not every one will find a place in this transformed labour market. For example, in Canada, it will not be easy to retrain the 250 000 truck drivers⁴ who may lose their job to autonomous trucks. Provision of a robust social safety net will be necessary to enable transition without major disruption and to ensure social stability. This explains why, in the tech policy world, the idea of introducing universal income schemes is intensely discussed and debated.

But AI & ET may pose an additional challenge to the classic responses to the downsides of globalization. Since labour currently counts for a significant portion of States’ tax bases, the shift from labor to capital threatens the capacity of States to keep providing both security and an appropriate social safety net. And because AI & ET “naturally gravitat[e] toward monopolies”⁵ and that China and the United States “are set up to capture a full 70 percent of the \$15.7 trillion that AI will add to the global economy by 2030”, “[o]ther countries will be left to pick up the scraps, while these AI superpowers will boost productivity at home and harvest profits from markets around the globe”⁶. This risks accentuating the divide between countries that can afford adequate services to their population and those that are dependent on, and vulnerable to, external interventions (either private or public).

The black box⁷ of inequality

While the macroeconomic conditions governing people’s lives may be difficult to grasp, AI & ET may bring another source of mass dissatisfaction: a new form of AI-induced discrimination. While direct discrimination is rather easy to identify and to police, indirect discrimination resulting from disparate impact of facially neutral rules or processes is often harder to detect and to fight. AI may not only compound the risks of direct and indirect discrimination by

reasons of the biases found in the data used to train the systems (“garbage in, garbage out”), it may also create a new form of discrimination based on correlations that humans may not have identified, given any signification or even understand⁸. These three forms of discrimination have already had an impact on predictive policing⁹, bail applications¹⁰, fraud detection and tax inspections, advertising¹¹, price discrimination, employment¹², loans and insurance decisions, etc.

Algorithmically based decisions are often opaque because the algorithms and data used are protected by commercial secret, because governments’ may need to protect their investigation techniques, or simply because they rely on deep learning algorithms that are unintelligible to the human mind. In such cases, not only may people be upset about being adversely affected by the decisions, but they may also react negatively to the lack of a proper justification for a decision hurting them.

Because of the consolidation of data streams, a negative decision in one aspect of one’s life (e.g. denial of a life insurance policy) may amplify the risk of other negative decisions in others (e.g. employment decision). This snowball effect may create a risk of general exclusion of the individual in question for reasons that escape him or her entirely. This reaches its epitome in the Chinese “social credit” systems¹³.

Cognitive biases that may turn ignorance into an instrument of geopolitical instability

Unchecked negative consequences resulting from major macroeconomic changes and AI-induced discrimination, combined with an easy access to reach millions through social networks create the perfect conditions for conspiracy theories, hate groups and populist movements to flourish. Indeed, those conditions breed the desire to find a “culprit” – an Other – to blame for one’s misfortune. Humans tend to infer that negative events are the results of an intentional agent¹⁴ – as opposed to the results of a complex set of random causes. The easiest an explanation comes to mind, be it because that explanation is emotionally charged or because it fits a readily available example, the more persuasive it will be¹⁵. Repetition of an explanation related to a negative event may create “availability cascades”, making that explanation increasingly plausible and more difficult for officials to deny without having to burden the cost of being labelled corrupt or naïve – thus provoking a “reputational cascade”¹⁶. Once a narrative has anchored itself, it is very difficult to displace. Echo chambers foster extremism and enmity.

Social networks are perfect instruments to accelerate negative informational cascades – especially with the possibility of “deepfakes” and bot farms amplifying messages. In attempting to address this problem, technological fixes may end up simply displacing it or even amplifying it. For example, algorithms that mainly promote “liked” contents may increase the effectiveness of availability cascades and the potential dispersion of conspiracy theories, hateful views and other divisive messages. Switching to algorithms more attuned to “personal connections” may lower the impact of “bad norms entrepreneurs”, but they may nonetheless have the unintended consequence of limiting the potential of established press organisations to infuse legitimate facts in the debates.

Cascades promoting conspiracy theories and mobilising populist movements may occur “spontaneously”, or they may be triggered by intentional agents. Indeed, the combination of inscrutability of complex decision-making processes, human cognitive biases and the social networks infrastructures can be weaponized to destabilise regimes by increasing polarisation within a society, or by bolstering a specific option one favours. The opacity of decisions affecting individuals is therefore most fertile grounds for a new form of “algitprop”¹⁷.

CONSIDERATIONS

- A World Artificial Intelligence Organisation (WAIO) to identify best practices, set standards and assist countries in developing effective AI legislation and regulations. Canada-Québec-France are discussing the creation of such an organisation. UK, in light of its current situation, would probably be another willing and desirable partner;
- Trade agreements need not only be fair and progressive in the abstract, their benefits for local industries have to be better explained by the representatives of the diverse stakeholders who participated in developing the national strategy (i.e. Canadian government, provincial governments, unions, members of the civil society that have been involved, etc.);
- Strong support and transition mechanisms (including (re-)training and “portable” benefits disconnected from specific employment situations) have to be put in place for the displaced workers, here and abroad to limit popular backlashes;
- Reviewing antitrust principles applicable to major digital platforms to best remove competitors’ barriers to entry;
- Treating major digital platforms as public utilities and regulate them accordingly;
- Promoting data trusts, limiting the types of data that can be collected and exploring alternatives to the private property model of data (e.g. the “common heritage of humankind” regime applicable to the human genome) to decrease barriers to entry and to weakens certain States of corporations’ dominant positions;
- Exploring de-territorialised tax revenues, including models of financial transactions taxes;
- Developing and exporting expertise on how to debias algorithmic decision-making tools, not only to promote human rights, but also to ensure that Canada occupies a position consistent with its values and established reputation on the international AI political chessboard;
- Working with social networks to decrease the risks of negative informational cascades;
- Supporting fact-checking organisations and creating incentives for media corporations to invest more in investigative journalisms and less in infotainment;
- Developing civic education programs adapted to the digital age and that include a strong digital literacy component – such program could also be an important Canadian export as part of a digital soft power strategy.

The Cold War has been won by the West in large part thanks to economic efficiency of market-based resources allocation over the centrally-managed alternative. AI & ET pose the threat of a new competition between digital authoritarianism fueled by public/private surveillance data

and freer (but maybe less efficient) liberal democracies¹⁸. Digital authoritarianism may be a way to avoid all sorts of disruptive informational cascades and to incentivize individual in conforming to authority by fulfilling their material needs while maintaining a tight social control. However, this is at the cost of sidelining the rule of law and fundamental human rights. Promoting inclusiveness, freedom and equality will require developing policies and political narratives that go well beyond championing efficiency and competitiveness in AI and ET development. This is a niche that Canada can hope to occupy in the New World Order.

¹ Global Affairs Canada, online: <https://www.international.gc.ca/gac-amc/priorities-priorites.aspx?lang=eng> .

² K.-F. Lee, *AI Super-Powers: China, Silicon Valley and the New World Order* (Houghton Mifflin Harcourt, New York *2018), pp. 150 and ff.

³ On this “Great Decoupling”, see E. Brynjolfsson & A. McAfee, “The Great Decoupling” (2013) 30 (1) *New Perspectives Quarterly* 61; E. Brynjolfsson & A. McAfee, *The Second Machine Age: Work, Progress, and Prosperity in a Time of Brilliant Technologies* (W. W. Norton & Company, New York: 2014) pp. 127 and ff. See also OECD Economic Outlook, Volume 2018 Issue 2, p. 51 and ff. Carolyn A. Wilkins, Senior Deputy Governor, Bank of Canada recently declared, “[t]here is compelling evidence that innovation has been an important reason behind rising income inequality in advanced economies in recent decades.” (See “At the Crossroads: Innovation and Inclusive Growth” (speech delivered at G7 Symposium on Innovation and Inclusive Growth, Montebello, Quebec, February 8, 2018 (<https://www.bankofcanada.ca/2018/02/crossroads-innovation-inclusive-growth/>)). She cited International Monetary Fund, *World Economic Outlook*, Chapter 3, “Understanding the Downward Trend in Labor Income Shares,” April 2017 and G. Michaels, A. Natraj and J. Van Reenen, “Has ICT Polarized Skill Demand? Evidence from Eleven Countries over Twenty-Five Years,” *Review of Economics and Statistics* 96, no. 1 (March 2014): 60–77 to support her claim.

⁴ See CANADA, Statistics Canada, *Labour statistics consistent with the System of National Accounts (SNA), by job category and industry* (Table: 36-10-0489-01 (formerly CANSIM 383-0031).

⁵ “As a technology and an industry, AI naturally gravitates toward monopolies. Its reliance on data for improvement creates a self-perpetuating cycle: better products lead to more users, and thus more users and data. Once a company has jumped out to an early lead, this kind of ongoing repeating cycle can turn that lead into an insurmountable barrier to entry” (Lee, pp. 168-69).

⁶ *Ibid.*

⁷ See generally: F. Pasquale, *The Black Box Society* (Harvard University Press, Cambridge (2015)).

⁸ F. Zuiderveen Borgesius, “Discrimination, artificial intelligence, and algorithmic decision-making”, Council of Europe, Directorate General of Democracy, Strasbourg, 2018, p. 5, 38-39. (<https://rm.coe.int/discrimination-artificial-intelligence-and-algorithmic-decision-making/1680925d73>)

⁹ See A.G. Ferguson, *The Rise of Big Data Policing: Surveillance, Race, and the Future of Law Enforcement* (NYU Press, New York, 2017). See also C. Haskins, “Dozens of Cities Have Secretly Experimented With Predictive Policing Software” (Motherboard (February 6, 2019) (https://motherboard.vice.com/en_us/article/d3m7jq/dozens-of-cities-have-secretly-experimented-with-predictive-policing-software)).

¹⁰ See J. Angwin, J. Larson, S. Mattu & L. Kirchner, “Machine Bias: There’s software used across the country to predict future criminals. And it’s biased against blacks” *Pro Publica* (May 23, 2016) (<https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>) and J. Larson, S. Mattu, L. Kirchner & J. Angwin “How We Analyzed the COMPAS Recidivism Algorithm”, *Pro Publica* (May 23, 2016) (<https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm>).

¹¹ D. Fine Marion, “Science Career Ads Are Disproportionately Seen by Men” *Scientific American* (July 25, 2018) (<https://www.scientificamerican.com/article/science-career-ads-are-disproportionately-seen-by-men/>).

¹² See, for example: Reuters, “Amazon scraps secret AI recruiting tool that showed bias against women” (Oct. 8, 2018) (<https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scrap-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G>)

¹³ N. Kobe, “The complicated truth about China’s social credit system” *UK Wired* (January 21, 2019) (<https://www.wired.co.uk/article/china-social-credit-system-explained>).

¹⁴ See E. Rosset, “It’s no accident: Our bias for intentional explanations” (2008) 108 *Cognition* 771; D. Kelemen & E. Rosset, “The Human Function Compunction: Teleological explanation in adults” (2009) *Cognition* doi:10.1016/j.cognition.2009.01.001; L. Bègue et al., “There is No Such Thing as an Accident” Especially When People Are Drunk”, (2010) 36(1) *Personality and Social Psychology Bulletin* 1301; J. K. Hamlin & A.S. Baron, “Agency Attribution in Infancy: Evidence for a Negativity Bias” (2014) 9 (5) *PLOS ONE* e96112; F. Keil & G. Newman, “Order, Order Everywhere, and Only an Agent to Think: The Cognitive Compulsion to

Infer Intentional Agents” (2015) 30 (2) Mind & Language 117; Y. Meng, T. Griffiths & F. Xu, “Inferring Intentional Agents From Violation of Randomness”, Proceedings of the 39th Annual Conference of the Cognitive Science Society (2017).

¹⁵ On “availability bias”, see, amongst others, the famous article by A. Tversky & D. Kahneman, “Judgment under Uncertainty: Heuristics and Biases” (1974) 185 (4157) Science 1124.

¹⁶ T. Kumar & C. Sunstein, “Availability Cascades and Risk Regulation” (1999) 51 Stan. L. Rev. 683.

¹⁷ I borrow the phrase from Tanya Montforte, Commentary - “Algitprop” (2019) 0 MAI 10.

¹⁸ Nicholas Wright, “How Artificial Intelligence Will Reshape the Global Order – The Coming Competition Between Digital Authoritarianism and Liberal Democracy” Foreign Affairs (July 10, 2018) (<https://www.foreignaffairs.com/articles/world/2018-07-10/how-artificial-intelligence-will-reshape-global-order>) and N. Wright, “Three Distinct Artificial Challenges for the United Nations”, (December 10, 2018) United Nation University, Center for Policy Research, AI and Global Governance platform (<https://ourworld.unu.edu/en/three-distinct-artificial-intelligence-challenges-for-the-un>) (“For the first time since the end of the Cold War, AI is enabling a plausible competitor to liberal democracy — digital authoritarianism. Specifically, AI’s greatest impact on competition in the global order will be to enable a new system of social organisation, one which provides a plausible path forward for big industrially sophisticated states to make their citizens rich *and* maintain rigid control over the citizenry.”)