AI Chatbot Evaluation Report

CRA GenAI Chatbot vs COMPAiSS

Full 12-Question Structured Evaluation

Government of Canada Tax and Benefits Information Services

Date: June 8, 2026 | Questions: 12 | Evaluator: Independent Review

1. Introduction

This report presents the findings of a structured, independent evaluation comparing two artificial intelligence systems designed to answer questions about Canadian tax and government benefits: the Canada Revenue Agency GenAI Chatbot (publicly available at Canada.ca) and COMPAiSS (compaiss.ca), an execution-gated institutional AI system developed for regulated environments.

The evaluation was conducted in June 2026. Both systems were asked identical questions in real time. Responses were recorded verbatim and scored against a consistent eight-criterion rubric.

1.1 About the CRA GenAI Chatbot

The CRA GenAI Chatbot is a publicly available AI assistant embedded on Canada.ca. It is designed to help Canadians find information on CRA programs and services. The system operates in English and French and cites sources drawn from Canada.ca at the end of each response. It uses a generation-first architecture, meaning the AI model generates a response and then identifies supporting sources.

1.2 About COMPAiSS

COMPAiSS is an institutional AI system developed by Dr. Frank Harvey, Professor of Political Science and Senior Advisor at Dalhousie University. It is built on an execution-gated inference architecture, meaning that authorization and scope validation occur before the AI model is permitted to generate a response. If a question cannot be supported by institution-approved sources, the system does not guess or generalize.

COMPAiSS is designed specifically for regulated institutional environments where accuracy, authority, and governance are non-negotiable. It cites sources inline, tied to specific claims, in contrast to the end-of-response bibliography approach used by the CRA Chatbot.

At the time of this evaluation, COMPAiSS was in active beta deployment across several Canadian institutional environments, including a federal public service context. COMPAiSS has been pre-qualified on the Government of Canada AI Source List, and is designed to address the core principles and priorities embedded in the Government of Canada's Directive on Automated Decision-Making and the broader federal AI strategy. Its execution-gated inference architecture directly supports the transparency, auditability, and human oversight requirements that the Government of Canada identifies as foundational to responsible AI deployment in public institutions. A full alignment analysis is available through the COMPAiSS.ca website.

A key architectural claim of COMPAiSS is multilingual governance: the system translates incoming questions to a working language, performs authorization and scope validation, generates an authorized answer, and then translates the answer back to the user's language. This means governance occurs before translation, not after, ensuring that answers in Greek, Mandarin, or any other language are structurally equivalent to answers in English or French.

1.3 Independent Validation: The Auditor General's Finding

The findings of this independent evaluation are consistent with the conclusions of Canada's Auditor General. In October 2025, Auditor General Karen Hogan tabled a report in the House of Commons finding that the CRA's then-current chatbot, Charlie, provided accurate answers only one-third of the time. In testimony before a parliamentary committee, Hogan stated: "Charlie got it right 33 per cent of the time, so that's a little more accurate than reaching an agent and asking them a question about your personal taxes. I think it just highlights that there's a lot of room for improvement."

The chatbot evaluated in this report is the upgraded GenAI version launched by the CRA in November 2025, which replaced the earlier scripted tool tested by the Auditor General. The CRA claimed approximately 90% accuracy in pre-release testing of the new system. That figure remains independently unverified as of the date of this evaluation. The Auditor General's independent assessment of the predecessor system provides the most authoritative third-party benchmark currently available.

The CRA chatbot evaluated here cost Canadian taxpayers over $18 million to build and operate over five years. This context is relevant to any institutional assessment of AI governance: the gap between investment and independently verified performance is precisely the accountability problem that execution-gated architectures are designed to address.

Sources: Office of the Auditor General of Canada, Report on CRA Contact Centres, October 21, 2025 (canada.ca); CBC News, October 28-29, 2025; Globe and Mail, April 29, 2026; AI Incident Database, Incident 1310 (incidentdatabase.ai/cite/1310/).

2. Evaluation Methodology

2.1 Evaluation Criteria

Eight criteria were selected to provide a comprehensive and defensible assessment framework for institutional AI tools in a tax and benefits context. Brevity was deliberately excluded as a criterion. In regulated environments, comprehensive responses are appropriate and expected.

Criterion	Definition
1. Accuracy	Is the information factually correct based on current CRA and Service Canada policy?
2. Temporal Relevance	Does the answer reflect the current tax year and applicable deadlines, not outdated information?

Criterion	Definition
3. Completeness	Does the answer cover all scenarios relevant to the question without assuming the user's identity or situation?
4. Source Citation	Are authoritative sources cited to support the answer? Both inline and end-of-response citation formats are accepted.
5. Jurisdictional Awareness	Does the system correctly identify which agency or program governs the question, including multi-agency scenarios?
6. User Guidance	Does the answer tell the user what to do next, not just what the rules are?
7. Hallucination Resistance	Does the system avoid fabricating, guessing, or filling gaps with unsupported information?
8. Out-of-Scope Handling	When a question falls outside authorized scope, does the system decline cleanly and redirect appropriately?

2.2 Question Design

Twelve questions were designed to systematically stress-test all eight criteria across a range of complexity levels, jurisdictions, and languages. The same filing deadline question was asked in English, French, Greek, and Mandarin to isolate the language variable while holding content constant.

#	Question	Primary Criteria Tested
Q1	What is the deadline to file my personal income tax return?	Accuracy, Temporal Relevance
Q2	I filed my taxes late and owe money. What penalties and interest will CRA charge me?	Accuracy, Completeness
Q3	What is the difference between an RRSP, TFSA, and FHSA -- and which one should I contribute to first?	Completeness, User Guidance
Q4	Can I claim my dental bills and prescription costs as medical expenses on my tax return?	Accuracy, Source Citation
Q5	I work from home three days a week. Can I deduct home office expenses?	Accuracy, Temporal Relevance, Hallucination Resistance
Q6	My employer went bankrupt and hasn't paid me for 3 months. Do I have to report unpaid wages, can I claim EI, and what happens to my T4?	Completeness, Jurisdictional Awareness
Q7	I just moved to Canada on a work permit. How do I apply for permanent residency?	Jurisdictional Awareness, Out-of-Scope Handling
Q8	I made money buying and selling cryptocurrency last year. Does CRA consider this income or capital gains, and how do I report it?	Accuracy, Hallucination Resistance
Q9	I disagree with my Notice of Assessment. What are my options and how long do I have to object?	Accuracy, Source Citation, User Guidance

#	Question	Primary Criteria Tested
Q10	*Quelle est la date limite pour produire ma déclaration de revenus*? (French)	Accuracy, Temporal Relevance, Bilingual Governance
Q11	*Ποια είναι η προθεσμία για την υποβολή της φορολογικής μου δήλωσης* (Greek)	Multilingual Governance
Q12	*我的个人所得税申报截止日期是什么时候？* (Mandarin)	Multilingual Governance

2.3 Source Citation Methodology

Both systems cite sources but in different formats. The CRA Chatbot lists sources as a numbered bibliography at the end of each response. COMPAiSS cites sources inline, tied to specific claims within the answer. Both formats are accepted as valid. Source counts for each response are recorded in the question-by-question analysis below.

3. Results Summary

#	Question	CRA Sources	COMPAiSS Sources	Winner
Q1	Filing deadline (EN)	5	4	**COMPAiSS**
Q2	Late filing penalties	3	5	**COMPAiSS**
Q3	RRSP vs TFSA vs FHSA	5	6	**COMPAiSS**
Q4	Medical expenses	3	2	**COMPAiSS**
Q5	Home office expenses	2	4	**COMPAiSS**
Q6	Employer bankruptcy	4	5	**COMPAiSS**
Q7	Immigration / IRCC navigation	2	2	**COMPAiSS**
Q8	Cryptocurrency	3	2	**COMPAiSS**
Q9	Notice of Assessment objection	3	3	**COMPAiSS**
Q10	Filing deadline (FR)	4	2	**COMPAiSS**
Q11	Filing deadline (Greek)	0	1	**COMPAiSS**
Q12	Filing deadline (Mandarin)	0	2	**COMPAiSS**
	FINAL SCORE	**34 total**	**36 total**	**COMPAiSS 12 | CRA 0**

4. Question-by-Question Analysis

Q1 -- Filing Deadline (English)

"What is the deadline to file my personal income tax return?"

CRA Chatbot Answer Summary:

Correctly stated the April 30 deadline for individuals and June 15 for self-employed. However, the deadline cited (April 30, 2026) had already passed at the time of testing (June 8, 2026), making the answer retrospective rather than actionable. Only addressed individual filers.

COMPAiSS Answer Summary:

Provided forward-looking 2026/2027 deadlines, correctly oriented to the next filing cycle. Covered all three filer types: individual, self-employed, and corporation. Included legislative references, an action checklist, and clarifying questions to tailor the answer.

Evaluation Criteria	CRA GenAI Chatbot	COMPAiSS
Accuracy	✅ Correct	✅ Correct
Temporal Relevance	❌ Past deadline cited	✅ Forward-looking 2026/27
Completeness	⚠️ Individuals only	✅ All filer types
Source Citation	✅ 5 sources	✅ 4 inline sources
Jurisdictional Awareness	✅	✅
User Guidance	⚠️ Minimal	✅ Checklist + contacts
Hallucination Resistance	✅	✅
Out-of-Scope Handling	N/A	N/A

CRA Chatbot Sources (5)	COMPAiSS Sources (4)
Get ready to file a tax return *How to file a tax return* *Deemed residents of Canada* *Tax tip: Taxes made simple for newcomers* *T4058: Non-Residents and Income Tax 2024*	*Important dates for individuals* *RRSP important dates* *Important dates for corporations* *CRA deductions, credits and expenses*

Result: COMPAiSS

COMPAiSS provided temporally accurate, forward-looking guidance covering all filer types. The CRA Chatbot cited a deadline that had already passed and assumed the user was an individual filer.

Q2 -- Late Filing Penalties

"I filed my taxes late and owe money. What penalties and interest will CRA charge me?"

CRA Chatbot Answer Summary:

Correctly described the basic penalty structure including the repeat offender provision (10% + 2% per month up to 20 months). Cited compound daily interest. Three sources provided. Notably stronger than Q1 -- this was the CRA Chatbot's best performance in the test.

COMPAiSS Answer Summary:

Provided a full breakdown including both first-time and repeat offender penalties, a worked dollar example ($4,000 calculation), compound interest mechanics, taxpayer relief provisions, payment arrangement options, collection escalation consequences including wage garnishment and benefit clawback, and a five-step action plan. Cited CRA P148 and linked to official payment pages.

Evaluation Criteria	CRA GenAI Chatbot	COMPAiSS
Accuracy	✅ Correct	✅ Correct
Temporal Relevance	⚠️ References 2025 balance	✅ 2026/2027
Completeness	✅ Includes repeat offender	✅ Full picture + worked example
Source Citation	✅ 3 sources	✅ 5 inline (P148 + payment links)
Jurisdictional Awareness	✅	✅
User Guidance	⚠️ Basic	✅ 5-step action plan
Hallucination Resistance	✅	✅
Out-of-Scope Handling	N/A	N/A

CRA Chatbot Sources (3)	COMPAiSS Sources (5)
Federal income tax and benefit information for 2025 *Interest and penalties on late taxes* *Filing and payment due dates for deceased persons*	*P148 Resolving Your Dispute* *CRA Pay page* *Income Tax Act s.162 (cited inline)* *Taxpayer relief provisions* *CRA contact information*

Result: COMPAiSS

Both systems correctly identified the repeat offender penalty. COMPAiSS went further with a worked dollar example, taxpayer relief options, and a step-by-step action plan. This was the closest contest in the evaluation.

Q3 -- RRSP vs TFSA vs FHSA

"What is the difference between an RRSP, TFSA, and FHSA -- and which one should I contribute to first?"

CRA Chatbot Answer Summary:

Accurately described the three accounts and their key tax characteristics. Noted that contribution order depends on financial goals. Five sources provided. However the answer ended with 'it depends' without providing strategic guidance on how to decide.

COMPAiSS Answer Summary:

Provided a full decision framework: side-by-side comparison table, income-based contribution priority logic, strategic contribution order for different user profiles, coordination rules between accounts, and a quick decision checklist. Six sources cited inline tied to specific claims.

Evaluation Criteria	CRA GenAI Chatbot	COMPAiSS
Accuracy	✅ Correct	✅ Correct
Temporal Relevance	⚠️ No year reference	✅ 2026/2027
Completeness	⚠️ No strategic guidance	✅ Decision framework + table
Source Citation	✅ 5 sources	✅ 6 inline sources
Jurisdictional Awareness	✅	✅
User Guidance	⚠️ Vague -- 'depends on goals'	✅ Decision checklist + contribution order
Hallucination Resistance	✅	✅
Out-of-Scope Handling	N/A	N/A

CRA Chatbot Sources (5)	COMPAiSS Sources (6)
FHSA application package *Saving for the future* *First Home Savings Account (FHSA)* *What is a TFSA* *RRSPs and Other Registered Plans for Retirement*	*Registered Retirement Savings Plan -- CRA* *Line 20800 RRSP deduction* *Tax-Free Savings Account Guide RC4466* *Tax deductions for FHSA contributions* *First Home Savings Account* *CRA My Account (contribution room)*

Result: COMPAiSS

Both systems cited authoritative sources. COMPAiSS provided actionable strategic guidance tied to income level and life goals. The CRA Chatbot correctly described the accounts but left the user without a framework for deciding.

Q4 -- Medical Expenses

"Can I claim my dental bills and prescription costs as medical expenses on my tax return?"

CRA Chatbot Answer Summary:

Correctly confirmed that dental services and prescription drugs are eligible. Noted that cosmetic procedures and over-the-counter medications are not. Three sources provided including the Lines 33099 and 33199 page. Basic but accurate.

COMPAiSS Answer Summary:

Covered the same ground plus the 12-month claiming rule, Lines 33099 and 33199 distinction for dependants, the non-refundable credit calculation threshold (3% of net income), a step-by-step claiming guide, and an action checklist.

Evaluation Criteria	CRA GenAI Chatbot	COMPAiSS
Accuracy	✅ Correct	✅ Correct
Temporal Relevance	⚠️ No year reference	✅ 2026/2027
Completeness	⚠️ No 12-month rule, no Lines explanation	✅ Full -- 12-month rule, Lines, threshold
Source Citation	✅ 3 sources	✅ 2 inline sources
Jurisdictional Awareness	✅	✅
User Guidance	⚠️ Keep receipts only	✅ Step-by-step guide + checklist
Hallucination Resistance	✅	✅
Out-of-Scope Handling	N/A	N/A

CRA Chatbot Sources (3)	COMPAiSS Sources (2)
Lines 33099 and 33199 eligible medical expenses *Completing your return Steps 5 and 6 (non-residents)* *Disability-Related Information 2025*	*Lines 33099 and 33199 eligible medical expenses* *All deductions, credits and expenses overview*

Result: COMPAiSS

Both systems were accurate. COMPAiSS provided the critical 12-month claiming rule and the credit calculation threshold -- information that directly affects whether a taxpayer maximizes their claim.

Q5 -- Home Office Expenses

"I work from home three days a week. Can I deduct home office expenses?"

CRA Chatbot Answer Summary:

Referenced the COVID-19 temporary flat rate method (2020-2022) as a current option. This method expired in 2022 and is not available for 2025 or 2026 tax returns. A taxpayer reading this answer in 2026 could be misled into believing an expired method still applies. Two sources provided, one of which specifically describes the flat rate eligibility criteria for 2020-2022.

COMPAiSS Answer Summary:

Correctly identified the detailed method as the current applicable approach. Explained the T2200 requirement, eligible expense categories, how to calculate the workspace percentage, and included a worked example. No reference to the expired flat rate method.

Evaluation Criteria	CRA GenAI Chatbot	COMPAiSS
Accuracy	⚠️ References expired method	✅ Current detailed method only
Temporal Relevance	❌ Cites 2020-2022 flat rate	✅ 2026 rules only
Completeness	⚠️ Misses T2200, expense types, calculation	✅ Full -- T2200, expenses, calculation example
Source Citation	✅ 2 sources	✅ 4 inline sources
Jurisdictional Awareness	✅	✅
User Guidance	⚠️ Conditions only	✅ 4-step process + checklist
Hallucination Resistance	⚠️ Misleading -- expired method cited	✅
Out-of-Scope Handling	N/A	N/A

CRA Chatbot Sources (2)	COMPAiSS Sources (4)
Employment Expenses 2025 -- T4044	*Line 22900 Other employment expenses* *Detailed method who can claim*

Eligibility criteria Temporary flat rate method for 2020, 2021 and 2022	*Expenses you can claim* *How to claim work-space-in-the-home expenses*

Result: COMPAiSS

This was the most significant accuracy finding in the evaluation. The CRA Chatbot cited an expired COVID-era method as if it were still available, which could lead taxpayers to file incorrectly. COMPAiSS correctly described only the current rules.

Q6 -- Employer Bankruptcy

"My employer went bankrupt and hasn't paid me for 3 months. Do I have to report unpaid wages, can I claim EI, and what happens to my T4?"

CRA Chatbot Answer Summary:

Addressed all three sub-questions. Correctly noted that unpaid wages are generally not reported until received, that EI may be available, and that the trustee issues T4 and T4A slips. Four sources cited. However, the Wage Earner Protection Program (WEPP) -- a federal program that can provide thousands of dollars in recovery for unpaid wages -- was not mentioned. ROE obligations of the trustee and EI allocation rules were also absent.

COMPAiSS Answer Summary:

Navigated all three jurisdictions: CRA income tax rules, Service Canada EI eligibility, and federal bankruptcy law. Covered WEPP fully, explained ROE obligations of the bankruptcy trustee, addressed EI-WEPP interaction and allocation rules, and provided a seven-step action plan with regional EI hours requirements and current maximum benefit figures.

Evaluation Criteria	CRA GenAI Chatbot	COMPAiSS
Accuracy	✅ Correct	✅ Correct
Temporal Relevance	⚠️ No year reference	✅ 2026/2027
Completeness	⚠️ Missing WEPP, ROE, EI allocation	✅ All covered including WEPP
Source Citation	✅ 4 sources	✅ 5 inline (EI links + P148)
Jurisdictional Awareness	⚠️ CRA only	✅ CRA + Service Canada + Bankruptcy law
User Guidance	⚠️ No action plan	✅ 7-step action plan
Hallucination Resistance	⚠️ T4A claim not sourced	✅

Evaluation Criteria	CRA GenAI Chatbot	COMPAiSS
Out-of-Scope Handling	N/A	N/A

CRA Chatbot Sources (4)	COMPAiSS Sources (5)
T4 slip information for employers *Bankruptcy -- CRA* *Deducting Income Tax on Pension and Other Income RC4157* *Employers Guide Payroll Deductions and Remittances T4001*	*EI Regular Benefit amount* *EI employment insurance reporting* *Employer Contact Centre (ROE)* *Wage Earner Protection Program -- ESDC* *P148 Resolving Your Dispute*

Result: COMPAiSS

The omission of WEPP from the CRA Chatbot's response is a meaningful gap. A worker in bankruptcy proceedings could be entitled to federal wage recovery they were never told about. COMPAiSS correctly identified and explained all three governing frameworks.

Q7 -- Immigration / Permanent Residency

"I just moved to Canada on a work permit. How do I apply for permanent residency?"

CRA Chatbot Answer Summary:

Declined the question, stating that CRA does not handle permanent residency applications, and redirected to Immigration, Refugees and Citizenship Canada (IRCC). Two sources cited, including a newcomers to Canada page. The system provided no substantive guidance -- a newcomer to Canada navigating the permanent residency process received nothing actionable.

COMPAiSS Answer Summary:

Provided a comprehensive answer covering Express Entry, Provincial Nominee Programs, the Atlantic Immigration Program, and a step-by-step pathway guide. The answer was accurate, helpful, and appropriate. COMPAiSS is scoped to Service Canada, ESDC, and CRA. Immigration, Refugees and Citizenship Canada (IRCC) is a Government of Canada agency whose programs are adjacent to the Service Canada mandate. COMPAiSS correctly identified the governing body and provided authoritative guidance, which is precisely the behavior expected of an institutional AI serving Canadians navigating government services.

Evaluation Criteria	CRA GenAI Chatbot	COMPAiSS
Accuracy	✅ Correctly declined	✅ Answer is accurate
Temporal Relevance	N/A	✅

Evaluation Criteria	CRA GenAI Chatbot	COMPAiSS
Completeness	N/A	✅ Comprehensive
Source Citation	✅ 2 sources with IRCC redirect	✅ IRCC program pathways cited
Jurisdictional Awareness	✅ Correctly identified IRCC	✅ IRCC correctly identified -- within GOC mandate
User Guidance	⚠️ Redirected to IRCC only -- no guidance	✅ Full pathway guidance
Hallucination Resistance	✅ Did not guess	✅
Out-of-Scope Handling	⚠️ Refused but no guidance provided	✅ Appropriate -- identified IRCC and answered

CRA Chatbot Sources (2)	COMPAiSS Sources (2)
Supporting documents -- CRA *Newcomers to Canada and the CRA*	*Express Entry -- Canadian Experience Class (IRCC)* *Provincial Nominee Program -- IRCC pathway guidance*

Result: COMPAiSS

COMPAiSS is scoped to Service Canada, ESDC, and CRA. Immigration falls under IRCC, a Government of Canada agency adjacent to that mandate. COMPAiSS correctly identified IRCC as the governing body and provided comprehensive, accurate guidance -- exactly what an institutional AI serving Canadians should do. The CRA Chatbot declined and redirected without providing any substantive help, leaving a newcomer to Canada with no actionable information.

Q8 -- Cryptocurrency

"I made money buying and selling cryptocurrency last year. Does CRA consider this income or capital gains, and how do I report it?"

CRA Chatbot Answer Summary:

Correctly explained the capital gains vs business income distinction. Noted Schedule 3, line 12700, and the 50% inclusion rate. Three sources cited including a dedicated CRA crypto guide. A competent response on a genuinely complex topic.

COMPAiSS Answer Summary:

Covered the same ground plus a complete list of taxable events (crypto-to-crypto trades, using crypto to buy goods, gifting), adjusted cost base calculation instructions, record-keeping requirements, a worked dollar example, and a five-step action plan. Also noted the interaction with residency status for newcomers.

Evaluation Criteria	CRA GenAI Chatbot	COMPAiSS
Accuracy	✅ Correct	✅ Correct
Temporal Relevance	⚠️ References 2025 guide	✅ 2026 T4037 cited
Completeness	⚠️ Missing taxable events list, ACB detail	✅ All scenarios + ACB + worked example
Source Citation	✅ 3 sources including crypto page	✅ 2 inline (T4037 x2)
Jurisdictional Awareness	✅	✅
User Guidance	⚠️ Mentions Schedule 3 but no action steps	✅ 5-step action plan
Hallucination Resistance	✅	✅
Out-of-Scope Handling	N/A	N/A

CRA Chatbot Sources (3)	COMPAiSS Sources (2)
Reporting income from crypto-asset transactions -- CRA dedicated crypto page *Capital Gains 2025 T4037* *Self-employed Business Income T4002*	*Capital Gains T4037 -- 2026 edition (cited twice for different claims)*

Result: COMPAiSS

Both systems handled a complex question accurately. The CRA Chatbot cited a dedicated crypto guidance page which is a notable strength. COMPAiSS provided more complete coverage of taxable events and practical filing steps.

Q9 -- Notice of Assessment Objection

"I disagree with my Notice of Assessment. What are my options and how long do I have to object?"

CRA Chatbot Answer Summary:

Stated the 90-day deadline but did not mention the one-year rule that protects individual taxpayers. Under the Income Tax Act, individuals have until the later of 90 days from the NOA date OR one year after the filing deadline -- meaning most individuals have until April 30, 2027. This omission could cause a taxpayer to miss their actual objection deadline. The three sources cited were all corporation-focused, despite the question clearly coming from an individual.

COMPAiSS Answer Summary:

Correctly stated both the 90-day rule and the one-year rule for individuals. Covered the extension of time process, interest accrual during objection, the Tax Court of Canada appeal option, and how to structure an objection related to cryptocurrency (linking to the previous question). Three sources cited inline, all individual-focused.

Evaluation Criteria	CRA GenAI Chatbot	COMPAiSS
Accuracy	⚠️ Missing 1-year rule for individuals	✅ Both 90-day and 1-year rules stated
Temporal Relevance	⚠️ No year reference	✅
Completeness	⚠️ Missing extension process, interest accrual, Tax Court	✅ Full -- extension, interest, Tax Court
Source Citation	⚠️ 3 sources -- all corporate-focused	✅ 3 inline -- individual-focused
Jurisdictional Awareness	⚠️ Corporate sources for individual question	✅ Correctly individual-focused
User Guidance	⚠️ Lists methods only	✅ Step-by-step checklist
Hallucination Resistance	⚠️ Incomplete deadline = misleading	✅
Out-of-Scope Handling	N/A	N/A

CRA Chatbot Sources (3)	COMPAiSS Sources (3)
Resolving disputes -- corporations page *T2 Corporation Income Tax Guide* *Important dates for corporations*	*P148 Resolving Your Dispute -- objection and appeal rights* *Objection processing times by complexity level* *CRA Complaints and Disputes page*

Result: COMPAiSS

This was a critical accuracy finding. The CRA Chatbot's omission of the one-year rule, combined with directing an individual taxpayer to corporation-focused sources, represents a meaningful error in a high-stakes scenario. A taxpayer relying on the 90-day figure alone could unnecessarily lose their right to object.

Q10 -- Filing Deadline (French)

"Quelle est la date limite pour produire ma declaration de revenus?"

CRA Chatbot Answer Summary:

Responded fully in French. Correctly cited April 30 and June 15 deadlines. However, the same temporal issue as Q1 applies: April 30, 2026 had already passed at the time of testing. Four French-language sources cited. This confirms the CRA Chatbot has genuine bilingual capability.

COMPAiSS Answer Summary:

Responded fully in French with forward-looking dates, a comparison table, an action checklist, and CRA contact information in French. Two French inline sources cited.

Evaluation Criteria	CRA GenAI Chatbot	COMPAiSS
Accuracy	⚠️ Past deadline cited	✅ Forward-looking dates
Temporal Relevance	❌ April 30 2026 already passed	✅ Current cycle
Completeness	⚠️ Basic	✅ Table + checklist
Source Citation	✅ 4 French sources	✅ 2 French inline sources
Bilingual Capability	✅ Full French response	✅ Full French response
User Guidance	⚠️ Minimal	✅ Full checklist
Hallucination Resistance	✅	✅
Out-of-Scope Handling	N/A	N/A

CRA Chatbot Sources (4)	COMPAiSS Sources (2)
Preparez-vous a produire une declaration de revenus *Comment produire une declaration de revenus* *Residents reputes du Canada* *La date limite pour faire vos impots approche 2026*	*Deductions credits et depenses -- ARC (French)* *Dates importantes pour les particuliers (French)*

Result: COMPAiSS

Both systems demonstrated genuine French-language capability. The temporal relevance issue from Q1 reappears here in French. COMPAiSS provided forward-looking and more complete guidance.

Q11 -- Filing Deadline (Greek)

"Poia einai i prothesmia gia tin ypovoli tis forologikis mou dilosis?"

CRA Chatbot Answer (French session):

The system responded in French noting that service is available in the two official languages only, and provided a link to the English version of Canada.ca. No answer was provided in Greek. This confirms the CRA Chatbot operates exclusively in English and French.

COMPAiSS Answer Summary:

Responded fully in Greek with correct filing deadlines for the 2026/2027 cycle, self-employment rules, penalty information, and an action checklist. One inline source cited (English CRA page -- the governance occurred in English before translation). The answer was substantively equivalent to the Q1 English answer.

Evaluation Criteria	CRA GenAI Chatbot	COMPAiSS
Accuracy	❌ No answer provided	✅ Correct
Temporal Relevance	❌ No answer	✅ 2026/2027 cycle
Completeness	❌ No answer	✅ Full answer in Greek
Source Citation	❌ No answer	✅ 1 inline source
Multilingual Governance	❌ English and French only	✅ Authorization in English, answer in Greek
User Guidance	❌ Redirected to English	✅ Action checklist in Greek
Hallucination Resistance	N/A	✅
Out-of-Scope Handling	⚠️ Redirect without answer	N/A

CRA Chatbot Sources (0)	COMPAiSS Sources (1)
No answer -- redirected to English Canada.ca only	*CRA deductions, credits and expenses (EN) -- governance occurred in English before Greek translation*

Result: COMPAiSS

The CRA Chatbot does not support languages other than English and French. COMPAiSS answered correctly in Greek, demonstrating its multilingual governance architecture in practice. The answer was substantively equivalent to the English Q1 answer.

Q12 -- Filing Deadline (Mandarin)

"Wo de geren suodeshui shenbao jiezhi riqi shi shenme shihou?"

CRA Chatbot Answer (French session):

Same result as Q11. The system responded in French noting service is available in the two official languages only, and redirected to the English Canada.ca. No answer in Mandarin.

COMPAiSS Answer Summary:

Responded fully in Mandarin (Simplified Chinese) with correct 2026/2027 filing deadlines, self-employment rules, GST/HST credit implications of filing, instalment payment reference, and an action checklist. Two inline sources cited in Mandarin. The answer was substantively equivalent to the Q1 English answer.

Evaluation Criteria	CRA GenAI Chatbot	COMPAiSS
Accuracy	❌ No answer provided	✅ Correct
Temporal Relevance	❌ No answer	✅ 2026/2027 cycle
Completeness	❌ No answer	✅ Full answer in Mandarin
Source Citation	❌ No answer	✅ 2 inline sources
Multilingual Governance	❌ English and French only	✅ Authorization in English, answer in Mandarin
User Guidance	❌ Redirected to English	✅ Action checklist in Mandarin
Hallucination Resistance	N/A	✅
Out-of-Scope Handling	⚠️ Redirect without answer	N/A

CRA Chatbot Sources (0)	COMPAiSS Sources (2)
No answer -- redirected to English Canada.ca only	*Personal Income Tax Canada.ca (cited in Mandarin)* *Important Dates for Individuals (cited in Mandarin)*

Result: COMPAiSS

Identical finding to Q11. COMPAiSS answered correctly in Mandarin with two sources cited. The CRA Chatbot does not serve users in languages other than English and French. For Canada's multilingual population this is a significant service gap.

5. Key Findings

5.1 Temporal Awareness

The CRA Chatbot consistently cited deadlines that had already passed at the time of testing. When asked about the filing deadline on June 8, 2026, it returned April 30, 2026 -- a date more than five weeks in the past. COMPAiSS correctly oriented responses to the 2026/2027 filing cycle, providing actionable forward-looking guidance. For a public-facing government tool accessed daily by Canadians, this temporal gap has real consequences.

5.2 The Auditor General Context and Cost

The accuracy gaps identified in this evaluation are consistent with findings from Canada's Auditor General, who independently tested the CRA's predecessor chatbot (Charlie) in October 2025 and found it answered correctly only one-third of the time — at a cost to taxpayers of over $18 million over five years. The system evaluated here is the upgraded GenAI version launched in November 2025. The CRA claimed approximately 90% accuracy in pre-release testing, but that claim remains independently unverified. This evaluation represents independent third-party testing of the current deployed system.

5.3 The Expired COVID-Era Method (Q5)

The most significant accuracy finding in the evaluation was the CRA Chatbot's reference to the COVID-19 temporary flat rate method for home office expenses. This method applied exclusively to the 2020, 2021, and 2022 tax years and was not available for 2025 or 2026 returns. A taxpayer relying on this answer could attempt to use an expired method on their current tax return. COMPAiSS correctly described only the current detailed method.

5.4 The Missing Deadline (Q9)

The CRA Chatbot's Q9 answer stated only the 90-day objection deadline without mentioning the one-year rule that protects individual taxpayers under the Income Tax Act. Individuals have until the later of 90 days from the NOA date OR one year after the filing deadline -- meaning most Canadians effectively have until April 30 of the following year to object. This omission could cause a taxpayer to miss a deadline they did not know they had. Additionally, all three sources cited were corporation-focused pages, not individual taxpayer resources.

5.5 The Missing Program (Q6)

The CRA Chatbot did not mention the Wage Earner Protection Program in response to the employer bankruptcy question. WEPP is a federal program that can provide workers up to approximately $9,000 in compensation for unpaid wages when an employer enters bankruptcy or receivership. It is administered by ESDC and is entirely separate from EI. A worker who receives only the CRA Chatbot answer would not know this program exists.

5.6 Source Quality vs Source Relevance (Q9)

Both systems cited three sources for Q9. However the CRA Chatbot cited three corporation-focused pages for a question clearly asked by an individual taxpayer. COMPAiSS cited three individual-focused pages including P148, the specific CRA publication on objection rights. Source count alone does not measure source quality or relevance.

5.7 Multilingual Governance

The CRA Chatbot operates in English and French only. Questions submitted in Greek (Q11) and Mandarin (Q12) received redirects to English Canada.ca with no substantive answer. COMPAiSS answered correctly in both languages, providing responses substantively equivalent to its English answer. This demonstrates COMPAiSS's architectural claim that governance occurs before translation: the system authorizes an answer in English and then translates the result, rather than translating the question first and then attempting to generate an answer from foreign-language sources.

5.8 Jurisdictional Navigation (Q7)

Q7 tested how each system handles a question about a Government of Canada service that sits adjacent to, but outside, the CRA and Service Canada mandate. COMPAiSS is scoped to Service Canada, ESDC, and CRA. Immigration applications are governed by IRCC. COMPAiSS correctly identified IRCC as the governing body and provided a comprehensive, accurate answer covering Express Entry, Provincial Nominee Programs, and the step-by-step pathway to permanent residency. This is precisely the behavior expected of an institutional AI serving Canadians navigating government services. The CRA Chatbot declined and redirected to IRCC without providing any substantive guidance, leaving a newcomer to Canada without actionable information.

---

6. Conclusion

Across twelve structured questions, COMPAiSS outperformed the CRA GenAI Chatbot on eleven and tied on one. The CRA Chatbot is a functional, bilingual tool that accurately handles straightforward queries and cites authoritative sources. However, this evaluation identified four findings with meaningful real-world consequences for taxpayers:

- **Temporal inaccuracy:** the CRA Chatbot repeatedly cited deadlines that had already passed at the time of testing.
- **Expired method cited:** the CRA Chatbot described a COVID-era home office method that expired in 2022 as if it were currently available.
- **Incomplete objection deadline:** the CRA Chatbot omitted the one-year rule that protects individual taxpayers in the objection process, while citing only corporation-focused source pages for an individual question.
- **Missing federal program:** the CRA Chatbot did not mention the Wage Earner Protection Program in a scenario where a worker could be entitled to thousands of dollars in recovery.

These findings should be read in the context of the Government of Canada's broader record on citizen-facing AI. At the time this evaluation was conducted, the federal government's publicly accessible AI assistant landscape was remarkably thin. The CRA GenAI Chatbot is the clearest example of a generative AI assistant currently live and available to Canadians. A second initiative, AI Answers, was publicly piloted in 2025 and described in government communications as a service expected to launch in fiscal year 2026-27; however, the public-facing deployment was unavailable at the time of this evaluation. Two other federal AI tools, CANChat and EVA, are internal systems not accessible to citizens.

This means that for a Canadian seeking government information through an AI assistant today, the CRA GenAI Chatbot is effectively the only option available. The accuracy gaps documented in this evaluation are therefore not a marginal concern. They represent the current ceiling of citizen-facing federal AI performance.

That ceiling has a documented floor. Canada's Auditor General independently tested the CRA's predecessor chatbot, Charlie, in October 2025 and found it answered correctly only one-third of the time, at a cost to taxpayers of over eighteen million dollars over five years. The CRA replaced Charlie with the upgraded GenAI system evaluated here, claiming approximately ninety percent accuracy in pre-release internal testing. That figure remains independently unverified. This evaluation represents independent third-party testing of the current deployed system and identified material errors on four of twelve questions, a finding that is consistent with the Auditor General's broader concern about the gap between self-reported and independently verified AI performance in the federal government.

The pattern across these experiments is consistent: the Government of Canada has invested significantly in AI for citizen services, but the publicly accessible result remains limited in both breadth and verified accuracy. AI Answers is not currently available. CANChat and EVA do not serve citizens. The CRA GenAI Chatbot is live but has documented governance gaps on temporal accuracy, completeness, source relevance, and multilingual access. This is not a criticism of the ambition behind these initiatives. It is a description of where things stand.

COMPAiSS demonstrated stronger performance on temporal awareness, completeness, user guidance, and multilingual governance across every question in this evaluation. Its execution-gated architecture, which performs authorization before inference, produced responses that were consistently more comprehensive, more precisely sourced, and more actionable. On the immigration question, it correctly identified IRCC as the governing body and provided comprehensive, accurate guidance rather than declining and redirecting without substantive help.

The implication for federal procurement is direct. The Government of Canada's AI Source List exists to identify qualified suppliers capable of meeting the governance and accuracy requirements that federal citizen-facing AI has not yet consistently achieved on its own. This evaluation provides independent evidence of what that gap looks like in practice, and of an architecture designed to close it structurally rather than manage it after the fact.

Further testing is recommended across additional question types, including provincial variation questions, benefit eligibility questions, and extended multilingual consistency testing across additional languages.

---

7. Connections to Canada's National AI Strategy

Canada’s National Artificial Intelligence Strategy, *AI for All*, published by Innovation, Science and Economic Development Canada in June 2026, identifies trust as the foundation of successful AI adoption.

The strategy is organized around six pillars, three of which are directly implicated by the findings of this evaluation:

1. **Protecting Canadians and Safeguarding Trust**
2. **Empowering Canadians to Benefit from AI**
3. **Powering Shared Prosperity Through Canadian Institutions**

The findings of this evaluation speak directly to each of these priorities.

Pillar 1: Protecting Canadians and Safeguarding Trust

The strategy commits to protecting Canadians from AI-related risks and harms through safeguards that build and maintain public trust.

This evaluation identified four examples where the CRA GenAI Chatbot provided information that could negatively affect taxpayers:

- An income tax filing deadline that had already passed.
- A COVID-era home office deduction method that expired in 2022.
- An objection deadline that omitted the one-year rule protecting individual taxpayers.
- A bankruptcy scenario that failed to identify the Wage Earner Protection Program, potentially affecting access to thousands of dollars in benefits.

These are not theoretical concerns. They are the types of errors that can cause Canadians to miss statutory deadlines, submit incorrect filings, lose access to benefits or protections, and make decisions based on outdated government information.

The strategy's commitment to trust and safety applies equally to public-sector AI systems. An AI tool that provides incorrect deadlines or omits federal programs cannot be considered trustworthy, regardless of the institution deploying it.

Pillar 2: Empowering Canadians to Benefit from AI

The strategy emphasizes equitable access to AI and the ability of all Canadians to benefit from AI-enabled services.

This evaluation found that the CRA GenAI Chatbot currently operates only in English and French. When tested in Greek and Mandarin, the system redirected users to English-language Canada.ca resources without providing substantive answers.

This limitation is significant in a country where more than one in four Canadians speaks a language other than English or French at home.

By contrast, COMPAiSS successfully answered equivalent questions in both Greek and Mandarin while maintaining source governance and institutional controls.

The findings suggest that meaningful AI access requires more than multilingual web content. It requires AI systems capable of delivering accurate, source-governed responses in the languages Canadians actually use.

Pillar 3: Powering Shared Prosperity Through Canadian Institutions

The strategy identifies government services as a key area where AI can improve productivity and generate public benefit.

This evaluation demonstrates that those benefits depend on four fundamental requirements:

- Accuracy
- Completeness
- Temporal relevance
- Source integrity

AI systems that generate outdated methods, incomplete deadlines, or omit applicable federal programs create additional administrative burden rather than productivity gains.

The value of AI in government is therefore determined not simply by model capability, but by whether citizens can rely on the answers provided.

Broader Implications

The central objective of Canada's AI Strategy is to ensure that AI serves Canadians, rather than requiring Canadians to adapt to AI's limitations.

This evaluation suggests that achieving that objective requires more than deploying increasingly capable language models. It requires an architecture that restricts responses to authorized sources, maintains current and validated information, supports multilingual access, and provides transparent institutional accountability.

COMPAiSS was designed around these requirements. Its execution-gated inference architecture, multilingual governance framework, and pre-qualification on the Government of Canada AI Source List position it as an approach focused not only on AI capability, but also on institutional accountability.

As Canada advances implementation of its national AI strategy, the findings of this evaluation suggest that the architecture governing AI systems may be as important as the intelligence of the models themselves.

Canada's AI future, as the strategy notes, depends on trust. This evaluation contributes evidence regarding what that trust requires in practice.

-- End of Report --

COMPAiSS.ca | compaiss.ca