

International Public Management Journal

ISSN: 1096-7494 (Print) 1559-3169 (Online) Journal homepage: www.tandfonline.com/journals/upmj20

Safeguarding online panel surveys from LLM pollution: evidence from Canadian federal survey contracts

Geneviève Morin, Étienne Charbonneau & Daniel Bromberg

To cite this article: Geneviève Morin, Étienne Charbonneau & Daniel Bromberg (04 Nov 2026): Safeguarding online panel surveys from LLM pollution: evidence from Canadian federal survey contracts, International Public Management Journal, DOI: 10.1080/10967494.2026.2716261

To link to this article: https://doi.org/10.1080/10967494.2026.2716261

View supplementary material

Published online: 04 Nov 2026.

Submit your article to this journal

Article views: 21

View related articles

View Crossmark data

Full Terms & Conditions of access and use can be found at
https://www.tandfonline.com/action/journalInformation?journalCode=upmj20

INTERNATIONAL PUBLIC MANAGEMENT JOURNAL
https://doi.org/10.1080/10967494.2026.2716261

Routledge
Taylor & Francis Group

RESEARCH ARTICLE

Check for updates

Safeguarding online panel surveys from LLM pollution: evidence from Canadian federal survey contracts

Geneviève Morin[a,b], Étienne Charbonneau[a] and Daniel Bromberg[b]

[a]École nationale d'administration publique; [b]Vermont Law and Graduate School

ABSTRACT

Bots and large language model (LLM) agents can now complete online surveys convincingly enough to pass as real respondents, a risk long flagged in other fields but rarely addressed in public management research. Drawing on disheartening empirical findings from previous efforts to validate responses to online panel surveys, this research note offers a portrait of efforts to ensure the integrity of data collected by firms that survey citizens on behalf of Canadian federal departments and ministries, a case study likely representative of the challenges facing online survey practices across countries and levels of governments.

ARTICLE HISTORY
Received 28 November 2025
Accepted 4 August 2026

Commercial online survey panels contain bots, a lot of bots. Until a few years ago, a pressing issue in the late 2010s and early 2020s was that a large proportion of survey takers on platforms like Lucid were professional survey takers. A recent study connecting survey takers from 2018 and 2019, to their browsing histories for numbers of visits per day or proportion of browsing time spent on survey portal, found that "in the Lucid sample, the estimates of survey professionals vary between 34.7% and 66.2%, and go up to 71.4% when, including subjects categorized as professionals by any of the measures" (von Hohenberg et al. 2026:441). Consequently, the use of AI-generated responses in online panel surveys presents not only a methodological challenge, but also raises important questions regarding the reliability of evidence used to inform public policy and administrative decision-making. This Research Note focuses specifically on commercial online panel surveys and should not be interpreted as applying equally to telephone, mail, or face-to-face public opinion research.

When we realized there was a problem

One of us recently engaged a reputable survey firm to conduct a survey experiment. It is the kind of firm that delivers electoral surveys before an election. Their survey had some measures to prevent low-quality respondents. After days of drilling down into the data, one of us realized that approximately two-thirds of respondents who failed at i) a ReCAPTCHA test up front, ii) answered a simple question about the task in the treatment/control they just perform, iii) selected '4′ when asked to select 'four' on a ten-item Likert scale, iv) did not speed-run the survey in under three minutes (there was a cognitive task to accomplished). That was an expensive sample from a survey that cost thousands of dollars. We did not know it at the time, but these measures had become obsolete. This research note sounds a warning to public managers and public management researchers who are tempted to use convenient, low-cost commercial online panels. This methodological warning does not limit itself to reproducing methods

CONTACT Geneviève Morin genevieve.morin@enap.ca CRC Comparative Public Management, École nationale d'administration publique, Quebec City, Canada.
Supplemental data for this article can be accessed online at https://doi.org/10.1080/10967494.2026.2716261.
© 2026 Taylor & Francis Group, LLC

for detecting and limiting bots in online panel surveys; more specifically, we present what firms did to protect the validity of government surveys. Public opinion research plays a central role in contemporary public policy and management by supporting policy development, program evaluation, and communication strategies (Trautendorfer, Schmidthuber, and Hilgers 2023). In the Canadian federal context, departments routinely commission surveys on issues such as immigration, public health measures including vaccination, taxation, service delivery experiences, and energy use (Government of Canada 2025). These instruments are used to capture citizens' perceptions, assess policy acceptance, and inform administrative and policy decisions. It relies on both qualitative and quantitative methods including surveys, interviews, and focus groups, and constitutes a key component of evidence-based governance. Consequently, the use of AI Bots in survey responses not only presents a methodological challenge, but it also presents a policy crisis.

Large language models (LLM) can be used in different ways in survey contexts, ranging from simple rephrasing of responses to fully automated completion by agents. "(T)he emerging phenomenon in which LLMs are involved in online tasks intended to measure human responses" has been dubbed "LLM pollution" (Rilla et al. 2025:1). At the time of this writing, an advanced agent can trick most suggestions made to safeguard online surveys (Liem 2025:2), including the venerable second generation ReCAPTCHA tests. Scholars have offered many preventative measures, such as the list below from Rilla et al. (2025:6), which suggest ten design and analysis measures to detect and mitigate LLM pollution.

The disruptive impact of these interventions varies; while some help mitigate risks, none fully eliminate them.

The remainder of this research note is divided into two sections. First, we present examples garnered from other disciplines about the measures they have taken to mitigate bots in survey data collection. Second, we present the extent to which survey firms made efforts to safeguard online panel surveys on behalf of ministries and departments of the Federal government of Canada.

Warning from other disciplines

One of the earliest empirical warnings of bot contamination in public opinion surveys was offered by fourteen authors, sharing their experiences with two surveys (Xu et al. 2022). At the time, the authors had not safeguarded their surveys from bots; rather they identified their interference and removed them as respondents. One of their surveys was left with 123 participants out of 1299, or about 9.4%, after excluding all participants with answers stained by "a) improbable time duration for completion of the survey, (b) illogical responses to open-ended questions, (c) suspicious date/time of survey completion, (d) suspicious name and email address, and (e) duplicate name and email address" (Xu et al. 2022:345). As will be illustrated, the bot detection strategies they offered in 2022, like open-ended questions, logic checks, honey pot items, and reCAPTCHA (Xu et al. 2022:347), all proved ineffectual from bot interference in the following years. Other examples include a recent article, in which a researcher built an autonomous synthetic respondent, that answers surveys with "(…) realistic reading times calibrated to the persona's education level, generates human-like mouse movements, and types open-ended responses keystroke-by-keystroke, complete with plausible typos and corrections" (Westwood 2025:2). Another recent article (Donkin et al. 2025) reporting two cases of bots infiltration on studies about Māori population health with recruitment strategies on social media and online had to exclude respondents to the point where they were left with 382 (47% of the sample) and 27 respondents (5% of the sample).

The detection of bots in survey data have triggered a range of interventions with various levels of sophistication. A basic approach, taken by health researchers was to include hard and soft flags. In which a hard flag is a clear and confirmed indicator that a response is problematic or invalid, and a soft flag is a slight or ambiguous sign that raises doubt. (Thompson and Utz

2025:S29). They ended up keeping 29 surveys (7%), out of an initial 457 (Thompson and Utz 2025:S30). The integrated red flag tool in Qualtrics did not help.

One of the most sophisticated approaches came in a 2024 study on video gaming. The researchers assessed the proportion of MTurkers that were human players in popular online video games, as they indicated in the survey (Eisele and Apruzzese 2024). The respondents were asked to provide a link to their gamertag -their alias-, share easy questions about their character and equipment and to answer a trivial question about the games they said they played (Eisele and Apruzzese 2024:4). In three studies, each using the same technique, the proportion of 'certified' gamers were 5%, 17% and 28%. Most striking, the bots tried to respond to the questions by providing a fake (or incorrect) gamertag or other modifications that logically fit the response but pragmatically made no sense (Eisele and Apruzzese 2024:5). This level of scrutiny is necessary to rule out bot responses but is rarely seen in Public Administration scholarship.

Höhne et al. (2025:5) directly tested the efficacy of six preventative measures: "1) three open narrative questions, 2) one picture CAPTCHA (counting cars), 3) two honey pot questions, 4) one attention check (clicking on survey logo), 5) one check-all-that-apply (CATA) question on survey location, and 6) completion times" (Höhne et al. 2025:6). Then, they designed four bots of various sophistication and asked them to respond to a survey one hundred times each. Even the most rudimentary rule-based bot did not fall for the honeypot questions, let alone the Captchas (Höhne et al. 2025:6,8). The authors warned researchers and practitioners relying on online survey data to use detection and prevention measures other than completion time, honeypot question and captchas (Höhne et al. 2025:8).

Recent work suggests that detecting AI-generated responses requires increasingly sophisticated methods (see Table 1). For example, Affonso (2026) proposes a "cognitive trap" approach that exploits structural limitations in AI agents, highlighting both the potential and the limits of current detection strategies. A 2026 study highlights that ensuring data validity increasingly depends on the integration of multiple detection and validation mechanisms, as individual controls are often insufficient to identify advanced automated responses (Liang et al. 2026:15) Table 2 shows measures discussed in the literature. Unfortunately, the speed at which LLMs are moving will most likely make these techniques ineffective.

The research above provides a guide to help prevent "LLM pollution," but it is only effective if it is utilized. In the following section we offer an in-depth review of public opinion research procured by the Canadian Federal Government. We use this example to demonstrate the consequential nature of bots in public opinion research.

Table 1. List of possible measures for detecting and mitigating LLM pollution.

Measure	Example Implementations	Responsibility	Purpose	Variants	Limitations
Third-party bot protection solutions	Use Cloudflare Bot Management, reCAPTCHA, etc.	Platforms, Researchers	Prevention	LLM Delegation LLM Spillover	Locally-running or human-intervening agents can circumvent the protection.
Norm signalling to participants	Remind participants of expectations and ask for authenticity affirmation	Platforms, Researchers, Community	Prevention	LLM Mediation LLM Delegation LLM Spillover	Making LLM presence too salient can increase suspicion or reactivity among participants.
Multimodal presentation of instructions	Present illustrated images or tutorial videos	Researchers	Prevention	LLM Mediation LLM Delegation	Some participants and automation tools provide screenshots to vision-capable models.
Input interface restriction	Block copying and pasting, or collecting responses via audio recording	Researchers	Prevention	LLM Mediation LLM Delegation LLM Spillover	Audio recording might have privacy concerns despite its effectiveness.
LLM-specific comprehension checks	Present modified versions of Theory of Mind tests, visual illusion quizzes, etc.	Researchers	Prevention	LLM Mediation LLM Delegation	Comprehensive checks effective for some models are not necessarily applicable to others.
Honeypot questions	Add very small or white text with specific instructions for LLMs	Researchers	Post-hoc detection	LLM Mediation LLM Delegation	Advanced LLM-based agents might be able to detect and ignore honeypot questions.
Behavioural logging and modelling	Use typing speed, contents, or mouse movements to identify polluted cases	Researchers	Post-hoc detection	LLM Mediation LLM Delegation	Automation tools might become capable of imitating human-like behaviours.
Commercial AI-generated text detectors	Apply GPTZero, undetectable AI, etc.	Researchers	Post-hoc detection	LLM Mediation LLM Delegation	Commercial models lack transparency and assurance against new models.
Institutional enforcement	Standardise policies and procedures for identifying, reporting, and refunding polluted cases	Platforms, Community	Prevention	LLM Mediation LLM Delegation LLM Spillover	Fairness between participants' access and researchers' control can be in tension.
Shared repository of best practices	Collect and disseminate reusable templates, code, etc.	Researchers, Community	Prevention, Post-hoc detection	LLM Mediation LLM Delegation LLM Spillover	Repositories may soon become outdated as model capabilities evolve.

Source: Rilla et al. (2025:6)

Table 2. Proven and unproven measures to prevent bot participation in online surveys.

Measure	Detailed definition	Main objective	References
Attention Checks assessment	Inclusion of specific items within the questionnaire designed to verify whether respondents are paying attention and providing thoughtful answers. Examples include instruction-based items or consistency checks between logically related questions.	Ensure attentiveness and detect inattentive or automated respondents.	M. D. Wang and Hau (2024); Taplin (2025).
CAPTCHA or Bot Protection	Technical safeguard used to differentiate human respondents from automated scripts or bots. It may include reCAPTCHA, hCaptcha, or custom challenges with server-side token validation.	Prevent automated submissions by verifying human interaction.	Gutub and Kheshaifaty (2023); Searles et al. (2023Searles, Prapty, and Tsudik :2025).
Certified or Verified Panels	Use of respondent panels managed by certified providers that validate respondent identity, uniqueness, and authenticity through IP tracking, demographic verification, and anti-fraud systems.	Guarantee respondent authenticity and prevent duplicates or fake accounts.	Pyrooz, Densley, and Sanchez (2025).
Email Verification	Verification process confirming that each email address is valid and active. May include single or double opt-in confirmation or automatic detection of disposable emails.	Confirm respondent legitimacy and filter out fake or temporary emails.	Walker, Murry, and Longoria (2023); J. Wang et al. (2023)
Exclusion Rules and Failure Rate	Set of predefined criteria used to exclude low-quality or invalid responses, such as failed attention checks or unrealistic completion times. The failure rate quantifies data cleaning outcomes.	Maintain data integrity by removing invalid or careless responses.	Taplin (2025); Pinzón et al. (2024)
Logs or Proof of Validation	Maintenance of audit logs recording IP addresses, timestamps, browser types, and CAPTCHA validation to verify participation authenticity.	Provide traceability evidence of validation processes.	Sharma et al. (2021)
Manual or Additional Checks	Human-based verification to review open-ended responses, detect duplicates, or inspect suspicious patterns missed by automation.	Add a final layer of quality assurance to detect subtle irregularities.	Comachio et al. (2025)
Provider and Control Placement	Identification of the technical provider (e.g., Google reCAPTCHA, Cloudflare Turnstile) and strategic placement of controls within the survey (registration, start, or submission).	Optimize bot detection without overburdening legitimate respondents.	Comachio et al. (2025)
Results of Quality Measures	Summary of performance indicators for all implemented quality measures, including rates of exclusion, attention check failures, and validated responses.	Evaluate effectiveness of protection measures and overall data quality.	Pinzón et al. (2024)
Researcher Technical Quality Controls	Comprehensive technical procedures such as IP filtering, VPN/proxy blocking, device/session limits, and response-time analysis to detect bots or repeated entries.	Ensure data reliability and prevent automated or fraudulent responses.	Comachio et al. (2025)
Panel Technical Quality Controls	Continuous panelist validation, rate-limiting controls, curbing panelist professionalism, and panelist quality assurance	Disclosing specific protocols on the parts of commercial panels	Westwood (2025)
Token Validation	Assignment of a unique token for each respondent to ensure single-use survey access and prevent duplicate submissions through server-side validation.	Prevent multiple entries and secure participant authentication.	

(*Continued*)

Table 2. Continued.

Measure	Detailed definition	Main objective	References
Cognitive Trap	A methodology for detecting AI agents with visual capabilities by leveraging the architectural constraints of vision-language models	Detect and filter out AI-generated or automated responses by creating conditions under which non-human agents are more likely to fail or behave inconsistently compared to human respondents.	Affonso (2026:9)

Table 3. Dollar value of reports.

Years	*n*	Contract value (current Canadian $)
2025	85	$13,784,116
2024	138	$27,292,049
2023	156	$25,527,239
2022	157	$32,952,562
2021	116	$11,597,428
2020	86	$10,454,031
Total	738	$121,607,425

Methodology

This section evaluates how quickly survey firms have adapted their practices to address the impact of LLMs on survey research. Specifically, we take an exploratory qualitative approach to document and analyze the measures implemented to counter bot use in Canadian federal government-procured survey research between 2020 and 2025 (Appendix A). The main objective is to understand the practices, tools, and strategies used to ensure the quality and integrity of data collected electronically and how they have changed over the previous five years.

Data collection

Secondary data was collected from a federal government portal which contains all survey reports conducted or commissioned by federal departments and agencies. All reports available for the period from 2020 to 2025 were systematically downloaded to form the analysis corpus (as of October 25th, 2025). The PDF documents were found on the Government of Canada website https://recherche-collection-search.bac-lac.gc.ca/eng/Help/porr

A total of 738 documents were downloaded manually (2020 $n = 86$; 2021 $n = 116$; 2022 $n = 157$; 2023 $n = 156$; 2024 $n = 138$; 2025 $n = 85$* (before the end of the year)).

Table 3 demonstrates that beyond the public policy implications these procured surveys cost Canadian taxpayers hundreds of millions of dollars.

Data analysis

The analysis was conducted by combining artificial intelligence and manual analysis. The transparent use of artificial intelligence to analyze documents related to survey integrity raises an apparent methodological paradox. In this context, AI is used as a controlled analytical instrument to support data extraction and organization, under direct human supervision and validation. This differs fundamentally from the hidden, undisclosed, and uncontrolled use of automated agents as survey respondents, which introduces risks of data contamination (Thaichon et al. 2026). The combined use of AI-assisted processing and manual validation was therefore intended to ensure

the validity and rigor of the interpretive process in four successive stages. The reports were initially analyzed to extract relevant passages answering four main questions, which were defined in advance and presented in Appendix B. The stages took place in the following order:

1. Evidano, an artificial intelligence tool for document corpus analysis, was used to process all reports. The analysis focused on identifying text related to survey costs, data collection methods, anti-bot protections, response quality checks, and validation procedures.
2. The extracted passages were exported and organized by reference year. They were then manually coded using a study-specific coding grid (Appendix B), which categorizes key features such as protection mechanisms (e.g., reCAPTCHA, IP filtering, server validation), attention checks, and respondent identity verification. The AI prompts are provided in Appendix C.
3. Two output files were generated: one containing verbatim excerpts from the reports and another containing AI-generated synthesized results.
4. Methodological triangulation was conducted to compare the original excerpts with the automated syntheses, ensuring consistency, accuracy, and fidelity to the source material. This review confirmed no discernible differences between automated and manual analyses.

Results

The results are organized around two analytical dimensions: (1) changes in data collection methods and modes, and (2) changes in the reported use of protection measures against automated or low-quality responses across the two periods. The end of 2022 is used as the demarcation at which point conversational AI is accessible to the public. This milestone corresponds to the public release of ChatGPT on November 30, 2022. This access triggered unprecedented adoption of LLMs and widespread dissemination in everyday and academic uses (Denning 2023; Nature Editorial 2023). The first few months following the launch were marked by explosive growth, often described as the fastest for a consumer application, and documented in both scientific literature and adoption analyses. This justifies the use of this date as an inflection point for our analyses (Florindo 2023; Farhat et al. 2023).

Table 4 shows that between the two periods, changes were observed in both the type and mode of data collection. The use of surveys increased from 47.4% in Period 1 to 64.9% in Period 2 ($p < 0.001$). This increase suggests a growing preference for large-scale standardized quantitative methods, likely facilitated by the development of digital survey platforms and

Table 4. Data collection methods and model, by period.

Data collection method	Period 2020–2022 (*n*)	Period 2020–2022 (%)	Period 2023–2025* (*n*)	Period 2023–2025*: (%)	Difference (%)	*z*	*p*-Value
Survey	170	47.4	246	64.9	+17.5	−4.79	<0.001
Interviews	93	25.9	88	23.2	−2.7	+0.85	0.394
Focus groups	62	17.3	69	18.2	+0.9	−0.32	0.749
Observation	4	1.1	7	1.8	+0.7	−0.79	0.428
Other (specify)	18	5.0	18	4.7	−0.3	+0.19	0.850
Total (*n*)	347		428				
Mode of data collection	**Period 2020–2022 (*n*)**	**Period 2020–2022 (%)**	**Period 2023–2025* (*n*)**	**Period 2023–2025*: (%)**	**Difference (%)**	***z***	***p*-Value**
Online	156	43.5	237	62.5	+19.0	−5.17	<0.001
Telephone	77	21.4	126	33.2	+11.8	−3.59	<0.001
Face-to-face	13	3.6	8	2.1	−1.5	+1.23	0.219
Mail	1	0.3	6	1.6	+1.3	−1.80	0.071
Total (*n*)	247		377				

*The 2023–2025 period includes only reports from early 2025.

automated data processing tools introduced after 2022. Interviews, focus groups, and observations remained relatively stable. These methods continue to play a complementary role but represent a smaller share of the overall methods reported. In terms of data collection methods, the results reveal a marked transition toward online and telephone surveys, which increased by +19.0% and +11.8%, respectively (both $p < 0.001$). These increases underscore the normalization of remote participation and digital communication tools in federal data collection processes. At the same time, face-to-face collection declined slightly (−1.5%, n.s.), consistent with post-pandemic organizational practices and hybrid work environments, while mail surveys remained marginal but showed a slight increase ($p = 0.071$). Overall, these results indicate a clear digitization of data collection practices after 2022, reflecting the growing integration of automation and artificial intelligence into federal research infrastructures.

As shown in Table 5, comparisons between Period 1 (2020–2022) and Period 2 (2023–2025) indicate a minor methodological adjustment in how firms working on the behalf of federal institutions treat survey data integrity. While overall changes remain modest, a few measures show statistically significant increases, suggesting a gradual adaptation of data quality protocols rather than a complete methodological shift. Attention checks increased slightly (+3.1%, $p = 0.003$), indicating an increased emphasis on internal consistency and respondent engagement. Similarly, exclusion rules and failure rate thresholds increased by 6.4% ($p = 0.008$), indicating a somewhat increased focus on post-collection validation and removal of questionable responses. The most notable, but still moderate, growth is seen in technical quality controls (+14.5%, $p < 0.001$) and token validation mechanisms (+5.4%, $p = 0.001$).

In this context, "technical quality controls" refer to a set of backend and procedural mechanisms designed to detect and prevent automated or fraudulent responses. These include IP filtering, VPN and proxy detection, device and session tracking, response-time analysis, and server-side validation processes to identify irregular participation patterns or duplicate entries. These measures suggest a cautious integration of more automated protective measures, even if their prevalence is still rare.

Other measures, such as CAPTCHA or anti-bot protection, manual checks, and email verification, show no significant variation between the two periods. Similarly, the use of certified

Table 5. Frequencies of protection measures against bots deployed in government-procured online panel surveys, by period.

Type of protection measure	Period 2020–2022 (n)	Period 2020–2022 (%)	Period 2023–2025* (n)	Period 2023–2025* (%)	Difference (%)	*z*	*p*-Value
Attention Checks Assessment	2	0.6	14	3.7	+3.1	2.92	0.003
CAPTCHA or Bot Protection	5	1.4	12	3.2	+1.8	1.61	0.108
Certified or Verified Panels	56	15.6	59	15.6	0.0	–0.01	0.991
Email Verification	16	4.5	12	3.2	–1.3	–0.92	0.359
Exclusion Rules and Failure Rate	32	8.9	58	15.3	+6.4	2.65	0.008
Logs or Proof of Validation	1	0.3	4	1.1	+0.8	1.29	0.198
Manual or Additional Checks	32	8.9	45	11.9	+3.0	1.31	0.189
Provider and Control Placement	2	0.6	6	1.6	+1.0	1.35	0.179
Results of Quality Measures	27	7.5	34	9.0	+1.5	0.72	0.475
Technical Quality Controls	21	5.8	77	20.3	+14.5	5.79	0.001
Token Validation	8	2.2	29	7.7	+5.4	3.37	0.001
Total (*n*)	202		350				

Note: * are calculated based on the number of reports per period: Period 1 (2020–2022) = 359 reports. Period 2 (2023–2025) = 379 reports.

or verified panels remained constant at 15.6%, highlighting continuity in how institutional partners and professional panels are used to ensure sample reliability. The changes observed appear to be more evolutionary than transformative, given the low growth rate. They represent marginal improvements in quality and the gradual implementation of behavioral, procedural, and technical controls rather than a systemic overhaul. These developments suggest that, while current responses may remain partial, some survey practitioners are engaging with emerging challenges, and slowly, progressively integrating new safeguards into their data collection processes.

Public servants needing to survey the public: what now?

The broader emergence of generative artificial intelligence tools around 2022 has reshaped data reliability and respondent authenticity. The use of LLMs has prompted a reassessment of quality control measures in some academic disciplines. However, as indicated by the analysis of the Canadian Federal data survey firms, in this case, use a measured approach rather than a radical methodological shift. In this sense, the period after 2022 represents not a break, but an initial advancement toward more systematic verification.

A more noticeable shift emerges around 2023 which can be interpreted as a methodological turning point in the adoption of additional safeguards but remains far from adequate to address the rapid expansion in bots-infected survey data. Further, the variability observed across firms in the adoption of quality control measures suggests that procurement practices may play a critical role in shaping data reliability. Governments may benefit from explicitly requiring advanced validation mechanisms and greater methodological transparency in survey contracts to ensure that collected data accurately reflects human respondents even if some of these measures are no longer safeguards in 2026. These findings should not be interpreted as an argument against the continued use of surveys or other public engagement mechanisms. Surveys, focus groups, and related consultation tools remain essential components of evidence-informed policymaking and public administration. Rather, this Research Note argues that governments and their contractors must adapt existing practices to ensure that the data generated through these mechanisms remain reliable, valid, and fit for purpose. Although this analysis focuses on surveys commissioned by the Canadian federal government, the challenges identified in this Research Note are unlikely to be unique to Canada. Governments around the world increasingly rely on commercial online panel providers to inform policy decisions, evaluate programs, and monitor public opinion. As LLM technologies continue to evolve, concerns surrounding respondent authenticity, data quality, and methodological transparency are likely to become increasingly salient across jurisdictions. Consequently, the recommendations presented in this Research Note, including stronger procurement requirements, layered validation strategies, and enhanced reporting standards, may be relevant to public organizations well beyond the Canadian context.

Recent methodological developments illustrate both the potential and the limits of current approaches. These include the combined use of multiple validation layers, such as token-based access controls, technical filtering mechanisms, and post-collection data screening procedures. While no single method appears sufficient to fully eliminate the risks associated with automated responses, the convergence of multiple controls points toward evolving best practices should be more systematically incorporated into government survey procurement and oversight processes.

This Research Note is subject to several limitations. First, the analysis is limited to publicly available survey reports commissioned by the Canadian federal government and may not fully capture practices adopted by other jurisdictions or private organizations. Second, the rapid evolution of LLM technologies means that some of the safeguards identified in this manuscript may become less effective over time. Future research could examine the extent to which governments in other countries are adapting their procurement and validation practices and evaluate the effectiveness of emerging approaches designed to preserve respondent authenticity in online panel surveys. This Research Note highlights a methodological issue that may warrant future

theoretical attention in areas such as evidence-informed policymaking, public sector innovation, and digital governance. To the four methodological recommendations listed in a recent meta-analysis on transparency and citizen satisfaction (Li, Wang, and Qu 2026:728), we would add a fifth: take precautions to assure that human being answered your surveys.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by Canada Research Chairs.

Notes on contributors

Geneviève Morin, PhD, is a postdoctoral researcher at the École nationale d'administration publique (ENAP) and Vermont Law and Graduate School (VLGS), and a research associate with the Canada Research Chair in Comparative Public Management.

Étienne Charbonneau, PhD, is Professor of Public Administration at the École nationale d'administration publique (ENAP), Canada, and Canada Research Chair in Comparative Public Management.

Daniel Bromberg, PhD, is Professor and Dean of the Graduate School at Vermont Law and Graduate School, United States.

References

Affonso, F. M. 2026. "Brief Commentary: A Framework for Detecting AI Agents in Online Research." *Journal of Consumer Research 53*(3):619–33. doi: 10.1093/jcr/ucag006.

Comachio, Josielli, Adam Poulsen, Adeola Bamgboje-Ayodele, Aidan Tan, Julie Ayre, Rebecca Raeside, Rajshri Roy, and Edel O'Hagan. 2025. "Identifying and Counteracting Fraudulent Responses in Online Recruitment for Health Research: A Scoping Review." *BMJ Evidence-Based Medicine 30*(3):173–82. doi: 10.1136/bmjebm-2024-113170.

Denning, Peter J. 2023. "Can Generative AI Bots be Trusted?" *Communications of the ACM 66*(6):24–7. doi: 10.1145/3592981.

Donkin, Liesje, Nathan Henry, Amy Kercher, Mangor Pedersen, Holly Wilson, and Amy Hai Yan Chan. 2025. "Effective Recruitment or Bot Attack? The Challenge of Internet-Based Research Surveys and Recommendations to Reduce Risk and Improve Robustness." *Interactive Journal of Medical Research 14*:e60548. doi: 10.2196/60548.

Eisele, Linus, and Giovanni Apruzzese. 2024. "Are Crowdsourcing Platforms Reliable for Video Game-Related Research?" A Case Study on Amazon Mechanical Turk." Paper Presented at the Companion Proceedings of the Annual Symposium on Computer-Human Interaction in Play (CHI PLAY Companion '24), October 14th-17th. Tempere, FI.

Farhat, Faiza, Emmanuel Sirimal Silva, Hossein Hassani, Dag Øivind Madsen, Shahab Saquib Sohail, Yassine Himeur, M. Afshar Alam, and Aasim Zafar. 2023. "The Scholarly Footprint of ChatGPT: A Bibliometric Analysis of the Early Outbreak Phase." *Frontiers in Artificial Intelligence 6*:1270749. doi: 10.3389/frai.2023.1270749.

Florindo, Fabio. 2023. "ChatGPT: A Threat or an Opportunity for Scientists?" *Perspectives of Earth and Space Scientists 4*(1):e2023CN000212. doi: 10.1029/2023CN000212.

Government of Canada. 2025. "Defining Public Opinion Research." https://www.canada.ca/en/treasury-board-secretariat/services/government-communications/public-opinion-research-government.html

Gutub, Adnan, and Nafisah Kheshaifaty. 2023. "Practicality Analysis of Utilizing Text-Based CAPTCHA vs. Graphic-Based CAPTCHA Authentication." *Multimedia Tools and Applications 82*(30):1–33. doi: 10.1007/s11042-023-15586-5.

Höhne, Jan Karem, Joshua Claassen, Saijal Shahania, and David Broneske. 2025. "Bots in Web Survey Interviews: A Showcase." *International Journal of Market Research 67*(1):3–12. doi: 10.1177/14707853241297009.

Liang, Xinyun Meg., Rebecca Andrews, Fay Hadley, and Peter Petocz. 2026. "Online Survey Data Integrity: Lessons Learnt from Investigating a Bot Attack on a Small-Scale Educational Study." *International Journal of Social Research Methodology*. Advance online publication doi: 10.1080/13645579.2026.2636205.

Li, Ruiying, Qiushi Wang, and Tingli Qu. 2026. "Transparency and Citizen Satisfaction: A Meta-Analysis and Future Research Agenda." *Public Management Review 28*(3):705–39. doi: 10.1080/14719037.2025.2456658.

Liem, Djinn Gie. 2025. "The Future of Online or Web-Based Research: Have You Been BOTTED?" *Appetite 213*:108058. doi: 10.1016/j.appet.2025.108058.

Nature Editorial. 2023. "Tools Such as ChatGPT Threaten Transparent Science; Here are Our Ground Rules for Their Use." *Nature 613*(7945):612. doi: 10.1038/d41586-023-00191-1.

Pinzón, Natalia, Vikram Koundinya, Ryan E. Galt, William O. 'R. Dowling, Marcela Baukloh, Namah C. Taku-Forchu, Tracy Schohr, Leslie M. Roche, Samuel Ikendi, Mark Cooper, et al. 2024. "AI-Powered Fraud and the Erosion of Online Survey Integrity: An Analysis of 31 Fraud Detection Strategies." *Frontiers in Research Metrics and Analytics 9*:1432774. doi: 10.3389/frma.2024.1432774.

Pyrooz, David C., James A. Densley, and Jose A. Sanchez. 2025. "Are Online Opt-In Panels Viable Data Sources on Hard-to-Reach Populations? Population and Relational Inferences on Gang Membership in the United States." *International Criminology 5*(4):438–54. doi: 10.1007/s43576-025-00174-4.

Rilla, Raluca, Tobias Werner, Hiromu Yakura, Iyad Rahwan, and Anne-Marie Nussberger. 2025. Recognising, Anticipating, and Mitigating LLM Pollution of Online Behavioural Research. Berlin, DE: Center for Humans and Machines, Max Planck Institute for Human Development.

Searles, Andrew, Yoshimichi Nakatsuka, Ercan Ozturk, Andrew Paverd, Gene Tsudik, and Ai Enkoji. 2023. "An Empirical Study & Evaluation of Modern CAPTCHAs." 32nd USENIX Security Symposium (USENIX Security '23), 3081–97. USENIX.

Searles, Andrew, Renascence Tarafder Prapty, and Gene Tsudik. 2025. "Understanding reCAPTCHAv2 via a Large-Scale Live User Study." Proceedings of the 2025 Symposium on Usable Security and Privacy (USEC). Internet Society/Conference Proceedings, San Diego, CA. doi: 10.14722/usec.2025.23021.

Sharma, Akash, Nguyen Tran Minh Duc, Tai Luu Lam Thang, Nguyen Hai Nam, Sze Jia Ng, Kirellos Said Abbas, Nguyen Tien Huy, Ana Marušić, Christine L. Paul, Janette Kwok, et al. 2021. "A Consensus-Based Checklist for Reporting of Survey Studies (CROSS)." *Journal of General Internal Medicine 36*(10):3179–87. doi: 10.1007/s11606-021-06737-1.

Taplin, Ross H. 2025. "Quantifying Data Quality After Removing Respondents Who Fail Data Quality Checks." *Current Issues in Tourism 28*(16):2570–81. doi: 10.1080/13683500.2024.2378611.

Thaichon, Park, Frank Mathmann, Sara Quach, and Liem Viet Ngo. 2026. "When AI Agents Take Surveys: Protecting Data Integrity in Business and Marketing Research." *Australasian Marketing Journal 34*(2):174–7. doi: 10.1177/14413582261421973.

Trautendorfer, Julia, Lisa Schmidthuber, and Dennis Hilgers. 2023. "Surveys Under the Lens: How Public Administration Research Can Benefit from Citizen Survey Data." *International Journal of Public Opinion Research 35*(3):edad019. doi: 10.1093/ijpor/edad019.

Thompson, Amber D., and Rebecca L. Utz. 2025. "Online Surveys: Lessons Learned in Detecting and Protecting Against Insincerity and Bots." *Quality & Quantity 59*(S1):23–39. doi: 10.1007/s11135-024-01973-z.

von Hohenberg, Bernhard Clemm, Tiago Ventura, Jonathan Nagler, Ericka Menchen-Trevino, and Magdalena Wojcieszak. 2026. "Survey Professionalism: New Evidence from Web Browsing Data." *Political Analysis 34*(3):432–50.

Walker, Lorraine O., Nicole Murry, and Kayla D. Longoria. 2023. "Improving Data Integrity and Quality from Online Health Surveys of Women with Infant Children." *Nursing Research 72*(5):386–91. doi: 10.1097/NNR.0000000000000671.

Wang, June, Gabriela Calderon, Erin R. Hager, Lorece V. Edwards, Andrea A. Berry, Yisi Liu, Janny Dinh, August C. Summers, Katherine A. Connor, Megan E. Collins, et al. 2023. "Identifying and Preventing Fraudulent Responses in Online Public Health Surveys: Lessons Learned During the COVID-19 Pandemic." *PLOS Global Public Health 3*(8):e0001452. doi: 10.1371/journal.pgph.0001452.

Wang, Melissa D., and Kit-Tai Hau. 2024. "Flagging Insufficient Effort Responses in Surveys: Stopping Rule to Prevent Insufficient or Excessive Removal of Doubtful Data." *Acta Psychologica 243*:104135. doi: 10.1016/j.actpsy.2024.104135.

Westwood, Sean J. 2025. "The Potential Existential Threat of Large Language Models to Online Survey Research." *Proceedings of the National Academy of Sciences 122*(47):1–10. doi: 10.1073/pnas.2518075122.

Xu, Yanfeng, Sarah Pace, Jaeseung Kim, Aidyn Iachini, L. Bailey King, Theresa Harrison, Dana DeHart, Sue E. Levkoff, Teri A. Browne, Ashlee A. Lewis, et al. 2022. "Threats to Online Surveys: Recognizing, Detecting, and Preventing Survey Bots." *Social Work Research 46*(4):343–50. doi: 10.1093/swr/svac023.