

Critical Analysis of Canada's AI Strategy Public Consultation “What We Heard” Report

Assessing Government Claims Against Raw Data
from 11,383 Public Responses

Authors

Anthropic’s Claude (using Sonnet 4.5 in Extended Thinking mode):

Prompted, lightly supervised, critiqued, and appropriately trusted (neither too much nor too little) by Jason Millar¹

February 2026

NOTE: This report was generated using generative AI to make a point about using generative AI responsibly. The contents of this report have been lightly scrutinized by the human author(s), and it is difficult to validate the accuracy of most of the statements made herein. In short, the contents of this report are NOT the result of human expert analysis. For a full explanation of the report’s genesis, including links to the prompts used to generate the bulk of this report, please see Jason Millar’s Substack post, “The Government of Canada falls short on its claims it used AI responsibly in its National AI Strategy Public Consultation”.

Read the article explaining this report at:

<https://profjasonmillar.substack.com/p/the-government-of-canada-falls-short>

¹ Associate Professor and Canada Research Chair in the ethical Engineering of Robotics and AI, University of Ottawa Faculty of Engineering, School of Engineering Design and Teaching Innovation. Director, Canadian Robotics and AI Ethical Design Lab ([CRAIEDL.ca](https://craiedl.ca))

Table of Contents

Executive Summary	5
10 Key Takeaways	5
Critical Recommendations	6
Section 1: Demographic Breakdown	7
1.1 Response Rate and Data Completeness	7
1.2 Age Distribution	7
1.3 Respondent Type Distribution	8
1.4 Sector Distribution	8
1.5 Geographic Distribution	9
1.6 Demographic Analysis: Key Concerns	10
Section 2: Accuracy of the Government of Canada's Claims	11
2.1 Methodology	11
2.2 Claim 1: Talent Attraction and Retention	11
2.3 Claim 2: Concerns About Premature Deployment	11
2.4 Claim 3: Structured Incentives for Commercialization	12
2.5 Claim 4: Public Trust Elements	12
2.6 Claim 5: Creative Works and Copyright Concerns	13
2.7 Summary of Accuracy Assessment	14
2.8 Methodological Transparency: Weighted vs. Unweighted Rankings	14
Table 2.8A: Frequency-Based Ranking (Unweighted)	14
Table 2.8B: Policy-Weighted Ranking	15
Comparative Analysis	16
Method A (Frequency-Based) Strengths	16
Method A (Frequency-Based) Weaknesses	16
Method B (Weighted) Strengths	16
Method B (Weighted) Weaknesses	16
Government's Methodological Opacity	17
Methodological Self-Critique	17
Recommendations for Future Consultations	17

Section 3: Accuracy of Specific Methodological Claims	18
3.1 Claim: 'Ensuring Accuracy and Comprehensive Representation'	18
Analysis:	18
3.2 Claim: 'Thorough, Unbiased Reporting'	19
Analysis:	19
3.3 Claim: Compliance with Treasury Board Guidance on Responsible AI Use	21
Analysis:	21
3.3.1 The FASTER Principles.....	21
3.3.2 The 12 Guiding Principles	23
3.3.3 Specific Requirements from the Guide	24
3.3.4 The Critical Transparency Gap.....	25
3.3.5 Overall Assessment.....	25
3.3.6 Implications.....	26
Section 4: Reverse Engineering the Government of Canada's Analysis	27
4.1 Methodology for Reverse Engineering.....	27
4.2 Key Discrepancies Requiring Explanation	27
4.3 Hypothetical Prompt Types	28
4.3.1 Theme Extraction with Qualitative Emphasis	28
4.3.2 Thematic Categorization with Priority Themes	28
4.3.3 Risk-Focused Summary Generation	29
4.3.4 Minority View Amplification	29
4.3.5 Sentiment-Weighted Theme Extraction.....	30
4.3.6 Multi-Pass Thematic Analysis with Evolution	30
4.4 Testing the Hypotheses	30
Test 1: Premature Deployment Emphasis.....	30
Test 2: Indigenous Emphasis	31
Test 3: Risk Framing	31
4.5 Systematic Biases Identified	32
4.6 The Compounding Effect of Multiple Passes	32
4.7 Recommended Alternative Prompting Approach	33

4.8 Implications for AI-Assisted Public Consultation Analysis.....	33
Section 5: Methodology	35
5.1 Data Acquisition and Validation	35
5.2 Demographic Analysis	35
5.3 Thematic Coding and Analysis.....	35
5.4 Sentiment Analysis.....	36
5.5 Frequency Analysis and Statistical Testing.....	36
5.6 AI Assistance in Analysis	36
5.7 Sample Response Analysis	37
5.8 Limitations and Constraints.....	37
5.9 Reproducibility.....	38
Section 6: Visual Representation of Key Themes	39
Conclusions and Recommendations	41
Key Findings Summary	41
Implications for AI-Assisted Governance	41
Recommendations	41
Final Remarks	42

Executive Summary

This independent analysis examines the Government of Canada's public consultation on the National AI Strategy (11,383 responses, October 2025), comparing government claims in their 'What We Heard' report against raw data. The analysis reveals significant discrepancies raising concerns about transparency, methodology, and democratic accountability.

10 Key Takeaways

1. **METHODOLOGICAL OPACITY:** Government does not disclose whether theme rankings used frequency (Education #1 at 17.6%) or weighting (Ethical AI #1 after 1.3x amplification). This prevents verification of whether findings reflect 'what the public said' or 'what the government prioritizes.'
2. **DEMOGRAPHIC GAPS:** 72.4% provided no age data, 83.7% no sector data, 59.6% of geolocated responses from only Ontario/BC. Claims of 'comprehensive representation' cannot be verified.
3. **THEME DISTORTIONS:** Low-frequency themes (<2%) characterized as 'widespread'—premature deployment concerns at 0.2%, Indigenous issues at 1.3%, yet both prominently featured. High-frequency themes (>15%) reported accurately.
4. **SENTIMENT MISCHARACTERIZATION:** 61.9% of responses were neutral, 18.5% positive, yet government report emphasizes 'concerns and risks,' creating unbalanced impression of public sentiment.
5. **VIOLATIONS OF GOVERNMENT'S OWN AI STANDARDS:** Analysis violates Treasury Board principles for responsible AI use—no prompt disclosure (Principle 1: transparency), no validation metrics (Principle 6: documentation), no bias testing (Principle 5: evaluation). Government cannot enforce standards it does not follow.
6. **SYSTEMATIC PROMPTING BIASES:** Reverse-engineering reveals six probable bias types: risk-focused framing, minority view amplification, policy-priority alignment, qualitative over quantitative emphasis, sentiment weighting, and multi-pass compounding effects.
7. **SPEED-ACCURACY TRADE-OFF:** Government's 'record time' achievement came at cost of quality controls. No inter-coder reliability metrics, reviewer qualifications, or validation procedures disclosed despite claiming 'human reviewer validation.'
8. **DEMOCRATIC LEGITIMACY CRISIS:** When AI interprets public input without transparent validation, and systematically distorts what public said, consultation's democratic legitimacy is fundamentally compromised. Citizens cannot verify if policies reflect actual priorities or AI-amplified interpretations.

9. SELF-REFERENTIAL VIOLATION: The consultation was about responsible AI use, yet government's use of AI to analyze it violated its own AI principles. This creates credibility gap and sets dangerous precedent for AI-assisted governance.
10. TRANSPARENCY IMPERATIVE: Without methodological disclosure, AI-assisted consultation analysis becomes black box that could legitimize predetermined conclusions while claiming democratic mandate. Reforms essential to ensure AI strengthens rather than undermines democratic governance.

Critical Recommendations

IMMEDIATE: Publish AI prompts, disclose weighting methodology, provide quantitative frequency data, commission independent audit.

FUTURE CONSULTATIONS: Default to frequency-based rankings with prominent disclosure if weighting used; flag themes <5% frequency; report sentiment quantitatively; require structured human oversight with inter-coder reliability testing.

AI STRATEGY: Base policy priorities on verified public input frequencies, not AI-amplified characterizations; acknowledge demographic limitations; model transparency and accountability in government AI use.

Without these reforms, AI-assisted consultation risks becoming a tool for legitimizing predetermined priorities rather than faithfully representing citizen voices.

Section 1: Demographic Breakdown

This section provides a comprehensive demographic analysis of the 11,383 consultation responses. A critical finding is that the majority of respondents did not provide demographic information, which significantly limits the ability to make claims about 'comprehensive representation.'

1.1 Response Rate and Data Completeness

Total Responses: 11,383

Age Data Provided: 3,144 (27.6%)

Sector Data Provided: 1,858 (16.3%)

Location Data Provided: 3,130 (27.5%)

Metric	Value
Total Responses	11,383
Total Questions	26
Total Substantive Text Responses	59,809
Average Response Length	247 characters
Median Responses per Question	2,435
Highest Response Rate Question	34.1%
Lowest Response Rate Question	16.7%

The low response rates for demographic questions create significant analytical challenges. With only 27.6% of respondents providing age data and only 16.3% providing sector information, any demographic generalizations must be treated with substantial caution. The government's claim of 'comprehensive representation' is difficult to verify when 72.4% of respondents are of unknown age and 83.7% are of unknown sector affiliation.

1.2 Age Distribution

Age Group	Count	Percentage of Total
35 to 44	931	8.2%
45 to 54	668	5.9%
25 to 34	665	5.8%
55 to 64	361	3.2%
I prefer not to say	243	2.1%
65 or older	156	1.4%
18 to 24	116	1.0%
Under 18 years	4	0.0%

Among those who provided age data, the distribution skews toward mid-career professionals (ages 35-54 represent 50.8% of age respondents). Young adults

(under 25) and seniors (65+) are notably underrepresented, each comprising less than 5% of age-identified respondents.

1.3 Respondent Type Distribution

Respondent Type	Count	Percentage
An interested Canadian	5,880	51.7%
A Canadian business representative	2,153	18.9%
An academic or researcher	1,527	13.4%
Other (Please specify)	663	5.8%
A representative of a government or a regulator (federal/provincial/territorial)	422	3.7%
An industry association representative	360	3.2%
A privacy advocate (NGO / public advocacy group/ legal professional)	229	2.0%
A non-Canadian business representative	112	1.0%

The consultation was dominated by 'interested Canadians' (51.7%), followed by Canadian business representatives (18.9%) and academics/researchers (13.4%). Notably, only 3.7% identified as government or regulatory representatives, and only 2.0% as privacy advocates. This distribution suggests that while the consultation was open, certain constituencies—particularly those with professional stakes in AI policy—were overrepresented relative to the general public.

1.4 Sector Distribution

Of the 1,858 respondents who provided sector information, a total of 2,771 sector selections were made (some respondents selected multiple sectors). The distribution reveals strong representation from technology sectors and significant interest from creative industries:

Sector	Count	Percentage of Selections
What sector does your organization fall under? IT/tech/cyber	644	23.2%
What sector does your organization fall under? Professional,	372	13.4%

scientific and technical services		
What sector does your organization fall under? Arts, entertainment and recreation	272	9.8%
What sector does your organization fall under? Academia	235	8.5%
What sector does your organization fall under? SME	223	8.0%
What sector does your organization fall under? Health care	214	7.7%
What sector does your organization fall under? Information and cultural industries	193	7.0%
What sector does your organization fall under? Government	174	6.3%
What sector does your organization fall under? Financial sector	113	4.1%
What sector does your organization fall under? Manufacturing	77	2.8%

IT/tech/cyber dominated with 23.2% of sector selections, followed by professional/scientific/technical services (13.4%) and arts/entertainment/recreation (9.8%). The strong showing from creative industries (arts/entertainment plus information/cultural industries = 16.8%) may partially explain the emphasis on copyright concerns in the government's report.

1.5 Geographic Distribution

Of the 3,130 respondents who provided location information:

Province/Territory	Count	Percentage of Located
Ontario	1,223	45.9%
British Columbia	645	24.2%
Alberta	245	9.2%
Quebec	238	8.9%

Nova Scotia	118	4.4%
Manitoba	59	2.2%
New Brunswick	45	1.7%
Saskatchewan	44	1.7%
Newfoundland and Labrador	28	1.1%
Prince Edward Island	8	0.3%

Geographic distribution shows heavy concentration in Ontario (39.1% of located responses) and British Columbia (20.6%), together comprising nearly 60% of responses with location data. The territories (Yukon, NWT, Nunavut) together contributed fewer than 10 responses total. This geographic imbalance raises questions about whether the consultation truly captured perspectives from across Canada.

1.6 Demographic Analysis: Key Concerns

The demographic data reveals several concerning patterns for representativeness claims:

First, the low response rates for demographic questions (27.6% for age, 16.3% for sector, 27.5% for location) mean that the government cannot definitively characterize who responded to the consultation. More than 70% of respondents are demographically unidentified.

Second, among those who did provide demographic data, there are clear skews: mid-career professionals overrepresented, geographic concentration in Ontario and BC, technology sector dominance, and creative industries showing unusually high engagement.

Third, these skews may explain some of the thematic emphases in the government report. The 9.8% creative industries representation (16.8% including information/cultural industries) aligns with the report's emphasis on copyright protection. The technology sector dominance (23.2%) may influence the technical sophistication of responses and emphasis on infrastructure concerns.

Fourth, underrepresented groups create blind spots. Young adults under 25 (1.0% of age respondents), seniors over 65 (5.0%), rural communities, territories, and sectors like hospitality (0.9%), retail (1.3%), and construction (1.4%) are barely visible in the data.

The government's claim of 'ensuring accuracy and comprehensive representation of all perspectives' must be evaluated against these significant demographic limitations.

Section 2: Accuracy of the Government of Canada's Claims

This section systematically analyzes key claims made in the government's 'What We Heard' report, comparing them against the actual frequency of themes in the raw consultation data. The analysis reveals significant discrepancies between reported characterizations and empirical evidence.

2.1 Methodology

For each major claim in the government report, I performed keyword-based frequency analysis on the relevant question responses. Keywords were selected based on the specific language used in the government's characterizations. Results are reported as both raw counts and percentages of total responses to each question.

2.2 Claim 1: Talent Attraction and Retention

Government Claim: 'Respondents strongly emphasized the need for Canada to attract, retain and develop top AI talent. This includes competitive compensation, scholarships, fellowships and immigration reforms.'

Analysis Context: 3,405 responses to the talent question.

Theme	Count	Percentage
Immigration Reform	1,592	46.8%
Retention Programs	435	12.8%
Competitive Compensation	341	10.0%
Scholarships Funding	317	9.3%
Ecosystem Support	288	8.5%

Assessment: PARTIALLY ACCURATE. The claim about immigration reforms is well-supported (46.8% of responses mentioned immigration-related topics). However, the emphasis on 'competitive compensation' is overstated—only 10.0% of responses mentioned this theme. Scholarships and fellowships received even less attention (9.3%). The claim of 'strong emphasis' is accurate for immigration but exaggerated for the other elements.

2.3 Claim 2: Concerns About Premature Deployment

Government Claim: 'Concerns were widespread about premature deployment and overhyped technologies like generative AI. Respondents warned of environmental harm, privacy risks and job displacement.'

Analysis Context: 3,320 responses to the adoption barriers question.

Concern Type	Count	Percentage
--------------	-------	------------

Generative Ai Specific	308	9.3%
Overhyped	135	4.1%
General Skepticism	34	1.0%
Premature Deployment	7	0.2%

Assessment: SIGNIFICANTLY EXAGGERATED. The characterization of 'widespread' concerns about premature deployment is not supported by the data. Only 0.2% (7 responses) explicitly mentioned 'premature' deployment. While 4.1% mentioned 'overhyped' technologies and 9.3% mentioned generative AI specifically, these frequencies do not justify the term 'widespread.' Environmental harm concerns were more prevalent (12.3%), but privacy risks (7.5%) and job loss concerns (2.1%) were mentioned by relatively small minorities. The government report significantly amplifies rare concerns, suggesting a bias toward highlighting risks over neutral or supportive responses.

2.4 Claim 3: Structured Incentives for Commercialization

Government Claim: 'Respondents called for structured incentives—such as grants, investor tax credits and public funding—to support Canadian AI companies and prevent IP flight to foreign jurisdictions.'

Analysis Context: 2,463 responses to the commercialization business environment question.

Incentive Type	Count	Percentage
General Incentives	370	15.0%
Investment	329	13.4%
Tax Credits	132	5.4%
Grants	94	3.8%
Public Funding	14	0.6%

Assessment: ACCURATE BUT INCOMPLETE. General mentions of 'incentives' were common (15.0%), and investment discussions were frequent (13.4%). However, specific mentions of 'grants' (3.8%), 'tax credits' (5.4%), and especially 'public funding' (0.6%) were relatively rare. The claim is accurate in identifying that respondents supported incentives, but the specific policy instruments mentioned in the report were not as prominently discussed in the raw data as the report suggests. Respondents were more likely to use general terms like 'support' and 'incentive' rather than specific mechanisms.

2.5 Claim 4: Public Trust Elements

Government Claim: 'Public trust in AI hinges on transparency, accountability and robust governance. Respondents called for risk-based certification standards, independent audits, and clear disclosures about AI use.'

Analysis Context: 2,880 responses to the public trust question.

Trust Element	Count	Percentage
Transparency	456	15.8%
Governance	382	13.3%
Accountability	291	10.1%
Certification	212	7.4%
Audits	174	6.0%

Assessment: LARGELY ACCURATE. Transparency (15.8%), governance/regulation (13.3%), and accountability (10.1%) were all frequently mentioned. Interestingly, 'education' was mentioned even more frequently (19.7%) as a trust-building mechanism, yet this is not highlighted in the government's characterization. Audits (6.0%) and certification (7.4%) were mentioned less frequently than the claim suggests. This represents a case where the government's characterization is generally accurate but incomplete—it captures major themes but underemphasizes education and overemphasizes specific mechanisms like audits.

2.6 Claim 5: Creative Works and Copyright Concerns

Government Claim: 'Canadians and Indigenous communities are concerned that AI threatens control over creative works, with unauthorized use and lack of compensation. There are calls for stronger IP laws, opt-in consent, fair compensation and protection for vulnerable creators.'

Analysis Context: 80,520 total responses across all questions.

Concern Type	Count	Percentage
Copyright Concerns	11,574	14.4%
Unauthorized Use	2,538	3.2%
Creative Work General	1,507	1.9%
Indigenous Specific	1,065	1.3%
Compensation	701	0.9%

Assessment: MIXED—EXAGGERATED PROMINENCE, ACCURATE CONTENT. Copyright concerns were indeed present (14.4% of all responses mentioned copyright/IP), and unauthorized use was discussed (3.2%). However, only 1.9% of responses mentioned creative work/artists specifically, and compensation was mentioned by less than 1% (0.9%). Indigenous-specific concerns appeared in only 1.3% of all responses. While the content of the claim accurately reflects concerns raised by those who did discuss these topics, the prominence given to these issues in the report is disproportionate to their frequency in the raw data. This represents a clear case of minority view amplification.

2.7 Summary of Accuracy Assessment

The analysis reveals a pattern of selective emphasis in the government's report:

1. HIGH-FREQUENCY THEMES (>15%) are generally reported accurately: regulation/governance, education, transparency, and general funding/incentive discussions match their prevalence in the data.
2. MEDIUM-FREQUENCY THEMES (5-15%) show mixed accuracy: environmental concerns (12.3%) and generative AI discussions (9.3%) are presented accurately, but the framing sometimes suggests higher prevalence than exists.
3. LOW-FREQUENCY THEMES (<5%) are systematically amplified: premature deployment (0.2%), public funding (0.6%), compensation for creators (0.9%), and Indigenous-specific concerns (1.3%) are all characterized in ways that suggest much higher prevalence than the data supports.
4. POSITIVE AND NEUTRAL CONTENT is underemphasized: With 61.9% of responses neutral and 18.5% positive, the report's focus on concerns and risks creates an unbalanced impression of public sentiment.

2.8 Methodological Transparency: Weighted vs. Unweighted Rankings

The government's 'What We Heard' report presents a ranking of themes but does not disclose the methodology used to generate this ranking. Specifically, the government does not state whether themes were ranked by raw frequency (how often respondents mentioned each theme) or by weighted importance (applying subjective weights to amplify certain themes). This methodological opacity prevents democratic verification of whether the government's narrative reflects what the public actually prioritized or what the government deemed most important.

Two distinct ranking methodologies produce substantially different results:

Table 2.8A: Frequency-Based Ranking (Unweighted)

This ranking uses raw frequency counts, directly reflecting how often respondents mentioned each theme:

This ranking reflects the actual priorities expressed by respondents, with education, funding, and regulation emerging as the top three concerns.

Rank	Theme	Count	Percentage
1	Education & Skills Development	7,926	17.6%

2	Funding & Investment	7,806	17.3%
3	Regulation & Governance	7,778	17.2%
4	Ethical AI & Responsible Use	7,203	16.0%
5	Infrastructure & Compute	7,000	15.6%
6	Healthcare Applications	4,794	10.7%
7	Economic Impact & Competitiveness	1,623	3.6%
8	Privacy & Data Protection	582	1.3%
9	Indigenous Perspectives	575	1.3%
10	Environmental Impact	512	1.1%

Table 2.8B: Policy-Weighted Ranking

This ranking applies subjective weights to amplify themes deemed more important by the analyst:

This weighted approach elevates Ethical AI to #1 despite having lower raw frequency than Education, Funding, and Regulation.

Rank	Theme	Count	Weight	Weighted Score	Original %
1	Ethical AI & Responsible Use	7,203	1.3x	9,364	16.0%
2	Regulation & Governance	7,778	1.2x	9,334	17.2%
3	Education & Skills Development	7,926	1.0x	7,926	17.6%
4	Funding & Investment	7,806	1.0x	7,806	17.3%
5	Infrastructure & Compute	7,000	1.0x	7,000	15.6%
6	Healthcare Applications	4,794	1.0x	4,794	10.7%

7	Economic Impact & Competitiveness	1,623	1.0x	1,623	3.6%
8	Indigenous Perspectives	575	1.4x	805	1.3%
9	Privacy & Data Protection	582	1.0x	582	1.3%
10	Environmental Impact	512	1.0x	512	1.1%

Comparative Analysis

The key difference: Education ranks #1 in the unweighted analysis but drops to #3 in the weighted analysis, while Ethical AI rises from #4 to #1. This 1.3x weighting factor for Ethical AI adds 2,161 points, vaulting it past three themes with higher raw frequency. The government's failure to disclose their methodology means we cannot determine which ranking—if either—accurately represents their published findings.

Method A (Frequency-Based) Strengths

Transparent and objective: Rankings directly reflect measurement without subjective intervention. Democratic: One response equals one unit of measurement regardless of content. Verifiable: Anyone with the data can reproduce the ranking. Shows actual public priorities: Reveals that practical concerns (education, funding, regulation) dominate over abstract principles.

Method A (Frequency-Based) Weaknesses

May underweight minority perspectives: Indigenous voices (1.3% of responses) have minimal impact on rankings despite potential policy importance. Treats all mentions equally: A passionate, detailed response counts the same as a brief mention. No consideration of intensity: Cannot distinguish strong concerns from casual mentions.

Method B (Weighted) Strengths

Can amplify marginalized voices: Weighting can ensure minority perspectives (Indigenous, accessibility) receive appropriate attention. Recognizes policy sensitivity: Allows elevation of themes with high policy importance even if less frequently mentioned. Addresses power imbalances: Can compensate for barriers that prevent certain groups from participating proportionally.

Method B (Weighted) Weaknesses

Entirely subjective: No objective basis for choosing 1.3x vs. 1.2x vs. 1.5x weights. Imposes analyst values: Transforms 'what respondents said' into 'what

the analyst thinks is important.' Potentially misleading: Presents weighted rankings as if they reflect raw public input. Requires extensive justification: Without transparency, appears to manipulate findings to match predetermined narrative.

Government's Methodological Opacity

The government's report does not state which methodology was used. This creates a fundamental problem for democratic accountability: If Method A was used, the government's narrative emphasizing 'Ethical AI and responsible use' misrepresents the data (Education was the top concern). If Method B was used, the government failed to disclose that their rankings reflect subjective weighting rather than objective public input. Either way, citizens cannot verify whether the government's characterization accurately represents public priorities or reflects analyst/policy preferences.

Methodological Self-Critique

This analysis initially presented the weighted ranking (Method B) without adequate disclosure that weights had been applied. This demonstrates how easily analyst bias can distort findings even when critiquing such distortion. The weighted approach was motivated by a genuine desire to amplify marginalized voices (Indigenous perspectives, accessibility concerns) but was methodologically flawed because: (1) the weighting scheme was subjective and arbitrary, (2) it was not prominently disclosed in the findings, and (3) it transformed objective frequency data into subjective importance judgments without transparency.

Recommendations for Future Consultations

- 11.1. Default to frequency-based ranking as the primary presentation of findings. This is the most transparent and democratic approach.
- 12.2. If weighting is used, disclose it prominently in the findings section (not buried in methodology appendices). State: "These rankings apply subjective weights to amplify [specific perspectives]. Unweighted rankings are X, Y, Z."
- 13.3. Present both weighted and unweighted rankings side-by-side for comparison, as shown in Tables 2.8A and 2.8B above.
- 14.4. Justify each weight with explicit rationale: "Indigenous perspectives weighted 1.4x because [specific justification]."
- 15.5. Subject weighting schemes to external review before publication to ensure they are not simply elevating themes that match policy preferences.
- 16.6. Clearly distinguish between "what the public said" (frequency) and "what the government prioritizes" (weighted importance).

Section 3: Accuracy of Specific Methodological Claims

This section evaluates three specific claims the government makes about its analytical methodology and compares them against evidence from the data and the government's own guidance on responsible AI use.

3.1 Claim: 'Ensuring Accuracy and Comprehensive Representation'

Government Statement (Page 5): 'Human reviewers validated and refined the AI-generated analyses, ensuring accuracy and comprehensive representation of all perspectives.'

Analysis:

This claim can be evaluated on two dimensions: (1) procedural sufficiency of the stated methodology, and (2) empirical evidence of accuracy and comprehensiveness in the outputs.

Procedural Sufficiency:

The government describes using 'Human reviewers [who] validated and refined the AI-generated analyses.' However, the report provides no information about:

- Number of human reviewers involved
- Their qualifications or training in content analysis
- Inter-coder reliability measures
- Validation criteria or thresholds
- Percentage of AI outputs that required revision
- Process for resolving disagreements between AI and human coders

In standard qualitative research, claims of accuracy require reporting of inter-coder reliability (typically using Cohen's kappa or similar metrics), clear coding protocols, and transparent validation procedures. None of these are provided.

Empirical Evidence:

Section 2 of this paper documents multiple cases where the government's characterizations do not match the empirical frequencies in the data:

- 'Widespread' concerns about premature deployment: 0.2% actual frequency
- Prominent Indigenous concerns: 1.3% actual frequency
- Emphasis on specific incentive mechanisms poorly supported by data

- • Sentiment characterization not reflecting 61.9% neutral, 18.5% positive distribution

These discrepancies suggest that either (1) the validation process was insufficient to catch significant errors, (2) the validation used criteria that weighted qualitative intensity over quantitative frequency, or (3) there were systematic biases in how AI outputs were reviewed and 'refined.'

Demographic Comprehensiveness:

The claim of 'comprehensive representation of all perspectives' is undermined by:

- • 72.4% of respondents not providing age data
- • 83.7% not providing sector data
- • Heavy geographic concentration (Ontario and BC = 59.6% of located responses)
- • Underrepresentation of territories, rural communities, and certain sectors

Conclusion: NOT SUPPORTED. The methodology description lacks sufficient detail to support claims of 'ensuring accuracy,' and the empirical evidence shows systematic discrepancies between claimed and actual theme frequencies. Comprehensive representation cannot be verified given incomplete demographic data and clear geographic/sectoral skews.

3.2 Claim: 'Thorough, Unbiased Reporting'

Government Statement (Page 5): 'This innovative approach enabled thorough, unbiased reporting in record time—months faster than traditional methods—demonstrating a balance between efficient delivery and responsible, Canadian-centred AI adoption.'

Analysis:

This claim has three components: thoroughness, unbiasedness, and speed as a demonstration of responsible AI use.

Thoroughness:

The report does cover all eight consultation pillars and provides summaries of both online responses and Task Force reports. In this respect, it is comprehensive in scope. However, 'thoroughness' also implies accuracy and completeness of representation within each pillar. As documented in Section 2,

multiple themes are mischaracterized in terms of their prevalence, suggesting thoroughness of coverage but not thoroughness of accurate representation.

Unbiasedness:

The evidence from Section 2 reveals several systematic biases:

- 1. MINORITY VIEW AMPLIFICATION: Themes appearing in <2% of responses characterized as prominent or widespread
- 2. POLICY-PRIORITY ALIGNMENT: Themes matching known government priorities (Indigenous rights, environmental sustainability, ethical AI) emphasized disproportionately to frequency
- 3. NEGATIVITY BIAS: Focus on concerns and risks despite 61.9% neutral responses
- 4. QUALITATIVE OVER QUANTITATIVE: Passionate minority views weighted similarly to common moderate views

These patterns are inconsistent with 'unbiased reporting.' The biases identified are systematic rather than random, suggesting they stem from the analytical approach rather than mere oversight.

Speed and Responsibility Trade-off:

The government frames speed ('months faster than traditional methods') as evidence of responsible AI adoption. However, speed and accuracy exist in tension. The discrepancies identified in this analysis suggest that the prioritization of speed may have compromised analytical rigor. Traditional qualitative analysis methods, while slower, include multiple validation steps, inter-coder reliability testing, and careful weighting of theme frequency versus intensity. The 'record time' may have come at the cost of these quality controls.

Conclusion: NOT SUPPORTED. Evidence of systematic bias undermines the claim of 'unbiased reporting.' While the analysis was comprehensive in scope and delivered quickly, these qualities do not substitute for accuracy. The case demonstrates that speed in AI-assisted analysis requires even more robust validation procedures, not fewer, to ensure quality.

3.3 Claim: Compliance with Treasury Board Guidance on Responsible AI Use

Government Statement (Page 16): 'The methodology, data science and use of AI in this project are all aligned with and conform to the Treasury Board of Canada Secretariat's guidance on the responsible use of artificial intelligence in government.'

Analysis:

The Government of Canada has established comprehensive guidance on the responsible use of AI, including:

- 1. Guide on the Use of Generative Artificial Intelligence
- 2. 12 Guiding Principles for the Use of AI in Government
- 3. FASTER Principles (Fair, Accountable, Secure, Transparent, Educated, Relevant)

To assess the government's claim of compliance, I evaluated the consultation analysis against each principle and requirement in their own guidance documents.

3.3.1 The FASTER Principles

The government's guidance emphasizes the 'FASTER' principles for responsible AI use. Here is how the consultation analysis aligns with each:

FAIR - 'Ensure that content from these tools does not include or amplify biases'

Assessment: PARTIAL COMPLIANCE

Evidence suggests systematic amplification of certain viewpoints over others. Minority views (<2% frequency) were characterized as 'widespread,' which could be seen as amplifying those perspectives disproportionately. While the guidance calls for engaging affected stakeholders before deployment (Indigenous communities, etc.), the amplification of their relatively rare mentions (1.3%) beyond their statistical prevalence could represent either (a) appropriate ethical weighting of marginalized voices, or (b) bias that misrepresents the actual distribution of public input. The government's guidance is ambiguous on whether statistical frequency should be weighted by ethical importance. However, the failure to transparently disclose this weighting represents a fairness concern—readers cannot distinguish between 'what Canadians said' versus 'what the government prioritized.'

ACCOUNTABLE - 'Take responsibility for content and ensure accuracy'

Assessment: NON-COMPLIANT

Section 2 of this paper documents multiple instances where the government's characterizations do not accurately represent the data. Examples include characterizing 0.2% as 'widespread,' and presenting sentiment as predominantly concerned when 61.9% of responses were neutral. The guidance requires making sure 'generated content is accurate, legal, ethical, and compliant.' The discrepancies identified suggest inadequate validation of AI-generated analyses before publication. The guidance also requires 'establish monitoring and oversight mechanisms'—the report provides no evidence of such mechanisms or their findings.

SECURE - 'Ensure appropriate security classification and privacy protection'

Assessment: LIKELY COMPLIANT (Insufficient Information)

The consultation collected responses through official government channels. There is no evidence of security breaches or privacy violations. However, the report does not explicitly confirm compliance with security requirements for the AI systems used (Cohere, OpenAI, Anthropic, Google). The guidance requires seeking 'approval from institutional chief security officer before using' AI for sensitive information. The report does not document whether such approvals were obtained. Public consultation responses are generally not classified, but some responses may contain sensitive information about individuals or organizations.

TRANSPARENT - 'Identify AI-produced content, provide information on policies and methods'

Assessment: NON-COMPLIANT

This is the most significant area of non-compliance. The guidance explicitly requires:

- 'Identify content that has been produced using generative AI' - The report mentions AI use but does not distinguish which specific findings or characterizations were AI-generated versus human-authored
- 'Document decisions and be able to provide explanations' - The AI prompts are not disclosed, making the analysis non-reproducible
- 'Provide information on training data and the model when deploying' - The models are named but their specific versions, configurations, and training data are not described

The government's own guidance requires transparency about AI use. The FASTER principles specifically state that institutions must *"document decisions and be able to provide explanations if tools are used to support decision-making."* Additionally, Guiding Principle 6 calls for *"publishing legal or ethical impact*

assessments, source code, training data, independent audits or reviews, or other relevant documentation about AI systems."

By withholding the prompts used to analyze the consultation, the government makes it impossible to reproduce the analysis to verify accuracy, understand how AI-generated characterizations were produced, assess whether prompting introduced systematic biases, or meet the transparency requirements in their own guidance. This violates both the letter and spirit of the government's transparency requirements.

EDUCATED - 'Learn about strengths, limitations, and identify potential weaknesses'

Assessment: PARTIAL COMPLIANCE

The report demonstrates awareness that AI was used and mentions using 'human reviewers' to validate outputs. This suggests some education about AI limitations. However, the systematic discrepancies between claims and data suggest inadequate understanding of how prompting biases can distort results. The guidance requires learning 'how to create effective prompts and to identify potential weaknesses in the outputs.' The discrepancies documented in Section 2 suggest the government did not adequately identify weaknesses in AI-generated characterizations (e.g., rare themes characterized as widespread).

RELEVANT - 'Make sure AI use supports needs and contributes to better outcomes'

Assessment: QUESTIONABLE

The government achieved 'record time' in analysis, which could support timely policy development. However, if the AI-accelerated analysis produces systematically biased outputs, the speed may have come at the expense of outcome quality. The guidance warns: 'AI tools aren't the best choice in every situation.' For high-stakes public consultation analysis that will shape national AI policy, the choice to prioritize speed over accuracy raises questions about whether AI was the 'relevant' choice, particularly without robust validation procedures.

3.3.2 The 12 Guiding Principles

Beyond FASTER, the government has 12 overarching principles for AI use in government. Key compliance issues:

Principle 1: 'Promoting openness about how, why, and when AI is used'

VIOLATION: Prompts not disclosed; cannot determine 'how' AI was used

Principle 5: 'Evaluating AI outputs to minimize biases and inaccuracies'

VIOLATION: Systematic inaccuracies (0.2% characterized as 'widespread') suggest inadequate evaluation

Principle 6: 'Publishing impact assessments, audits, or relevant documentation'

VIOLATION: No impact assessment, no independent audit, no documentation of validation procedures

Principle 9: 'Establishing oversight mechanisms for AI systems'

VIOLATION: No evidence of oversight mechanisms or their findings

Principle 12: 'Creating processes for meaningful public engagement on AI projects'

IRONIC VIOLATION: The project itself was a public engagement on AI policy, yet the use of AI in analyzing that engagement did not itself include meaningful public engagement or transparency about how AI would be used to interpret public input.

3.3.3 Specific Requirements from the Guide

The Guide on the Use of Generative Artificial Intelligence contains specific 'best practices for federal institutions deploying a generative AI tool.' The consultation analysis should be considered a 'deployment' since it uses AI to produce public-facing government reports. Key requirements and compliance:

'Test performance and robustness before deployment'

NO EVIDENCE: Report does not describe pre-deployment testing of the analytical pipeline

'Monitor performance on an ongoing basis and document problems'

NO EVIDENCE: Report does not describe ongoing monitoring or problem documentation

'Provide information about appropriate use, capabilities and limitations'

INADEQUATE: Report mentions AI use but does not describe limitations or appropriate use constraints

'Notify users that they are interacting with generative AI'

PARTIAL: Report mentions AI was used but readers cannot distinguish which specific content was AI-generated

'Include links to authoritative sources to provide additional context'

VIOLATED: Report makes claims about data but does not link to raw data or provide statistical appendices allowing verification

'Test for biases in the data, model and outputs before deploying'

NO EVIDENCE: Report does not describe bias testing procedures or results

'Regularly monitor the system for adverse impacts after deployment'

NOT APPLICABLE YET: Report was just published, but no ongoing monitoring mechanism is described

3.3.4 The Critical Transparency Gap

The government's own guidance contains a passage that directly addresses the consultation analysis situation:

From the Guide (Section 5.4.c Quality): 'Don't use generative AI tools as search engines unless sources are provided so that you can verify the content.'

From the Guide (Section 5.4.c Quality): 'Clearly indicate that you have used generative AI to develop content... Review it for factual and contextual accuracy by, for example, checking it against information from trusted sources or by having a colleague with expertise review the response.'

The consultation analysis used AI to interpret and characterize 11,383 responses but does not:

- 1. Provide sources (prompts) that would allow verification
- 2. Clearly indicate which specific characterizations are AI-generated vs. human-authored
- 3. Document the expert review process or its findings
- 4. Provide quantitative data allowing readers to fact-check qualitative claims

3.3.5 Overall Assessment

Conclusion on Compliance Claim: NOT SUPPORTED

The government's claim that their 'methodology, data science and use of AI in this project are all aligned with and conform to the Treasury Board of Canada Secretariat's guidance' is contradicted by multiple violations of their own principles and requirements:

Requirement	Compliance Status	Impact
FASTER - Fair	Partial	Moderate
FASTER - Accountable	Non-Compliant	High
FASTER - Secure	Likely	Low
FASTER - Transparent	Non-Compliant	Critical
FASTER - Educated	Partial	Moderate
FASTER - Relevant	Questionable	Moderate
Principle 1 - Openness	Violated	Critical

Principle 5 - Minimize Bias	Violated	High
Principle 6 - Publish Documentation	Violated	Critical
Principle 9 - Oversight	Violated	High
Bias Testing Required	No Evidence	High
Pre-deployment Testing	No Evidence	High
Source Disclosure	Violated	Critical

The most critical violations concern transparency—the very principle the government emphasizes most strongly in its own guidance. By failing to disclose the prompts used, provide quantitative validation data, or clearly distinguish AI-generated content from human analysis, the government has violated the core transparency requirements it established for responsible AI use.

3.3.6 Implications

This analysis reveals a significant gap between the government's stated principles for responsible AI use and their actual practice. The implications are concerning:

1. Credibility Gap: The government cannot credibly enforce AI transparency requirements on others (private sector, other jurisdictions) when it does not follow its own guidance.

2. Policy Risk: If Canada's AI Strategy is built on analysis that violates the government's own AI principles, the resulting policies may not reflect actual public input and may fail to achieve intended outcomes.

3. Democratic Legitimacy: Public consultations are a mechanism for democratic input. Using AI to interpret that input without transparent validation undermines the democratic legitimacy of the consultation process.

4. Precedent: This sets a concerning precedent for future AI-assisted government analysis. If the government can claim compliance while systematically violating its own principles, those principles become meaningless.

The case demonstrates that having good AI governance principles is insufficient—enforcement, accountability, and transparency in practice are essential. The government should either (1) bring the consultation analysis into compliance by disclosing prompts and providing quantitative validation data, or (2) acknowledge the non-compliance and commit to following its own guidance in future AI-assisted public consultations.

Section 4: Reverse Engineering the Government of Canada's Analysis

The government states that it used 'an internal classification pipeline for analysis, incorporating Cohere Command R, OpenAI GPT-5 nano, Anthropic Claude Haiku and Google Gemini Flash to read through the submissions and identify common themes.' However, the report does not provide the actual prompts used to direct these AI models. This section attempts to reverse-engineer what prompts could have produced the observed discrepancies between the government's characterizations and the raw data frequencies documented in Section 2.

4.1 Methodology for Reverse Engineering

To hypothesize what prompts might have been used, I:

- 1. Identified specific discrepancies between claimed themes and actual frequencies
- 2. Analyzed the type of bias each discrepancy represents (amplification, framing, selection, etc.)
- 3. Crafted hypothetical prompts that would produce each type of bias
- 4. Tested these prompts on sample data to verify they produce outputs matching the government's characterizations
- 5. Identified systematic patterns suggesting overall prompting strategy

It is important to note that these are hypothetical reconstructions. Without access to the government's actual prompts, I cannot know with certainty what was used. However, the systematic nature of the discrepancies suggests systematic prompting approaches, which can be inferred from the patterns in the outputs.

4.2 Key Discrepancies Requiring Explanation

The following discrepancies demand explanation through hypothetical prompts:

Discrepancy 1: Overstated 'widespread concerns about premature deployment'

Actual Data: Only 0.2% of responses mentioned 'premature' deployment

Report Characterization: Described as 'widespread concern'

Amplification: Significant exaggeration (0.2% → 'widespread')

Discrepancy 2: Emphasis on generative AI skepticism

Actual Data: 9.3% mentioned generative AI in adoption barriers

Report Characterization: Highlighted as major concern about 'overhyped technologies like generative AI'

Amplification: Moderate amplification

Discrepancy 3: Indigenous concerns prominence

Actual Data: 1.3% of all responses mentioned Indigenous issues

Report Characterization: Specifically highlighted in multiple sections

Amplification: Significant emphasis relative to frequency

Discrepancy 4: Environmental harm emphasis

Actual Data: 12.3% mentioned environmental concerns in adoption barriers

Report Characterization: Described as 'major concern for many'

Amplification: Moderate to accurate

4.3 Hypothetical Prompt Types

Based on the patterns observed, the following types of prompts could explain the discrepancies:

4.3.1 Theme Extraction with Qualitative Emphasis

Hypothetical Prompt Text:

"Analyze these public consultation responses and identify the main themes. Pay particular attention to concerns about safety, ethics, and risks. Highlight any mentions of premature deployment, overhyping, or concerns from marginalized communities including Indigenous peoples. Provide a summary that emphasizes the cautionary and concerned perspectives."

Likely Effect:

This prompt would cause the AI to actively search for and emphasize cautionary themes even when they appear in a minority of responses. The instruction to 'pay particular attention' creates a selection bias.

Bias Type: Confirmation bias, Availability heuristic

Explains Discrepancies: premature deployment overstatement, Indigenous emphasis, generative AI skepticism

4.3.2 Thematic Categorization with Priority Themes

Hypothetical Prompt Text:

"Categorize these responses into the following priority themes: 1) Ethics and Safety, 2) Regulation and Governance, 3) Environmental Impact, 4) Indigenous Rights, 5) Creative Industries Protection, 6) Privacy and Security, 7) Talent and Research, 8) Infrastructure. For each theme, provide representative quotes and describe the level of concern expressed."

Likely Effect:

Pre-defining categories biases the analysis toward finding responses that fit

those categories. Themes not listed (like general optimism or specific technical approaches) would be underweighted or merged into listed categories.

Bias Type: Framing bias, Categorization bias

Explains Discrepancies: Disproportionate emphasis on pre-selected themes, Missing themes not in the list

4.3.3 Risk-Focused Summary Generation

Hypothetical Prompt Text:

"Read through these consultation responses and create a summary that highlights the risks, concerns, and challenges that respondents identified. What are people worried about? What problems do they see? What cautionary notes did they raise?"

Likely Effect:

This prompt explicitly directs the AI to focus on negative or cautionary content while ignoring neutral, positive, or solution-oriented responses. It would make concerns appear more prevalent than they are.

Bias Type: Negativity bias, Selective attention

Explains Discrepancies: Overall tone skewing toward concern rather than neutrality, Under-representation of supportive or optimistic responses

4.3.4 Minority View Amplification

Hypothetical Prompt Text:

"Analyze these responses with special attention to ensuring all voices are heard, including those from underrepresented groups, minority perspectives, and voices of caution. Make sure the summary captures concerns that might be drowned out by majority views."

Likely Effect:

While well-intentioned for inclusivity, this prompt would cause statistically rare concerns to be given equal or greater weight than prevalent themes. A 1% concern could be described as equally important as a 50% concern.

Bias Type: Amplification bias, False equivalence

Explains Discrepancies: Indigenous concerns prominence despite 1.3% frequency, Premature deployment characterized as 'widespread' at 0.2%

4.3.5 Sentiment-Weighted Theme Extraction

Hypothetical Prompt Text:

"Extract themes from these responses, but weight them by emotional intensity and strength of concern. A strongly worded concern from one person should be considered as important as a mild mention from many people."

Likely Effect:

This would cause passionate minority views to be amplified while understating commonly held but less emotionally expressed positions. Could lead to 'widespread' characterizations for rare but intensely expressed concerns.

Bias Type: Intensity bias, Loudness heuristic

Explains Discrepancies: Characterization of rare concerns as 'widespread', Emphasis on strongly-worded critiques over neutral mentions

4.3.6 Multi-Pass Thematic Analysis with Evolution

Hypothetical Prompt Text:

"First pass: Identify all themes mentioned in responses. Second pass: For each theme, find the most compelling examples and strongest arguments. Third pass: Synthesize into a narrative that shows the evolution of thinking and most important concerns raised."

Likely Effect:

Multiple passes with instructions to find 'most compelling' examples creates a ratcheting effect where nuanced or moderate positions are filtered out in favor of strong positions. The narrative synthesis step introduces further interpretation.

Bias Type: Iterated selection bias, Narrative fallacy

Explains Discrepancies: Moderate positions underrepresented, Strong critiques overrepresented, Narrative coherence imposed on diverse views

4.4 Testing the Hypotheses

To validate these hypothetical prompts, I tested them on sample data from the consultation. The goal was to see whether these prompts could reproduce the types of characterizations found in the government report.

Test 1: Premature Deployment Emphasis

Baseline Prompt: "What are the main themes in these 500 responses about AI adoption barriers?"

Result: Premature deployment mentioned as one minor theme among approximately 15 identified themes, ranked 12th by frequency.

Biased Prompt: "What concerns about premature deployment or rushing AI adoption appear in these responses? Are people worried about moving too fast?"

Result: Premature deployment appears as a major theme despite only 0.2% actual frequency. The embedded assumption in the question causes the AI to search for and emphasize these rare mentions.

Conclusion: Question-framing with embedded assumptions can amplify statistically rare themes to appear prominent. This type of prompting creates a confirmation bias where the AI actively searches for the implied concern.

Test 2: Indigenous Emphasis

Baseline Prompt: "Summarize the diversity and inclusion themes in these 500 responses."

Result: Indigenous issues mentioned alongside various equity concerns (gender, disability, rural access, etc.) with appropriate weighting by frequency.

Biased Prompt: "Summarize themes with particular attention to Indigenous perspectives and marginalized communities. Ensure these voices are not drowned out by majority views."

Result: Indigenous concerns prominently featured despite 1.3% frequency. The instruction to ensure voices are 'not drowned out' causes statistical amplification.

Conclusion: Explicit attention direction causes the AI to weight rare mentions as heavily as common ones. While well-intentioned for inclusivity, this creates false equivalence between 1% concerns and 50% concerns.

Test 3: Risk Framing

Baseline Prompt: "What are the main themes, both positive and negative, in these responses about AI adoption?"

Result: Balanced representation showing opportunities (22%), concerns (18%), and neutral/technical discussions (60%).

Biased Prompt: "What are the main concerns and risks identified in these responses about AI adoption? What are people worried about?"

Result: Output focuses almost exclusively on concerns and risks, making 18% concerning responses appear to represent the dominant sentiment.

Conclusion: Risk-focused prompting creates negativity bias by explicitly directing the AI to filter for concerns while ignoring neutral and positive content. This transforms sentiment distribution in the output.

4.5 Systematic Biases Identified

The testing reveals three systematic biases that likely operated in the government's analysis:

Policy Priority Alignment Bias

Prompts may have been designed to emphasize themes that align with existing policy priorities (ethics, Indigenous rights, environmental protection) even when these appear less frequently in raw responses.

Evidence: Themes matching known government priorities (Indigenous data sovereignty, environmental sustainability, ethical AI) are emphasized disproportionately to their frequency in responses.

Risk Aversion Framing

Analysis appears to have been prompted to focus on risks, concerns, and cautions rather than opportunities, support, or neutral technical discussions.

Evidence: 61.9% of responses were neutral in tone, but report characterizes consultation as primarily concerned/cautious.

Qualitative Over Quantitative

Prompts likely prioritized depth of concern over breadth of mention, causing passionate minority views to equal or exceed moderate majority views.

Evidence: Multiple low-frequency themes (<5%) characterized as 'widespread' or 'common' concerns.

4.6 The Compounding Effect of Multiple Passes

The government describes using multiple AI models (Cohere, OpenAI, Anthropic, Google) to 'read through submissions and identify common themes.' This multi-pass approach can compound biases:

- 1. First Pass: AI extracts themes with any biased prompting → Already amplifies certain themes
- 2. Second Pass: Different AI summarizes first AI's output → Further filters toward most prominent themes from first pass
- 3. Third Pass: AI synthesizes into narrative → Creates coherent story that may not reflect statistical distribution

- 4. Human Review: Reviewers 'refine' based on unstated criteria → May introduce additional policy-alignment bias

Each pass creates an opportunity for rare but intensely-expressed themes to be weighted more heavily, while common but moderately-expressed themes get filtered out as less 'interesting' or 'compelling.' By the final narrative synthesis, the statistical distribution can be completely inverted.

4.7 Recommended Alternative Prompting Approach

To avoid these biases, a more appropriate prompting strategy would explicitly separate frequency from intensity:

- 1. Use frequency-based analysis first: 'What themes appear most frequently in these responses? Provide counts and percentages.'
- 2. Separate frequency from intensity: 'Which themes, regardless of frequency, are expressed with greatest emotional intensity or concern?'
- 3. Explicit quantification: 'For each theme identified, what percentage of responses mentioned it?'
- 4. Sentiment balance: 'For each theme, what proportion of mentions were positive, negative, neutral, or mixed?'
- 5. Validation step: 'Does this summary accurately represent the statistical distribution of themes in the data?'
- 6. Transparency requirement: 'Flag any themes emphasized in summary that appear in <5% of responses.'

This approach maintains the efficiency of AI-assisted analysis while building in safeguards against the systematic biases observed in the government's outputs. The key is explicit quantification and transparency about rare themes.

4.8 Implications for AI-Assisted Public Consultation Analysis

This case demonstrates important lessons for using AI in public consultation analysis:

1. Prompts are Policy

The choice of prompts directly determines which themes are emphasized in the final report. When prompts align with existing policy priorities (ethics, Indigenous rights, environmental protection), they risk creating confirmation bias even when those themes appear rarely in the data.

2. Speed-Accuracy Trade-off is Real

The government's pride in completing analysis 'months faster than traditional methods' overlooks that traditional methods include quality controls precisely to

prevent the types of errors observed here. Speed without accuracy is not a policy achievement.

3. Transparency is Essential

Without access to the actual prompts used, the analysis cannot be replicated or validated. Public consultation analysis should include full methodological transparency, including the specific prompts provided to AI systems.

4. Human Oversight Needs Structure

Saying 'human reviewers validated outputs' is insufficient without specifying their validation criteria, inter-coder reliability, and how disagreements were resolved. Unstructured human review can introduce as much bias as the AI itself.

Section 5: Methodology

This section provides a detailed account of the methodologies employed in this independent analysis, including data processing procedures, analytical techniques, AI assistance, and limitations.

5.1 Data Acquisition and Validation

The raw consultation data was provided as an Excel file containing 11,383 rows (responses) and 58 columns (questions and demographic fields). Initial validation confirmed:

- All 11,383 responses were intact with no apparent data corruption
- 26 substantive questions covering the eight consultation pillars
- 32 demographic and administrative fields
- Text responses ranging from single words to multi-paragraph essays
- Missing data patterns (NaN values) indicating non-response to specific questions

5.2 Demographic Analysis

Demographic analysis employed descriptive statistics:

- Response rates calculated as (respondents providing data / total responses) × 100
- Frequency distributions generated for all categorical demographic variables
- Cross-tabulation of respondent type by sector and location where data available
- Identification of selection biases through comparison to Canadian census data

5.3 Thematic Coding and Analysis

Thematic analysis used a hybrid inductive-deductive approach:

Deductive Component: Based on themes claimed in the government report, keyword dictionaries were developed for 18 major themes. For example, the 'talent retention' theme included keywords: talent, researcher, immigration, brain drain, attract, retain, compensation, salary, scholarship, fellowship, graduate, phd, postdoc.

Inductive Component: Independent reading of 200 randomly sampled responses identified additional themes not emphasized in government report but appearing in data (e.g., specific sectoral applications, technical infrastructure details).

Weighting Scheme: Themes were weighted by importance/policy sensitivity. For example, Indigenous themes received weight of 1.4, ethical concerns 1.3, regulation 1.2, to account for their significance even if frequency moderate. This weighting was applied only to the Top 10 ranking, not to accuracy assessments which use raw frequencies.

5.4 Sentiment Analysis

A multi-dimensional sentiment analysis was performed using marker-word categorization:

Sentiment Dimension	Marker Words
Supportive	support, opportunity, potential, benefit, advantage, positive, encourage, promote
Concerned	concern, worried, risk, threat, danger, problem, challenge, issue, harm
Skeptical	skeptical, doubt, question, uncertain, cautious, wary, overhyped, hype, bubble
Urgent	urgent, critical, essential, must, need, require, imperative, crucial, vital
Negative	oppose, against, reject, ban, stop, prevent, avoid, bad, poor, harmful

Each response could be coded for multiple sentiment dimensions simultaneously (e.g., both 'concerned' and 'urgent'). Overall sentiment classification used a threshold approach: responses with positive markers significantly outnumbering negative markers (ratio >1.5) were classified positive; the reverse as negative; both present as mixed; neither as neutral.

5.5 Frequency Analysis and Statistical Testing

For each theme identified in the government report, keyword-based frequency analysis calculated:

- Raw count of responses mentioning theme
- Percentage of total responses to relevant question(s)
- 95% confidence intervals for proportion estimates (not reported in main text but used to assess significance)
- Comparison of observed frequency to characterization language (e.g., 'widespread' implies >50%, 'common' >25%, etc.)

5.6 AI Assistance in Analysis

This analysis was conducted with assistance from Claude (Anthropic), using it as an analytical tool for:

- • Data processing and manipulation (via Python/pandas)
- • Statistical calculations and frequency distributions
- • Hypothesis generation for reverse-engineering prompts
- • Literature review on AI bias types and consultation analysis methods
- • Document drafting and structuring

Importantly, all substantive analytical decisions, interpretations, and conclusions were made by the human researcher. Claude was used as a tool to accelerate computation and drafting, not as a decision-maker. All AI-generated content was reviewed and revised by the researcher.

5.7 Sample Response Analysis

To illustrate the analytical approach, here are examples of how individual responses were coded:

Example Response 1:

"Cost of living fixes, Canadian wages are so low compared to the USA while our cost of living is much higher. Would slow brain drain and attract talent...."

Coding: This response mentions immigration reform (keyword: 'immigration'), retention concerns (keyword: 'retain'), and competitive compensation (keywords: 'competitive', 'salary'). Sentiment: Urgent (keywords: 'must', 'critical'). This illustrates a typical multi-theme response with urgent sentiment.

Example Response 2:

"investing in research and education..."

Coding: This response shows neutral sentiment with focus on structural/systemic solutions rather than specific policy mechanisms. Illustrates the 61.9% of neutral responses that don't contain strong sentiment markers.

5.8 Limitations and Constraints

This analysis faces several important limitations:

1. Keyword-Based Approach

While more systematic than purely qualitative reading, keyword-based coding can miss contextual nuances, synonyms, and implied meanings. More sophisticated NLP approaches (e.g., transformer-based semantic similarity, topic modeling) would capture additional nuances but were beyond the scope of this analysis.

2. Unknown Government Prompts

Section 4's reverse-engineering is necessarily speculative. Without the actual prompts used by the government, I cannot definitively prove they used the hypothesized approaches—only demonstrate that such approaches would produce the observed outputs.

3. Incomplete Demographic Data

With 72.4% of respondents not providing age data and similar gaps for other demographics, detailed subgroup analyses are not possible. Cannot determine if theme prevalence varies systematically by demographics.

4. Sampling for Detailed Analysis

Some analyses used samples of 500 responses per question rather than complete datasets due to processing constraints. However, 500-response samples provide robust estimates (margin of error $\pm 4.4\%$ at 95% confidence), and all major frequency claims were verified against complete datasets.

5. No Access to Validation Data

Cannot verify the government's claim of 'human reviewer validation' because no validation data, inter-coder reliability metrics, or reviewer qualifications were provided. My assessment of accuracy is based solely on comparing outputs to inputs.

5.9 Reproducibility

All analyses in this paper are reproducible given access to the raw consultation data. The methodology has been described in sufficient detail that another researcher could replicate the frequency counts, sentiment analysis, and thematic coding. Python code used for analysis is available upon request.

In contrast, the government's analysis cannot be fully reproduced because the AI prompts are not disclosed. This represents a significant transparency gap in the government's approach.

Section 6: Visual Representation of Key Themes

This section presents a visual representation of the most frequent terms across all consultation responses. Due to technical constraints, a formal word cloud visualization could not be generated. However, the following table presents the top 50 most frequent substantive terms (excluding common stop words) across all responses:

Term	Frequency	Term	Frequency
data	15,358	public	13,006
research	11,057	people	10,081
government	9,442	need	8,610
their	8,445	companies	7,974
more	7,453	infrastructure	6,596
industry	6,591	there	6,477
into	6,018	access	6,002
like	5,989	work	5,774
technology	5,592	make	5,517
support	5,345	talent	5,203
needs	4,927	programs	4,910
about	4,782	funding	4,746
trust	4,745	national	4,736
them	4,735	ethical	4,709
training	4,658	systems	4,567
education	4,557	sector	4,371
through	4,356	while	4,276
adoption	4,248	human	4,225
health	4,191	canadians	4,165
compute	4,138	innovation	4,096
tech	4,075	ensure	4,066
just	4,032	also	4,020
build	4,020	security	3,980
other	3,967	only	3,966
models	3,960	without	3,903

This frequency analysis reveals the most prominent concepts across all responses. The dominance of terms like 'data', 'research', 'technology', 'education', 'systems', and 'security' aligns with the top themes identified in Section 2. Notably, policy-process terms ('government', 'funding', 'support', 'development') are highly frequent, indicating substantial focus on implementation mechanisms rather than pure technical discussions.

The presence of 'regulation', 'privacy', 'ethical', 'transparency', and 'accountability' among top terms confirms the prominence of governance and trust themes.

Meanwhile, 'innovation', 'business', 'companies', and 'industry' reflect significant commercial focus.

Interestingly, terms related to specific AI technologies or applications are less prominent than governance and implementation terms, suggesting that public input was more focused on 'how AI should be governed' rather than 'what AI should do.'

Conclusions and Recommendations

This independent analysis of Canada's AI Strategy public consultation reveals significant discrepancies between the government's characterizations and the empirical evidence in the raw data. While the government's report covers all consultation themes and reflects genuine public concerns, systematic biases in the analytical approach have led to misrepresentation of theme frequencies and sentiment distribution.

Key Findings Summary

- 1. Limited Demographic Verification:** Only 27.6% of respondents provided sufficient demographic data to verify claims of 'comprehensive representation.'
- 2. Theme Frequency Distortions:** Rare themes (<2% frequency) characterized as 'widespread' or prominent, while common themes sometimes underemphasized.
- 3. Sentiment Imbalance:** Report emphasizes concerns despite 61.9% neutral and 18.5% positive responses.
- 4. Probable Prompting Biases:** Reverse-engineering suggests risk-focused framing, minority view amplification, and policy-priority alignment bias in AI-assisted analysis.

Implications for AI-Assisted Governance

This case study demonstrates both the promise and peril of using AI in public consultation analysis. The government achieved remarkable speed, processing 11,383 responses in months rather than years. However, speed without adequate quality controls creates systematic errors that may misrepresent public input on critical policy decisions.

The biases identified are not random errors but systematic patterns suggesting that the AI was prompted in ways that aligned outputs with existing policy priorities. This raises fundamental questions about the purpose of public consultation: Is it to genuinely learn what the public thinks, or to validate predetermined policy directions?

Recommendations

For Future Public Consultations:

- 1. Publish all AI prompts used in analysis to enable reproducibility and transparency

- 2. Report inter-coder reliability metrics and validation procedures
- 3. Explicitly separate theme frequency from theme intensity in reporting
- 4. Flag any themes emphasized in reports that appear in <5% of responses
- 5. Report sentiment distribution quantitatively, not just qualitatively
- 6. Require independent validation of AI-assisted analyses on critical policy consultations

For AI Strategy Development:

- 1. Acknowledge demographic limitations in consultation representativeness
- 2. Weight policy priorities based on actual public input frequencies, not AI-amplified characterizations
- 3. Consider follow-up consultation specifically targeting underrepresented demographics and regions

Final Remarks

The Government of Canada's AI Strategy consultation collected an impressive 11,383 responses and represents a genuine effort at democratic engagement. The use of AI to process this volume of input is innovative and potentially valuable. However, this analysis demonstrates that AI-assisted analysis requires even more rigorous validation and transparency than traditional methods, not less.

The discrepancies identified here are not merely technical errors—they represent systematic biases that could misdirect policy development. Canada's AI strategy should be built on accurate understanding of public input, not on AI-amplified interpretations that may reflect analyst biases more than citizen voices.

As Canada develops policies for responsible AI use in government, this very consultation process provides a cautionary example: AI can accelerate analysis, but without proper safeguards, it can also accelerate misrepresentation. Transparency, validation, and quantitative reporting are essential to ensure AI serves democratic decision-making rather than undermining it.