**AI** Responsible
Artificial Intelligence
Institute

Advancing Trusted AI

June 2022

# The Responsible AI Certification

Program Guidebook

June 2022

Responsible Artificial Intelligence Institute (RAII) - Responsible AI Certification - Program Guidebook

June 2022

# Table of Contents

# 1. About

## 1.1 Introduction

This guidebook contains all information pertaining to RAII's Certification Program such as how it was developed, how it works, and the policies and procedures that must be upheld by those who earn and audit the certification. Each of these, and all other sections were written in accordance with the required scheme development criteria in IAF (2022) and ISO (2019) scheme development documents.

## 1.2 Why Responsible AI Matters

For several years, governments, companies, and civil society organizations have grappled with how to govern AI systems in a consistent manner. As the number of organizations putting forward principles has multiplied, an international consensus has emerged on what constitutes responsible AI - the practice of designing, developing, and deploying AI with good intention to empower employees and businesses, and fairly impact customers and society (Eitel-Poter et al., 2021). These efforts have led to the adoption of principles by organizations around the world. These principles are now informing a burgeoning and expansive set of AI-related standards and regulations. However, as countries begin proposing and enacting various laws and regulations, we return to the issue of consistency as many of these laws and regulations differ in scope, stringency, and severity for non-compliance. This inconsistency problem is what the Responsible Artificial Intelligence Institute's (RAII) work has addressed through its *Responsible AI Certification Program* and various supporting governance tools.

## 1.3 The Responsible AI Institute (RAII)

RAII is an independent and community-driven non-profit organization building a Responsible AI Certification program and various supporting governance tools such as an AI Regulatory Tracker and Global AI Use Case Map to help organization's design, implement, and monitor their AI use and compare various use cases in their industry. RAII's main governance tool, the Responsible AI Certification Program, assesses an organization's AI system against six dimensions: Systems Operations, Explainability and Interpretability, Accountability, Consumer Protection, Bias and Fairness, and Robustness.

RAII is a member of the World Economic Forum's (WEF) Global AI Action Alliance (GAIA), which comprises over 100 government entities, civil society organizations, private companies, and academic institutions dedicated to responsible AI use. RAII drew on insights from the WEF GAIA community, and many other organizations and countless experts across several AI subfields to help shape its Responsible AI Certification Program assessment dimensions and assessment questions within. The following section describes RAII's entire Responsible AI Certification Program development process.

# 2. Certification Program Development

RAII developed its Responsible AI Certification Program through five consecutive steps: (a) extensive review, implementation framework, and assessment creation; (b) assessment testing, adjustments, and working groups; (c) assessment scoring; (d) RAII Council formation; and (e) accreditation.

## 2.1 Extensive Review, Implementation Framework, and Assessment Creation

First, RAII conducted an extensive review of existing and developing AI-related laws, regulations, standards, frameworks, uses and harms, views of notable AI stakeholders across 18 countries and the European Union, and industry-specific guidance documents (BSI, 2020; Committee of Ministers, 2020; FDA, 2021; GPAI, 2020; HLEGAI, 2019; ISO, 2022; Jillson, 2021; OECD, 2019: 2022; Office of the Comptroller of the Currency's, 2021; OSFI, 2020; Smith, 2020; IEEE, 2017; Treasury Board Secretariat of Canada, 2019; UNESCO, 2021; World Economic Forum, 2022). This extensive review helped RAII identify recurring themes, principles, and gaps within various aspects of the extensive review. These recurring themes, principles, and gaps laid the foundation for what became RAII's Implementation Framework - the framework used as a basis for all assessment dimensions and subdimensions and their corresponding questions. The Implementation Framework is based on six recurring AI dimensions derived from the extensive review. The framework's six dimensions are: (i) System Operations; (ii) Explainability and Interpretability; (iii) Accountability; (iv) Consumer Protection; (v) Bias and Fairness; and (vi) Robustness. Each dimension contains a list and description of its subdimensions.

## 2.1.1.  System Operations

The system operations dimension explores the functioning of the AI system and key design choices related to the model and its data. The subdimensions assess five key areas: system scope and function, which examines the system's origin, capabilities, breadth of deployment, and domain; human-in-the-loop, which examines the autonomy level of the system and associated risk; model is fit for purpose, which examines whether the AI system's outcome match its intent; data relevancy and representativeness, which examines the data's composition and use; and data quality, which examines the dataset's creation and quality.

> **System Scope and Function:** The contexts, use cases, and limitations of the AI system.

> **Human-in-the-Loop:** The extent of staff interaction with an AI system's decision-making process.

> **Model is Fit for Purpose:** The sector/industry in which the AI system operates and that sector/industry's associated risk level alongside what the AI system is programmed to do.

> **Data Relevance and Representativeness:** The extent to which an AI system is used within or outside an organization and how many people it affects.

> **Data Quality:** The strength of the AI system's performance and accuracy alongside the types of data it uses.

## 2.1.2.  Explainability and Interpretability

The explainability and interpretability dimension ensures that the AI system's workings and uses can be explained and documented in terms that humans - including users, data subjects, and others - can understand. This involves

inspecting the complexity of the system – like its capabilities, how it was trained - plus any steps taken by the team to bolster the system's explainability (like prioritizing simple models during the design process, implementing integration tests to understand how individual components interact with each other). It also involves analyzing how the system presents information to its users and data subjects: how it communicates the outcome and the reasoning behind that outcome, whether it provides notification that an AI system was involved in the generation of that outcome, and whether it offers and communicates opportunities for redress.

> **Communication about the Outcome:** The extent to which people are appropriately informed about the inputs and outputs of the AI system.

> **Notification:** The processes, if any, that are in place to notify a person when an automated decision has been made about them.

> **Recourse:** The mechanisms available to end users to appeal the AI system's decisions and/or outputs.

> **Understanding the AI System's Decisions or Functions:** The extent to which the organization documents, reviews, and/or publishes additional system information.

## 2.1.3. Accountability

The accountability dimension examines whether the organization has set up clear oversight processes for the development and implementation of the AI system. These oversight processes should ensure that the organization is held accountable for designing a system that is explainable, fair, and not manipulative, as well as for clearly communicating the system's functions and limitations to its users. The accountability dimension also verifies that the AI system development team has documented design choices, reviewed system failures, and conducted an appropriate scenario planning exercise.

> **Organizational Governance:** The organization's documentation requirements for various AI system changes, oversight processes, and implementation methods.

> **Team Governance:** Independent review processes and ongoing monitoring of an AI system throughout its lifecycle.

## 2.1.4.  Consumer Protection

The consumer protection dimension evaluates the risk the AI system poses to individuals and the steps the organization and development team have taken to mitigate these risks. The assessment studies transparency - whether data policies, system risks, testing results, and appropriate uses are communicated to users and data subjects. It also estimates the maximum potential harm of the AI system and checks whether the team has completed appropriate mitigation exercises such as harms mapping and root cause analysis. The assessment is also concerned with privacy, cataloging what sensitive data (like personal data, demographic information, or business data) is used during training and deployment, and what strategies the team has employed to protect that data.

> **Transparency to the User and Data Subject:** The degree to which AI system users are informed that AI is assisting with decisions.

> **Harm to Individuals:** The degree to which the AI system could harm its users.

> **Protections:** The extent to which the AI system protects an individual's or a group's privacy.

## 2.1.5. Bias and Fairness

The bias dimension assesses whether the AI system was designed in a manner that promotes fairness and avoids bias. The extent to which the organization and development team have engaged with bias and fairness issues, such as by conducting research, situating the system in its historical and cultural context, hiring team members with relevant expertise, and providing opportunities for workers displaced by the system, is considered. The assessment also reviews any bias training that the organization has provided to the AI system's users. Finally, the team's testing procedures are analyzed: tests that employ appropriate fairness definitions and that consider multiple types of potential bias should be performed on an ongoing basis.

> **Bias Impacts:** The degree to which the organization has put mitigation processes in place to combat unintended bias and similar issues.

> **Bias Training:** The types of educational resources provided to those designing, developing, using, and impacted by the system.

> **Bias Testing:** The extent to which the interaction among the AI system's components - including models, algorithms, and datasets - has been tested.

## 2.1.6. Robustness

The robustness dimension investigates if the AI system is safe and effective. Its questions ascertain whether the system is adequately protected against data drift, as well as whether it is robust enough to handle edge cases and extreme scenarios. This dimension also checks what testing, like accuracy tests or unit tests, are completed and at what frequency.

> **Data Drift:** The organization's processes and procedures for combatting the degradation of the AI system's performance due to changing data and variable relationships.

> **System Acceptance Test Performed:** The extent to which the AI system has been exposed to and tested across several edge cases.

> **Contingency Planning:** The extent to which the organization is prepared for adversarial attacks, load inputs, and other edge cases and extreme scenarios.

*Figure 1. RAII Implementation Framework*

**1 Systems Operations**

1.1 System Scope and Function
1.2 Human-in-the-Loop
1.3 Model is Fit for Purpose
1.4 Data Relevance and Representativeness
1.5 Data Quality

**2 Explainability & Interpretability**

2.1 Communication About the Outcome
2.2 Notification
2.3 Recourse
2.4 Understanding the AI System's Decisions or Functions

**3 Accountabilty**

3.1 Organizational Governance
3.2 Team Governance

**4 Consumer Protection**

4.1 Transparency to the User and Data Subject
4.2 Harms to Individuals
4.3 Protections

**5 Bias & Fairness**

5.1 Bias Impacts
5.2 Bias Training
5.3 Bias Testing

**6 Robustness**

6.1 Data Drift
6.2 System Acceptance Test is Performed
6.3 Contingency Planning

After RAII completed its extensive review and created its Implementation Framework, RAII created the first assessment version based on the findings from the extensive review and Implementation Framework dimensions and subdimensions. A complete explanation of the assessment can be read in Section 4 below.

## 2.2 Assessment Testing, Adjustments, and Working Groups

After RAII completed the first assessment version, three organizations were used to test the assessment on. Each organization was from a different industry with different AI systems (each organization signed NDAs and have asked to remain anonymous). Testing the assessment across different industries and AI systems helped RAII identify several assessment questions that needed to be edited, removed, moved to a different subdimension, or questions that needed to be added.

Alongside the organizations' testing, RAII convened working groups around each industry represented across the three organizations. These working groups were each co-chaired by leading AI-experts in the working groups' designated industry and engaged other organizations within the industry to examine relevant research and industry knowledge surrounding the AI system(s) being assessed. These working groups' examination's helped inform RAII's assessment adjustments. With the assessment tested across three industries with different AI systems, RAII now needed to develop a scoring method for its assessment.

## 2.3 Assessment Scoring

To develop the assessment's scoring method, RAII examined existing certification program's scoring methods. Some of these examined certification programs included: FairTrade, the Marine and Forest Stewardship Council, Rainforest Alliance, GlobalGAP, LEED, and B Corp (B Corp, 2021; FLOCERT, 2021; Forest Stewardship Council [FSC], 2022; GlobalGAP, 2019; LEED, 2012; Marine Stewardship Council [MSC], 2020; Rainforest Alliance, 2022). These existing certification programs adopted varying scoring methods,

such as ranking assessed criteria from one to five and earning certification if the assessed body achieved a rank above a certain threshold across the assessment; earning one credit per assessed criteria that required earning a certain amount of credits to pass; or assigning numeric values to each possible answer choice then averaging the total earned over the total possible numeric value, requiring a certain average to pass; among others. While each of these varying scoring methods accomplish the same goal - to determine how and when an entity earns certification - they all take slightly different approaches to reach that determination. After examining these various scoring methods, RAII created a scoring method using some parts of the examined scoring methods. A full description of the Responsible AI Certification Program can be found below in Section 4.

## 2.4 RAII Council Formation

Once RAII had developed, tested, and created a scoring method for its assessment, it needed to establish a formal method for approving assessment changes and overseeing the Certification program's governing policies (provided below). To perform both of these duties, RAII formed the RAII Council. The RAII Council is composed of RAII's Executive Director and 10-15 other members from industry, academia, and civil society, each with AI-related technical, legal, or policy expertise. The exact process the RAII Council requires for assessment changes and policy oversight are detailed below in the *Certification Program Change Process* (all further RAII Council duties are found in their *Terms of Reference*) With all major components of the Certification Program completed, at least in their initial versions, RAII sought out formal accreditation from national accreditation agencies.

## 2.5 Calibration

In collaboration with the RAII Working Council, RAII calibrates the RAII Implementation Framework to each certification-seeking use case by adding, modifying, or removing assessment questions; updating scoring, and updating required documentation.

## 2.6 Accreditation

RAII is currently going through the Scheme Owner accreditation process with the Standards Council of Canada (SCC) as outlined by the International Accreditation Forum.

# 3. Responsible AI Certification Program

The Responsible AI Certification Program section comprises five subsections that describe its: (3.1) scope; (3.2) process, (3.3.) assessment, question types, and scoring method; (3.4) final report; and (3.5) certification timeline.

## 3.1 Scope of Certification Program

The Responsible AI Certification Program certifies organizations' AI systems against RAII's AI System Assessment. RAII's AI System Assessment evaluates an AI system's residual risk and compliance to current and forthcoming regulation, standards, and best practices identified by our working groups and regulatory tracking. The assessment evaluates and scores AI system's across the six dimensions and various subdimension outlined in our Implementation Framework. RAII continually tests its assessment on AI systems and validates these test results through a multidisciplinary community of industry experts, policymakers, academics, and other subject matter experts (NIST, 2022; OECD, 2019: 2022). RAII recommends that certification seeking organizations familiarize themselves with our Certification Program's requirements throughout their AI system's development process. This familiarization gives these organizations' the greatest likelihood of passing the certification once the formal process is started.

## 3.2 Certification Process

Before the formal assessment is completed, RAII examines the certification seeking organization's AI system and maps its operations against several previously identified harms and practices. This mapping is used to tailor the

assessment to the AI system. RAII understands that not all AI systems work the same way and therefore should add/remove assessment questions that would otherwise be unanswered. Once this mapping and assessment tailoring are complete, the AI system begins the formal assessment process.

## 3.3  Assessment, Question Types, and Scoring Method

The System Level Assessment is the foundation of RAII's Certification Program, it includes four question types, is run by auditors, and, if passed, earns an organization's AI system the Responsible AI Certification. Auditors will ask all the assessment questions and require documentation submissions to support all question answers. This assessment provides the certification seeking organization with a comprehensive final report detailing each subdimension and dimension's scores, areas of strength and improvement, and tailored recommendations for how to improve each area of improvement. The System Level Assessment is broken into four question types and scoring method: (i) screening questions; (ii) filtering questions; (iii) assessment questions; (iv) bonus questions; and (v) scoring and certification level.

### 3.3.1.  Screening Questions

Screening questions are set to ensure organization's can provide all necessary information and documentation about the AI system seeking certification. If an organization cannot provide all necessary information and documentation, they are not permitted to be formally assessed and certified. The required documentation is used to support assessment question answers. The current list of screening questions and required documentation for formal certification is provided in Appendix A.

### 3.3.2. Filtering Questions

Once an organization has provided all necessary information and documentation about their AI system, the AI system team is asked to fill out a few filtering questions. These filtering questions are used to filter which assessment questions are asked about the AI system, such as what type of AI system is getting assessed and which industry it operates in. The current list of filtering questions is provided in Appendix B.

### 3.3.3. Assessment Questions

Assessment questions are what determines whether an organization's AI system becomes certified, and, if certified, to what level is it certified. All assessment questions are spread across six dimensions, with each dimension being broken down into two or more subdimensions. These six dimensions are: Systems Operations; Explainability and Interpretability; Accountability; Consumer Protection; Bias and Fairness; and Robustness. Each of these questions are scored according to RAII's scoring method (explained in-detail below). The current list of assessment questions with justifications for why each question is asked is provided in Appendix C.

### 3.3.4. Bonus Questions

Bonus questions are asked at the end of the assessment and serve as a way for RAII to see if your organization's AI system goes "above and beyond" what we reasonably expect an organization's AI system to demonstrate. These bonus questions are scored after the assessment questions are scored and do not contribute to the AI system's passing threshold. Bonus question scores are only added to the AI system's score after they have passed the assessment. The current list of bonus questions is provided in Appendix D.

## 3.3.5. Scoring and Certification Level

Each question is scored using a 0-5 range, with 0, 1, 3, and 5 being the only digits assigned to answer choices (See Table 1 for scoring rubric). Once the assessment questions are completed, all scores within each subdimension are totaled, producing a total score for each dimension. The total score for each dimension is then represented as a percentage (total dimension score earned/ total dimension score available). For an AI system to earn certification, that AI system must earn 50%+ of available score in each dimension.

*Table 1. Scoring Rubric*

| Score | Description |
|-------|-------------|
| 0 | Needs Improvement |
| 1 | Satisfactory |
| 3 | Good |
| 5 | Excellent |

If an AI system earns 50%+ of available score in each dimension, each dimension score is totaled to get the total assessment score. This total assessment score is then represented as a percentage (total assessment score earned/total assessment score available). The assessment score percentage is used to determine the AI system's certification level (See Table 2 for assessment score percentages and their corresponding certification level).

*Table 2. Certification Percentage, Level, and Logo*

| Total Score | Level Obtained | Corresponding Mark |
|---|---|---|
| 0-49.9% | Not Certified | N/A |
| 50-59.9% | Certified |  |
| 60-69.9% | Silver |  SILVER |
| 70-79.9% | Gold |  GOLD |
| 80+% | Platinum |  PLATINUM |

# 3.4 Certification Report

After the organization's AI system team has completed the certification assessment and earned their corresponding score, a certification report will be issued to the organization. This report provides a complete breakdown of

their assessment scores across all dimensions and subdimensions and offers recommendations for each dimension and subdimension's areas of improvement. Below is an example of this report's dimension and corresponding subdimensions' breakdown and score summary.

*Figure 2. Certification Score Report*
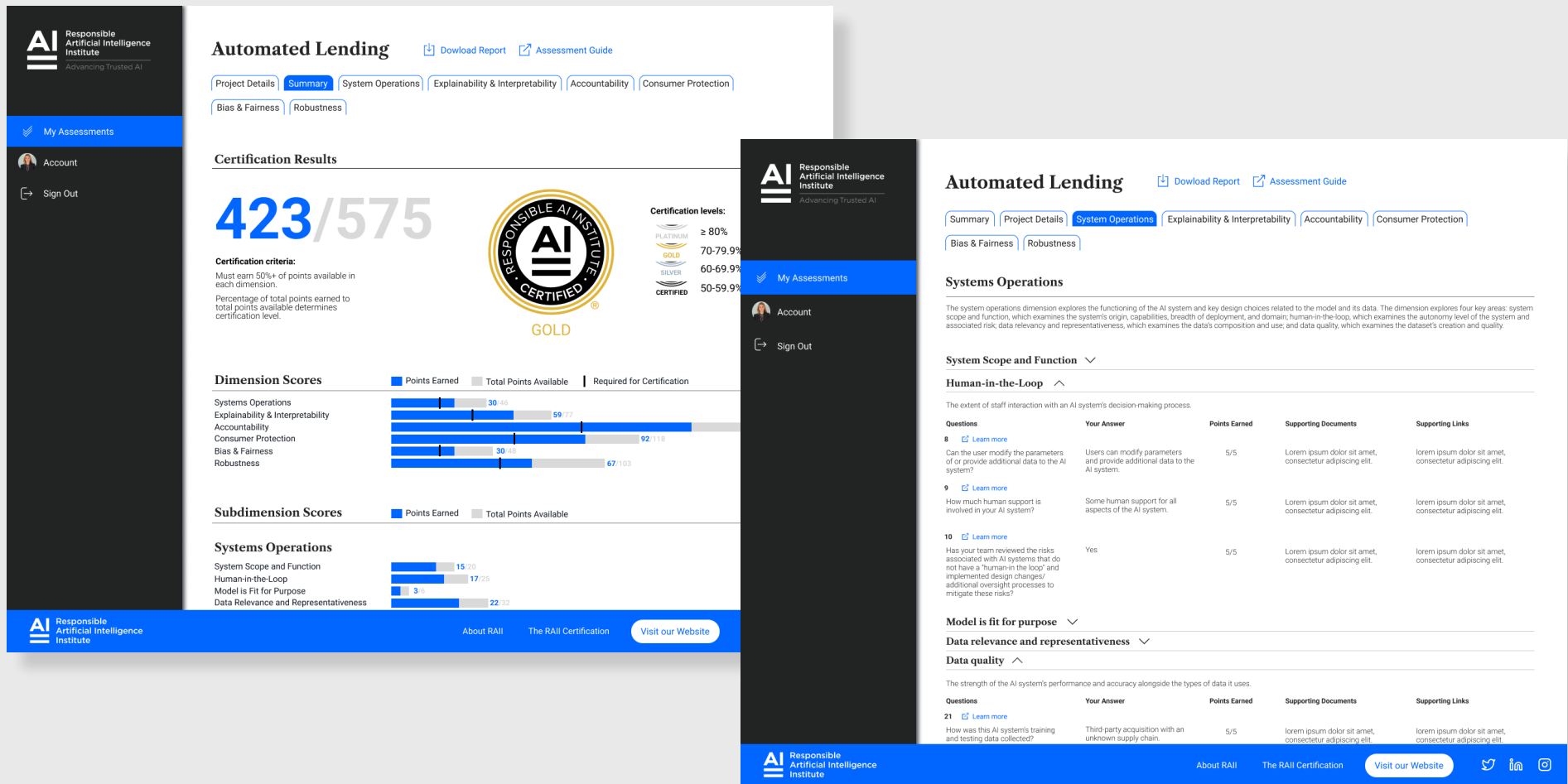
# 3.5 Certification Timeline

The certification timeline begins as soon as the certification seeking organization hires accredited auditors to certify their AI system. Once hired, the organization must submit all required documentation to and coordinate with their hired accredited auditor to complete the certification process. The exact timeline of certification will depend on how quickly an organization can submit all required documentation to and coordinate with their chosen accredited auditors. RAII anticipates that the certification process can be completed within 90 calendar days once an organization has submitted all required documentation to and coordinated with their chosen accredited auditors.

Once an organization's AI system passes the certification, the organization must comply with the following Responsible AI Certification Program policies. Organization's found not complying with the Responsible AI Certification Program policies will have their certification revoked until they can demonstrate full compliance with the Responsible AI Certification Program policies.

# 4. Policies and Processes for the Responsible AI Certification Program

## 4.1 Certification Permissions

Once an organization's AI system has passed all required steps to become RAII Certified, that organization is permitted to display the RAII Certification logo on their website and any public facing products using the certified AI system. If the RAII Certification logo is displayed on the organization's website, it must also contain a description of the AI system that is RAII Certified and a disclaimer that this logo does not mean an organization, and therefore, all AI systems used within that organization, are RAII Certified.

If an organization has displayed the RAII Certification without passing the initial certification process, that organization will be privately asked to remove the certification logo from display. If this private ask is disregarded, RAII will publicly ask the organization to remove the RAII Certification logo from display.

## 4.2 Recertification Process

After organizations' have passed the initial certification process, those organizations must complete an annual recertification process (sometimes referred to as surveillance or renewal audits). This recertification process is generally conducted by the same auditing firm that conducted the organization's initial certification process but with an emphasis on opportunities for the organization to improve in areas of weakness identified in the initial certification process. This focus is to ensure an organization is continuously improving their responsible AI practices.

The recertification process score will replace the initial certification score and its corresponding certification level (i.e., if an organization underperforms, its certification level will be downgraded, if an organization performs equally the certification level will remain unchanged, and if an organization outperforms the certification level will be increased).

The annual recertification process will begin within 30 days prior to the anniversary date of the initial certification processes completion date. If an organization cannot meet this requirement, the organization must provide a justification to their auditor.

# 4.3 Certification Suspension Policy

Certification suspension can occur for two reasons: (i) by failing a recertification process or (ii) if a clear link has been established between an organization's AI system and illegal or harmful activity by a reputable governmental regulatory authority. Regardless of the certification suspension reason, the organization cannot switch auditors during either certification suspension reason until the suspension is lifted or the organization chooses to forgo their certification completely.

## 4.3.1. Failing the Recertification Process

Following the annual recertification process, if an organization's recertification score falls below the minimum certification score requirement (as outlined above), their certification will be suspended for a minimum of six months. After six or more months have passed, the organization can begin the recertification process again. If this recertification process is passed, their suspension will be lifted and they will regain their certification permissions. If the organization fails the recertification process, the six month suspension restarts.

## 4.4 Clear Link to Illegal or Harmful Activity by Government Regulatory Authority

If a clear link has been established between an organization's AI system and illegal or harmful activity by a reputable governmental regulatory authority, then that AI system's RAII certification will be suspended. To end this suspension, the AI system's practices must be either recertified following the recertification process. If the organization fails the recertification process, their certification permissions will be revoked.

## 4.5 Dispute Process

Disputes can be submitted only when an organization disagrees with either their initial certification or recertification assessment result. All disputes must be submitted to the auditing firm that completed the initial certification or recertification assessment. The auditing firm will review, determine the merits of, and decide on the outcome of each dispute. All disputes must follow the three below procedures for audit firm review:

### 4.5.1. Submission of Dispute Claim

Detailed explanation of the dispute with one or more supporting arguments for each dispute area (i.e., if an organization disagrees with two or more scores, they must provide supporting arguments for each of those score disagreements). An organization is permitted to submit one dispute per certification process (either initial or recertification). Any failure to comply with this Dispute Process will void the dispute's submission.

### 4.5.2. Dispute Outcome Timeliness

Audit firms will have up to 120 calendar days to review and decide on dispute submissions. If this timeframe is passed, the dispute will automatically be decided in the organization's favour.

### 4.5.3. Unsolved Disputes

Unresolved disputes between an organization and audit firm that have not been will be sent to the scheme owner (RAII). RAII will facilitate these disputes with the dispute claimant and review the same submission documentation as the auditing firm. RAII decides on the final outcome of these disputes.

## 4.6 Certification Program Update Process

As the AI field and related regulations continue to develop, the Responsible AI Certification Program, particularly its assessment questions, must stay up-to-date. Following existing certification programs' method of addressing continuous developments in their field, RAII has made the RAII Council responsible for approving all Certification Program Changes every six months (FLOCERT, 2021; FSC, 2022; GlobalGAP, 2019; MSC, 2020; Rainforest Alliance, 2022). The RAII Council will ensure the following process is adhered to when approving Certification Program changes.

### 4.6.1. Complete Gap Analysis

RAII will complete a gap analysis between the assessment and newly enacted regulations (found in their AI Regulatory Tracker), published academic and government reports, and AI policy developments. Any identified gap will either require a new, removed, or edited question or answer choices proposed change.

### 4.6.2. Propose Changes

All identified proposed changes from the gap analysis must be documented and submitted for testing and validation with existing and/or new use cases.

### 4.6.3. Test and Validate Proposed Changes

Proposed changes will be tested and validated with existing and/or new use cases. When tested with new use cases, the proposed changes will not be disclosed to the use case owner nor calculated in their assessment scores. The testing is completely blind to give RAII the least biased results. Once the use case test has concluded, the use case owner will be made aware of the proposed changes and their testing. At this point, the use case owner will be asked to discuss what they thought of these proposed changes. If positive, the proposed changes will move onto the next process step, if negative, the proposed changes will either be discarded or reworked.

### 4.6.4. Inform Certified Organizations and Accredited Auditors about Changes

Positive proposed changes will be incorporated into the assessment as scored questions. Once the assessment is updated, a full list of changes will be sent to already certified organizations to prepare them for annual recertification and accredited auditor to anticipate upcoming training and improve their competence around the Responsible AI Certification Program. Any organization already certified will not be expected to meet any new changes until their next recertification nor will their most recent certification score be changed.

### 4.6.5. Train Accredited Auditors on Changes

Accredited auditors will be trained on the new changes as soon as the changes are approved by the RAII Council.

### 4.6.6. Publish Changes

Once all other steps are completed, a complete list of changes will be sent to all certified organizations, accredited auditors, and any organization or audit currently in the process of becoming certified or accredited.

## 4.7 Criteria to Become and Maintain Accredited Auditor Status

To become certified, an organization must hire an accredited auditor to audit against the Responsible AI Certification Program. RAII trains and accredits auditors on how to audit against the Responsible AI Certification Program and provides additional training when Program changes are made.

Auditors must complete RAII's training courses. These training courses teach auditors which documents organizations' are required to submit before the certification audit can begin, how to score questions according to those documents, and how to follow all policies and procedures outlined in this Guidebook.

Once accredited, auditors are required to complete additional training whenever changes are made to the Certification Program as described in the Certification Program Change Process.

# 5.  References

B Corp. (2021). BIA Scoring & Benchmarking. B Impact Assessment Knowledge Base. https://kb.bimpactassessment.net/support/solutions/articles/43000625465-bia-scoring-benchmarking

BSI. (2020). Overview of standardization landscape in artificial intelligence (pp. 1–11). BSI. https://storage.pardot.com/35972/282873/White_paper_on_AI_standards.pdf

Committee of Ministers. (2020). Recommendation CM/Rec(2020)1 of the Committee of Ministers to member States on the human rights impacts of algorithmic systems (pp. 1–15). Council of Europe. https://rm.coe.int/09000016809e1154

Eitel-Poter, R., Corcoran, M., & Connolly, P. (2021). Responsible AI: From principles to practice (pp. 1–15). Accenture. https://www.accenture.com/us-en/insights/artificial-intelligence/responsible-ai-principles-practice

FDA. (2021). Artificial Intelligence and Machine Learning in Software as a Medical Device (pp. 1–8). FDA. https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-learning-software-medical-device

FLOCERT. (2021). Certification Standard Operating Procedure (p. 18). FLOCERT. https://www.flocert.net/wp-content/uploads/2022/01/certification-procedure.pdf

Forest Stewardship Council. (2022). Become Certified. FSC Canada. https://ca.fsc.org:443/en-ca/certification/become-certified

GlobalGAP. (2019). General Regulations Part I – General Requirements (p. 35). GlobalGAP. https://www.globalgap.org/.content/.galleries/documents/190201_GG_GR_Part-I_V5_2_en.pdf

GPAI. (2020). Data Governance Working Group A Framework Paper for GPAI's work on Data Governance (pp. 1–28). The Global Partnership on Artificial Intelligence. https://gpai.ai/projects/data-governance/gpai-data-governance-work-framework-paper.pdf

High-Level Expert Group on Artificial Intelligence. (2019). Ethics guidelines for trustworthy AI (pp. 1–41). European Commission. https://data.europa.eu/doi/10.2759/346720

IAF. (2022). CRITERIA FOR EVALUATION OF CONFORMITY ASSESSMENT SCHEMES (Issue 1; pp. 1–11). International Accreditation Forum. https://iaf.nu/iaf_system/uploads/documents/IAF_MD_25_Criteria_for_the_Evaluation_of_CAS_07012022.pdf

IEEE. (2017). Ethically Aligned Design: A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems (No. 2; Intelligent Systems, Control and Automation: Science and Engineering, pp. 1–226). Institute of Electrical and Electronics Engineers. https://doi.org/10.1007/978-3-030-12524-0_2

ISO. (2019). How to develop scheme documents—Guidance for ISO technical committees (pp. 1–13). International Organization for Standardization. https://www.iso.org/cms/render/live/en/sites/isoorg/contents/data/publication/10/04/PUB100439.html

ISO. (2022). ISO/IEC CD 42001.2: Information Technology—Artificial intelligence—Management system. ISO. https://www.iso.org/cms/render/live/en/sites/isoorg/contents/data/standard/08/12/81230.html

Jillson, E. (2021). Aiming for truth, fairness, and equity in your company's use of AI. Federal Trade Commission: Business Blog. http://www.ftc.gov/business-guidance/blog/2021/04/aiming-truth-fairness-equity-your-companys-use-ai

LEED. (2012). MINIMUM PROGRAM REQUIREMENTS (MPRS) For LEED Canada Rating Systems (pp. 1–35). Canada Green Building Council. https://portal.cagbc.org/cagbcdocs/LEED_Canada-Minimum_Program_Requirements-120731-EN.PDF

Marine Stewardship Council. (2020). MSC Fisheries Certification Process (p. 184). Marine Stewardship Council. https://www.msc.org/docs/default-source/default-document-library/for-business/program-documents/fisheries-program-documents/msc-fisheries-certification-process-v2-2.pdf?sfvrsn=9294350_9

NIST. (2022). AI Risk Management Framework: Initial Draft (pp. 1–23). National Institute of Standards and Technology. https://www.nist.gov/itl/ai-risk-management-framework

OECD. (2019). The OECD Artificial Intelligence (AI) Principles. OECD.AI Policy Observative. https://oecd.ai/en/ai-principles

OECD. (2022). OECD Framework for the Classification of AI systems (No. 323). OECD Publishing. https://doi.org/10.1787/cb6d9eca-en

Office of the Comptroller of the Currency's. (2021). Comptroller's Handbook: Model Risk Management (pp. 1–109). Office of the Comptroller of the Currency's. https://www.occ.gov/publications-and-resources/publications/comptrollers-handbook/files/model-risk-management/index-model-risk-management.html

OSFI. (2020). Developing Financial Sector Resilience in a Digital World: Selected Themes in Technology and Related Risks (pp. 1–35). Office of the Superintendent of Financial Institutions Canada. https://www.osfi-bsif.gc.ca/Eng/fi-if/in-ai/Pages/tchrsk-let.aspx

Rainforest Alliance. (2022). RAINFOREST ALLIANCE POLICY: Adjusted Timelines for the Transition and Full Certification Cycles of the 2020 Rainforest Alliance Certification Program (p. 6). Rainforest Alliance. https://www.rainforest-alliance.org/wp-content/uploads/2022/03/adjusted-timelines-for-transition-full-certification-cycles-2020-certification-program.pdf

Smith, A. (2020). Using Artificial Intelligence and Algorithms. Federal Trade Commission: Business Blog. http://www.ftc.gov/business-guidance/blog/2020/04/using-artificial-intelligence-algorithms

Treasury Board Secretariat of Canada. (2019). Directive on Automated Decision-Making. Government of Canada. https://www.tbs-sct.gc.ca/pol/doc-eng.aspx?id=32592

UNESCO. (2021). Recommendation on the Ethics of Artificial Intelligence (UNESCO Digital Library, p. 44). UNESCO. https://unesdoc.unesco.org/ark:/48223/pf0000381137

World Economic Forum. (2022). Unpacking AI Procurement in a Box: Insights from Implementation (pp. 1–34). World Economic Forum. https://www.weforum.org/whitepapers/unpacking-ai-procurement-in-a-box-insights-from-implementation/

**Responsible Artificial Intelligence Institute (RAII) - Responsible AI Certification - Program Guidebook**
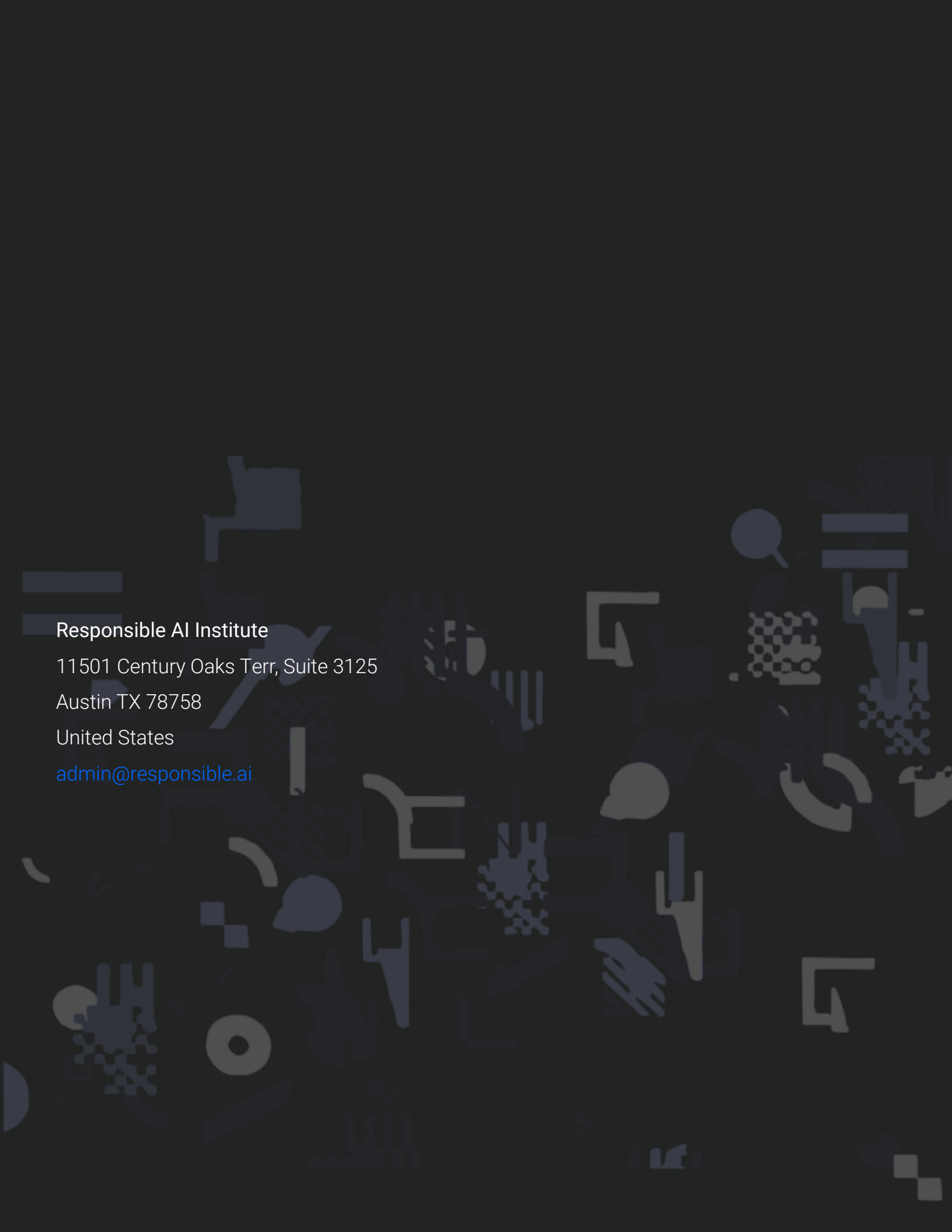
June 2022

**Authors**

- > Benjamin Faveri, Research and Policy Analyst, RAII
- > Amanda Lawson, Research and Policy Analyst, RAII
- > Tina Lassiter, Policy Researcher and Design, RAII
- > Gabriel Bouffard, Business Strategy and Design, RAII

**Corresponding Author**

- > Benjamin Faveri, benjamin@responsible.ai

**Recommendation Citation:**

Faveri, B., Lawson, A., Lassiter, T., & Bouffard, G. (2022). Responsible AI Certification Program Guidebook (pp. 1-33). Responsible Artificial Intelligence Institute. LINK