A Global Public Opinion on Articial Intelligence policy micro-report

Deepfakes and electoral integrity

The Schwartz Reisman Institute for Technology and Society (SRI), in collaboration with the Policy, Elections and Representation Lab (PEARL) at the Munk School of Global Affairs & Public Policy at the University of Toronto, published a survey in 2024 examining opinions about artificial intelligence (AI) in 21 countries. The Global Public Opinion on Artificial Intelligence survey (GPO-AI) reveals varying, diverse and region-specific attitudes about the use of artificial intelligence. Topics of focus include job loss, deepfakes, and justice.

Introduction

Deepfakes have enormous potential to disrupt democratic processes. These pieces of synthetic media often take the form of photos, videos or audio recordings that have been altered using artificial intelligence (AI) to depict someone saying or doing something they did not do.

Deepfake content that targets political figures and institutions presents a particular and real threat to democratic systems. One well-known example involved robocalls in which American President Biden appeared to tell citizens not to vote in the 2024 New Hampshire Democratic Party presidential primary. In a second example, a deepfake audio recording in Slovakia was released 48 hours before its 2023 election, purporting to be a recording of a party leader discussing how to rig the election.

However, the Global Public Opinion on Artificial Intelligence (GPO-AI) survey by the University of Toronto shows that most people are unaware of deepfakes. Further, respondents in this survey were divided on which actors or methods are best placed to tackle the detection, response and regulation of deepfakes. These responses underscore the need for policies to address voter awareness and deepfake use in the political sphere.

There is a lot of fake news, where politicians tell lies with their voice, but this is only material created by AI.

- Respondent (Poland)

[Mnóstwo fakenewsów, gdzie polityk mówi kłamstwa swoim głosem, ale jest to tylko materiał stworzony przez AI]

Background

Numerous jurisdictions have initiated efforts to tackle the challenge of deepfake regulation, deploying several different strategies. In February 2024, Brazil became one of the first countries to specifically regulate the use of deepfakes, falsified audio clips and generative imagery in a political context.

Several proposed bills in Canada would have introduced regulation for some aspects of deepfakes. The proposed AI and Data Act would have made it an offence to make an AI system available that is likely to cause serious physical or psychological harm, or with intent to defraud the public. The proposed Online Harms Act included protections against deepfake pornography and other harmful online content. However, both bills died on the order paper when Prime Minister Trudeau prorogued Parliament in January 2025.

Similarly, in the United States, the proposed DEEPFAKES Accountability Act would have introduced criminal penalties for non-compliant synthetic media production and required digital watermarks and disclosures. However, this Act was not passed before the presidential transition in January 2025. States including California, Minnesota, Texas and Washington have enacted laws to regulate deepfakes in political contexts which aim to prevent reputational harm to candidates and protect voters from being misled.

In the European Union, while the focus has been on the broader regulation of AI, deepfakes are also addressed. The Artificial Intelligence Act mandates clear labelling of AI-generated content and categorizes systems designed to influence elections as "high risk," subjecting them to stricter scrutiny. The United Kingdom's Online Safety Act has a section dedicated to duties to protect content of democratic importance, defined in part as content that "is or appears to be specifically intended to contribute to democratic political debate in the United Kingdom."

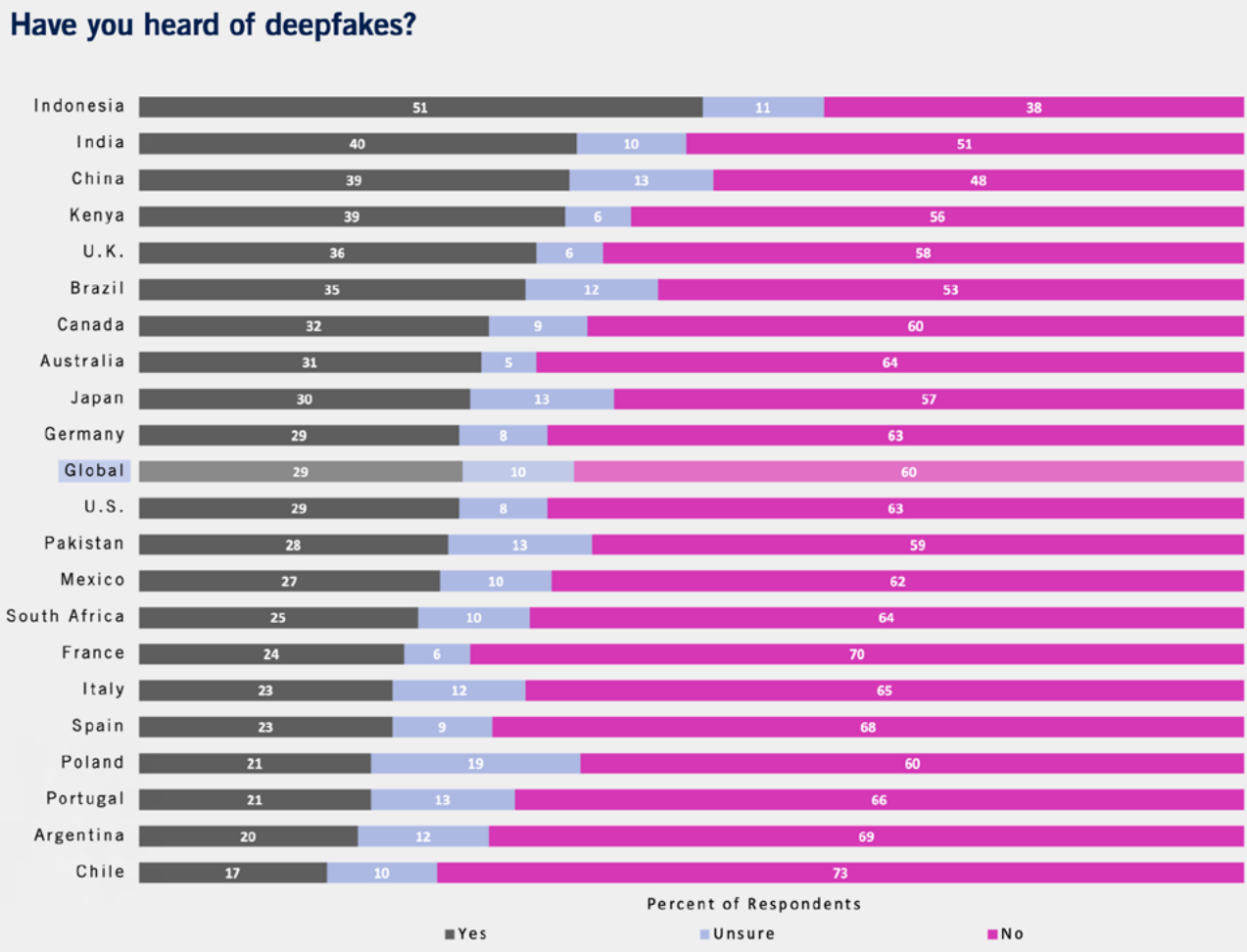
The following findings serve to underline the importance of implementing additional policy interventions beyond regulation. Such interventions could strengthen the governance of deepfakes, providing further protection both in states that have yet to enact deepfake regulation (e.g., Canada and the United States) as well as those that have some deepfake regulation in place (e.g., the European Union and United Kingdom).

GPO-AI Findings

Most respondents have not heard of deepfakes.

There is relatively low global awareness of deepfakes: less than a third of survey respondents (30%) have heard of the term "deepfakes" (Figure 6.1). This lack of awareness is fairly consistent across countries. This finding further reinforces the argument for comprehensive and balanced educational campaigns to equip citizens with the necessary awareness to counter disinformation campaigns. This is discussed further in the policy recommendations section.

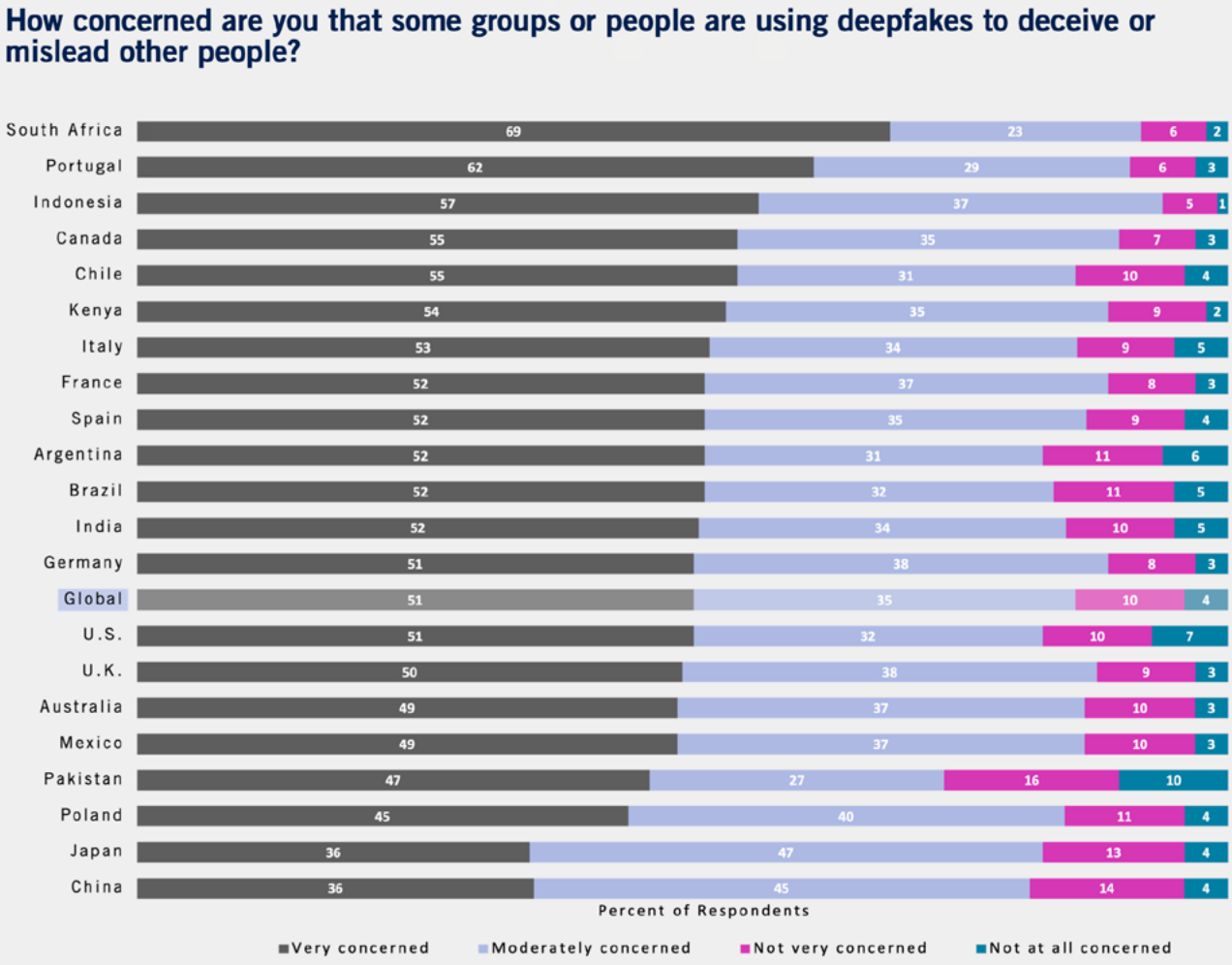
6.1 Deepfake awareness (%)



Most respondents are concerned once they learn about the existence of deepfakes.

When provided with a definition of deepfakes, the vast majority of global respondents (86%) are either very or moderately concerned about their use to deceive and mislead people (Figure 6.2). This not only points to the obvious—that deepfakes are a tool with clear and obvious use cases for deceit—but also to the importance of ensuring information campaigns are balanced to avoid inciting broad mistrust in the media. Though citizens need to be aware of deepfakes to mitigate disinformation, total mistrust in the media would undermine crucial democratic information sources.

6.2 Concern about deepfakes (%)

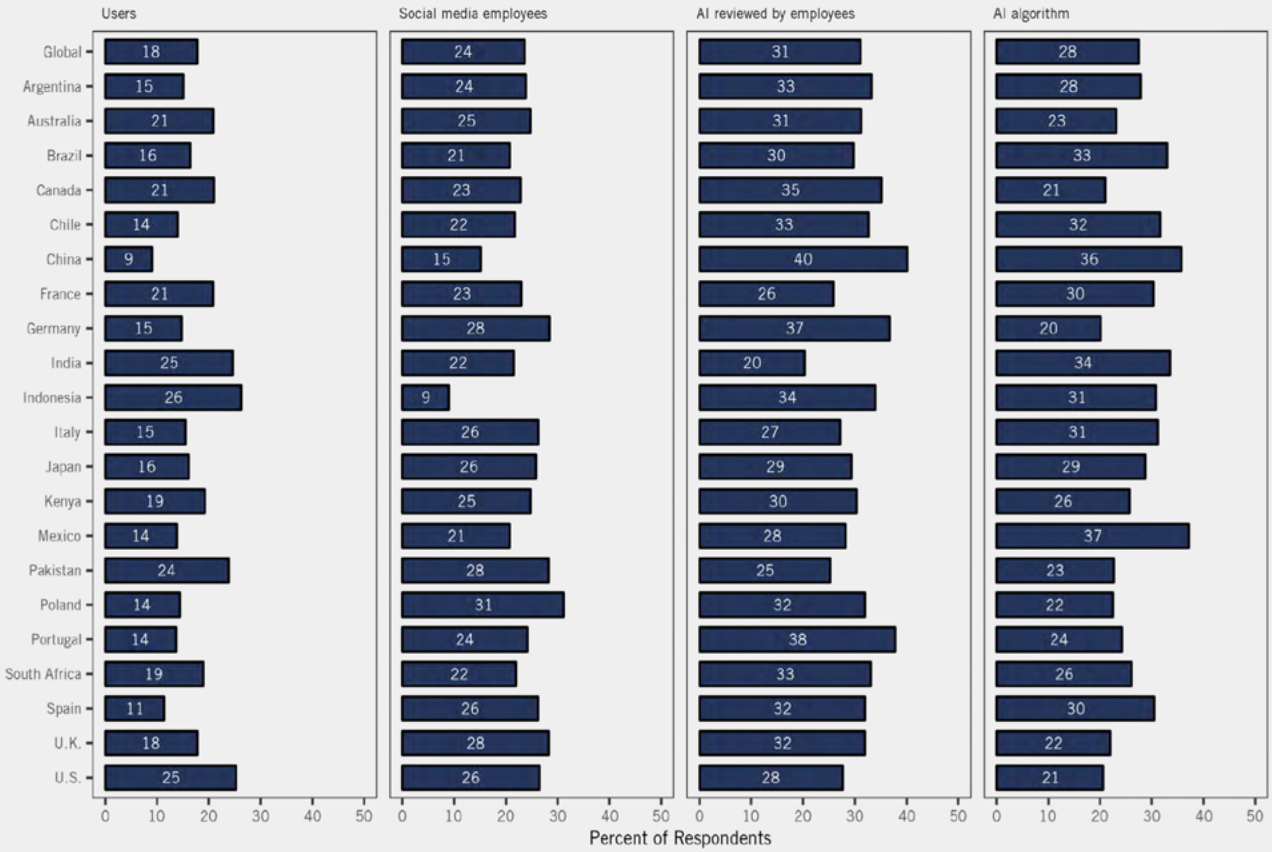


Respondents are divided about which actors or methods are best placed to tackle detection, response and regulation.

Globally, respondents believe the best methods for detecting deepfakes on social media are algorithms and employee review, but they are unsure about who to trust for detection and regulation generally (Figure 6.3). Technology companies are the most trusted to detect and counter deepfakes, followed by governments and university researchers (Figure 6.4). However, fewer than 40% of respondents choose each of these options and support for all three varies widely between countries. North Americans and Europeans generally trust technology companies less than the global average. This raises concerns that a single actor lacks sufficient trust to tackle deepfakes single-handedly. Targeted interventions from multiple actors are therefore likely to elicit more public trust. Our second policy recommendation proposes developing guidelines specific to the use of deepfakes in the political sphere, suggesting how governments can bring together multiple actors to tackle the risks deepfakes pose in this area.

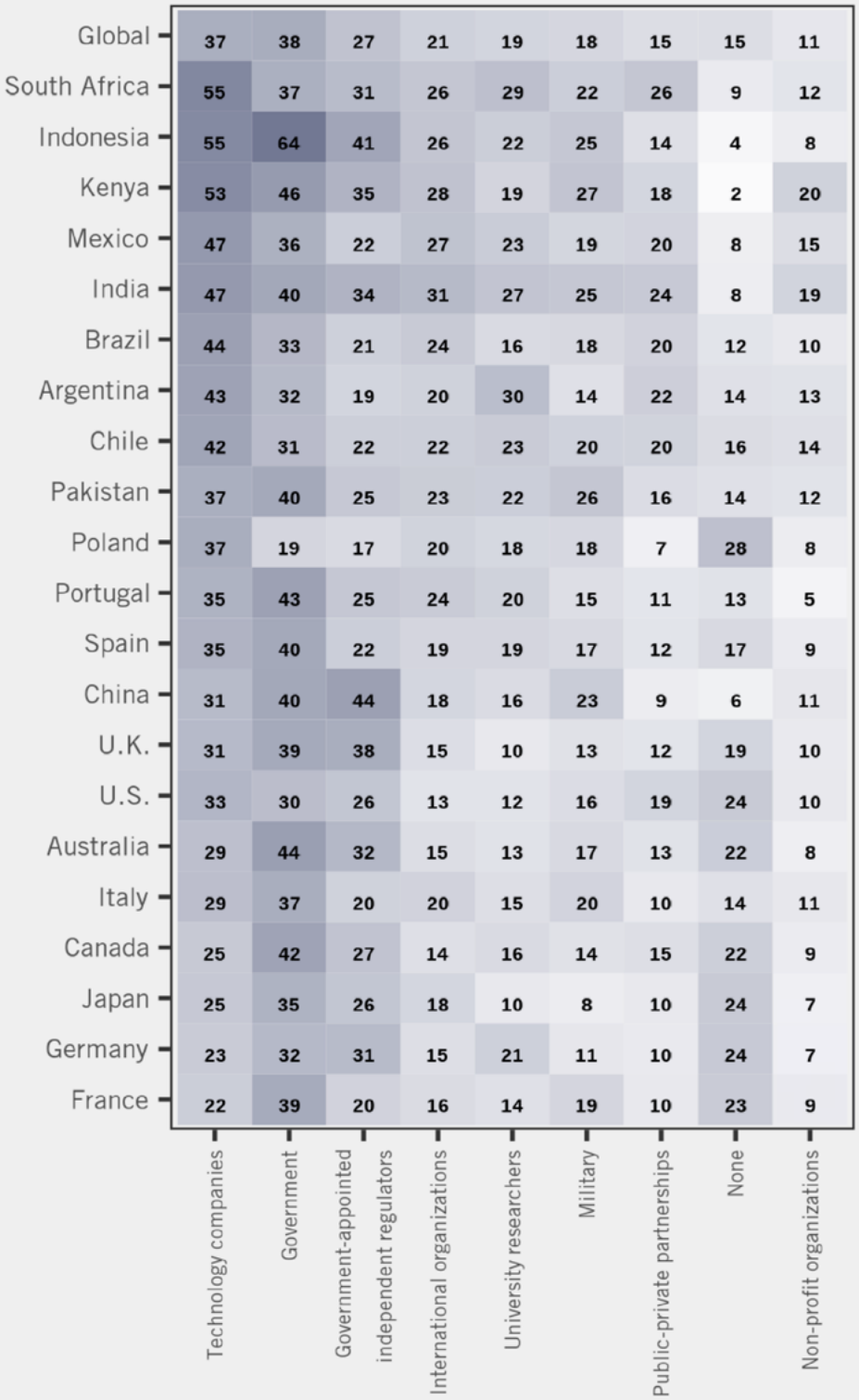
6.3 Best actors for detecting deepfakes (%)

When it comes to detecting deepfakes on social media, in your opinion, who is best placed to do so?



Which of the following actors do you trust to detect and counter deepfakes?

Which of the following actors is best placed to regulate the use of deepfakes?



Policy recommendations

1. Launch public campaigns to increase voter awareness of deepfakes

One of the most troubling statistics about deepfakes in the GPO-AI report is the extremely low level of awareness. Only 30% of global respondents had heard of them at the time the survey was conducted (October and November 2023). In the context of elections, this suggests that a majority of voters may not be aware that it is possible for audio or visual content of political candidates to be fabricated.

This lack of awareness presents an opportunity for malicious actors to interfere in democratic processes. Voters may come into contact with this type of media in the months leading up to a major election. Knowledge of this possibility could decrease the risk of their opinion being swayed by false media.

Since governments bear responsibility for upholding democratic rights and values, they are in a strong position to counter the risks posed by deepfakes through information campaigns. Governments may pursue this goal by devising and implementing public information campaigns that alert the public to the existence of deepfakes and their potential harm to political discourse. These campaigns should be balanced so as to increase public awareness of this issue without heightening concern to a disproportionate level. Such campaigns may also benefit from providing strategies for identifying deepfakes and encouraging people to report them.

In the longer term, digital literacy programs are essential for educating voters on how and where to seek out political information and how to recognize and avoid fraudulent content. Governments should be careful not to overstep in these programs by dictating specific media outlets the public should turn to for accurate information, as this could lead to an unintentional erosion in democratic freedoms. Instead, general guidelines that urge voters to turn to verified and trusted news sources, and to fact-check any political information they see in the media, can reduce citizens' susceptibility to misinformation and disinformation.

2. Develop guidelines specific to the use of deepfakes in the political sphere

Given the novelty of this threat and the dangers it presents to democracy, there are several easy wins governments can implement to help reduce potential harm. Harms from deepfakes can be mitigated through regulations prohibiting non-consensual or malicious political use. This might include explicitly forbidding the creation, distribution and dissemination of deepfake content depicting political figures and/or the messaging of political parties. Such guidelines could also include reporting obligations, such as obliging political candidates and affiliated entities to disclose the use of deepfakes or synthetic content in campaign materials portraying their own party.

These guidelines could also include recommendations on a standardized labelling framework enabling identification of synthetic content across various communication channels, including broadcast media, digital platforms and printed materials. The labelling should clearly indicate that the content has been generated or manipulated using AI technology. Although identifying every instance of deepfake content would be exceptionally difficult, taking this step will nevertheless serve to help the public identify AI-generated content and raise awareness of its existence. However, labelling should go hand-in-hand with broader digital literacy and informational campaigns clarifying that the lack of a label is not a guarantee that the content is real or trustworthy.

Finally, governments can ensure these guidelines have teeth by including effective enforcement mechanisms that can be rapidly deployed. Where possible, governments can do this by allocating resources to enforcement and establishing specialized enforcement agencies tasked with monitoring various digital spaces to detect and remove deepfakes. Governments can collaborate with technology companies to develop advanced algorithms and tools capable of accurately identifying deceptive media. Mechanisms should also be put in place to appeal the removal of content that has been incorrectly flagged as fake to protect against undue censorship or unpredicted biases.

Deepfakes and electoral integrity

A Global Public Opinion on Articial Intelligence policy micro-report

About PEARL

The Policy, Elections & Representation Lab (PEARL) at the Munk School of Global Affairs & Public Policy at the University of Toronto investigates key questions related to political decision-making, representation, the societal and political implications of COVID-19 and the impact of technology on governance. PEARL team members use empirical methods based primarily on survey data, experimental research, and social media data, to understand how society and politics are shaped by attitudes and behaviours. Their work has been published in leading academic journals, featured by the media, and used by a wide range of stakeholders, including policymakers around the world.

About SRI

The Schwartz Reisman Institute for Technology and Society (SRI) at the University of Toronto is an interdisciplinary research hub that examines the social impacts of advanced technologies like artificial intelligence. SRI integrates research across a wide range of disciplines to foster insights towards safe and responsible AI innovation, developing policy-oriented solutions to better align powerful technologies with human values and harness their potential to improve life—for everyone.

Contributors:

Special thanks to our authors and contributors, without whom this targeted policy report would not have been possible.

Maggie Arai
Isaac Gazendam
Hugh Needham
Sofiya Yusypovych

Acknowledgement:

We gratefully acknowledge Professor Peter Loewen, former director of the Munk School of Global Affairs & Public Policy, former associate director of the Schwartz Reisman Institute for Technology and Society, an award winning political scientist and administrator for his work leading the GPO-AI survey on which this report is based.

UNIVERSITY OF TORONTO | SCHWARTZ REISMAN INSTITUTE FOR TECHNOLOGY AND SOCIETY | PEARL POLICY, ELECTIONS & REPRESENTATION LAB | munk school OF GLOBAL AFFAIRS & PUBLIC POLICY