



Responsible  
Artificial Intelligence  
Institute

Advancing Trusted AI



2022/06/07

# Responsible AI Certification

## Program Scheme Document (Sample)



**Responsible Artificial Intelligence Institute (RAII)**

**Responsible AI Certification - Program Scheme Document (Sample)**

*\* Preliminary Draft - Subject to Change \**

June 6, 2022

Unpublished work © 2022 Responsible Artificial Intelligence Institute.

All Rights Reserved.

# Table of Contents

---

## **04 Background**

04 About RAI

04 About RAI's Certification

06 How the RAI Certification Program Works

06 The Scoring Behind the RAI Certification

08 In This Document

## **10 Scheme Document Sample**

11 Systems Operations

19 Explainability and Interpretability

29 Accountability

34 Consumer Protection

39 Bias and Fairness

43 Robustness

## **48 References**

# Background

---

## About RAI

The Responsible AI Institute (RAI) is an independent and community-driven non-profit organization building tangible governance tools for trustworthy, safe, and fair Artificial Intelligence (AI). RAI's primary objective is to develop the RAI Certification Program, one of the world's first responsible AI (rAI) certification programs, to support organizations as they deploy and scale AI systems.

As a member of the World Economic Forum's Global AI Action Alliance (GAIA), RAI joins hands with over 100 government entities, civil society organizations, private companies, and academic institutions to identify and implement tools and best practices that promote responsible AI.

## About RAI's Certification

As AI systems are becoming increasingly prevalent, governments, companies, and civil society organizations are grappling with approaches to govern AI systems in a consistent manner. Recent research has suggested that certification programs for AI could serve as an important complement to laws and regulations. Organizations around the world have put forward responsible AI principles. Accordingly, a general, international consensus on what constitutes responsible AI has emerged.

The RAI Certification Program takes the guesswork out of what it means to be responsible, by translating globally adopted principles, standards, and regulations into clear implementation requirements. RAI Certification benefits all stakeholders by setting a clear bar for global best practices to implement AI responsibly. RAI certification benefits global regulators by enabling compliance and alignment with their regulatory approaches. Additionally, it benefits investors, executives, and compliance/procurement officers by providing assurance that their AI systems are built on recognized global best practices. Meanwhile, it benefits consumers (who are often “data subjects” of the AI system) by helping protect their privacy rights and civil liberties.

The RAI Certification Program is based on a maturity assessment that evaluates AI systems. Recognizing that not all AI systems are the same, this program tailors its tests to specific industries and functions. RAI’s initial focus industries and functions are: finance, health care, HR, and procurement.

Relying on input and validation from these industries and others, RAI 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). Informed by those researching, designing, building, deploying, using, and overseeing AI, the RAI team has aggregated extensive information to understand:

- > What responsible AI is;
- > Why we need responsible AI; and
- > How certification can support responsible AI adoption.

# How the RAI Certification Program Works

The System Level Assessment is the foundation of RAI's Certification Program. Conducted by auditors, it includes a total of 89 assessment questions, and, if passed, grants an organization's AI system the Responsible AI Certification. The System Level Assessment is broken into four question types and scoring methods: (i) screening questions; (ii) filtering questions; (iii) assessment questions; (iv) bonus questions; and (v) scoring and certification level. The details of each category are explained in the upcoming RAI Certification Guidebook. This scheme document is concerned with assessment questions - (iii) above.

Auditors will ask the assessment questions and require documentation submissions to support the responses. The assessment provides the certification-seeking organization with a comprehensive final report detailing scores for each dimension and subdimension, areas of strength and improvement, and tailored recommendations for how to improve each area of improvement.


## The Scoring Behind the RAI Certification

Assessment questions are scored on the following rubric:

Score	Description
0	Needs Improvement
1	Satisfactory
3	Good
5	Excellent

Some assessment questions (such as “Did you establish mechanisms to inform data subjects on the reasons and criteria behind the AI system’s outcomes?” below in this sample document) are scored out of 3 points, while others are scored out of a possible 5 points. Note that each question in the assessment has an “Other” textbox response option whereby an organization can answer the question without selecting one of the other presented answer choices. In this case, the auditor will use their best judgment to assign a score for that response based on both the provided answer and required documentation (e.g. documentation of the system model or information-sharing policy).

If an AI system earns 50%+ of the 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. The below table includes assessment score percentages and their corresponding certification levels:

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

# In This Document

The RAI Certification Program Scheme Document Sample is a selection of eight questions from the broader assessment scheme document, which contains the 89 scored assessment questions of the RAI Certification Assessment.

Designed to display a cross-section of the broader Assessment, this sample features one or more questions from each of RAI's six Dimensions of Responsible AI: System Operations, Explainability and Interpretability, Accountability, Consumer Protection, Bias and Fairness, and Robustness. Throughout the scheme document, the following information is provided for each question:

- > **Dimension and Sub-Dimension:** Notes which of the six dimensions and 20 sub-dimensions of the RAI Implementation Framework the question belongs to.
- > **Question:** The question as it appears in the RAI assessment.
- > **Responses:** The response options and the points available for each response.
- > **Intent:** A brief summary of the intent behind the question, including what it is meant to ascertain about the AI system and how it relates to rAI.
- > **Rationale:** An explanation of the rationale behind the question, including applicable guiding standards, regulations, academic and industry research, and an example of the principles behind the question in practice.
- > This version of the assessment scheme questions is calibrated to the human resources (HR) field. So, each example in this section is focused on applications of AI in HR contexts.



- > **Documentation:** A description of the evidence required for each response (e.g. AI system team member profiles or AI data label) and why RAI requires it for certification.

All sections were written in accordance with the required scheme development criteria in the IAF (2022) and ISO (2019) scheme development documents. Further information about RAI's Certification Program and delivery processes can be found in our upcoming *RAI Certification Program Guidebook*.

# Scheme Document Sample

---

Scheme Document Sample - Navigation Menu:

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 and Interpretability](#)
  - 2.1 [Communication About the Outcome](#)
  - 2.2 Notification
  - 2.3 [Recourse](#)
  - 2.4 Understanding the AI System's Decisions or Functions
3. [Accountability](#)
  - 3.1 [Organizational Governance](#)
  - 3.2 Team Governance
4. [Consumer Protection](#)
  - 4.1 [Transparency to the User and Data Subject](#)
  - 4.2 Harm to Individuals
  - 4.3 Protections
5. [Bias and 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

# 1. Systems 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 four 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; data relevancy and representativeness, which examines the data's composition and use; and data quality, which examines the dataset's creation and quality.

[Click to return to Scheme Document Sample](#)

2022/05/18

SYSTEMS OPERATIONS

## 1.1 System Scope and Function

The contexts, use cases, and limitations of the AI system.

[Click to return to  
Scheme Document  
Sample](#)

2022/05/18

**QX. “Who provides and modifies the AI system?”****SYSTEMS OPERATIONS: SYSTEM SCOPE AND FUNCTION**

Score	Responses
5	Internal team - can modify
3	Internal team - cannot modify
3	Third party - can modify
0	Third party - cannot modify

**Intent**

This question seeks to determine if a system user has control over the system’s operations and configuration to efficiently address issues that arise.

**Rationale**

As discussed under the *European Commission Liability for Artificial Intelligence* (2019), it is critical that an AI system user “be required to abide by duties to properly select, operate, monitor and maintain the technology in use”. Additionally there are several standard-related discussions that highlight the importance of course correction in AI systems (IEEE, 2017; ISO, 2013: 2015; NIST, 2022). Furthermore, the World Economic Forum’s (2022) *Procurement in a Box* report describes the importance of “conduct[ing] routine audits internally and externally.”

The ability to modify an AI system is critical to reducing errors, biases, and other harmful or sub-optimal consequences. RAI recognizes that these consequences can occur throughout an AI system’s life cycle stages. Due to the importance of modifying the AI system, RAI recommends that internal teams have the ability to modify any AI system decision, especially where potential for significant harm exists.

[Click to return to Scheme Document Sample](#)

2022/05/18

**SYSTEMS OPERATIONS: SYSTEM SCOPE AND FUNCTION**

For accountability and liability purposes, it is more effective to hold those responsible for an AI system accountable when it is provided and modified internally and not subject to third party confidentiality or information withholding. In this question, RAI scores internal team system provision and modification higher as research shows that typically, internal teams can innovate, adjust, and align a system with internal objectives more efficiently than third parties (Kessler et al., 2000).

In the HR context, an AI system that aims to match job applicants with appropriate jobs could systematically reinforce biases by establishing positive correlations with proxy variables of protected categories. One famous example is the algorithm that found job applicants named Jared who played lacrosse were best suited for a job (Bogen & Rieke, 2018). While the AI system might have specifically taken out protected categories (like women) and obvious proxy variables (like attending an elite school), it failed to see that other characteristics (being named “Jared,” lacrosse) were used as proxy variables. The example shows that “machine learning may discover relationships that we do not understand” and, “a statistically valid assessment may inadvertently leverage ethically problematic correlations” (Raghavan et al., 2020). For such cases, it is best to have human oversight by the user that allows modifications of the AI system. Such modification should preferably be internal to make sure the result is approved from within and tailored to the specific needs, values, and objectives of the organization.

**Required Documentation**

System model documentation showing support for the chosen response.

[Click to return to Scheme Document Sample](#)

## 1.2 Human-in-the-Loop

The extent of staff interaction with an AI system's decision-making process.

SYSTEMS OPERATIONS

2022/05/18

[Click to return to  
Scheme Document  
Sample](#)

**QX. “How much human support is involved in your AI system?”****SYSTEMS OPERATIONS: HUMAN-IN-THE-LOOP**

Score	Responses
5	Full human support for all aspects of the AI system
3	Some human support for all aspects of the AI system
1	Some human support for most aspects of the AI system
1	No human support for most aspects of the AI system
0	No human support for all aspects of the AI system

**Intent**

This question seeks to determine the degree of control a human user has in influencing a system’s outputs.

**Rationale**

This question tests the degree of end-to-end automation and human oversight in the AI system (“human-in-the-loop”), with AI systems that have the most human support scoring the highest. RAI’s scoring assumes there is higher risk in fully-automated AI systems compared to those with a human-in-the-loop.

WEF (2022) defines “human-in-the-loop” as the understanding that “human agents should assess, review, and remain accountable for the algorithm’s outputs, ensuring compliance with legal obligations and alignment with user expectations” Relevant research underscores the importance of human influence over AI systems (Jotterand & Bosco, 2020; Demartini et al., 2017). This is also a foundational principle in existing responsible AI standards, such as ISO 9001 (2015) and

[Click to return to Scheme Document Sample](#)



2022/05/18

**SYSTEMS OPERATIONS: HUMAN-IN-THE-LOOP**

Organisation for Economic Co-operation and Development (OECD)'s *Framework for the Classification of AI Systems* (OECD, 2022).

The United Nations (UN) Educational, Scientific and Cultural Organization (UNESCO) notes the importance of human-in-the-loop oversight as well as involving the right, culturally-sensitive human (UNESCO, 2021), which is interrogated further in RAI's Bias and Fairness dimension. The European Union (EU)'s *Ethics Guidelines for Trustworthy AI* (High-Level Expert Group on Artificial Intelligence, 2019) include "human agency and human oversight" as one of seven key requirements for Trustworthy AI. It further includes "human-machine interface" in its proposed AI regulations as key to ensuring that high-risk systems remain responsive long-term. Additionally, Article 14 of the proposed *EU AI Act* states that high-risk AI systems shall include "appropriate human-machine interface tools, that they can be effectively overseen by natural persons during the period in which the AI system is in use" (European Commission, 2021).

Human support is critical in guiding an AI system's operations into alignment with its purpose, objectives, and values, and ensuring continuous improvement in the suitability, adequacy, and effectiveness of the AI management system, a priority reflected across international standards. Google's *Responsible AI Practices* list the need to implement "continued monitoring [to] ensure your model takes real-world performance and user feedback into account" and to update training and testing data "frequently based on who uses your technology and how they use it" (Google, n.d.).

In Human Resources (HR)-related use cases, human support can help reduce employment discrimination risks. For example, job advertisements have sometimes used AI microtargeting advertising tools that fail to advertise to individuals of historically-marginalized

[Click to return to Scheme Document Sample](#)

2022/05/18

**SYSTEMS OPERATIONS: HUMAN-IN-THE-LOOP**

groups. With human-machine interfaces, individuals can recognize these bias patterns and adjust their risk mitigation approach. Thus, where an AI system might fail to identify and deter harmful outcomes, human-machine interface is key to ensuring the smooth functioning of an AI system and identifying and mitigating potential issues (Sonderling, 2021).

**Required Documentation**

Documentation of the system model to verify the level of end-to-end automation, where human support is built in, and to what degree.

[Click to return to  
Scheme Document  
Sample](#)

## 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.

[Click to return to Scheme Document Sample](#)

2022/05/18

## EXPLAINABILITY AND INTERPRETABILITY

## 2.1 Communication About the Outcome

The extent to which people are appropriately informed about the inputs and outputs of the AI system.

[Click to return to  
Scheme Document  
Sample](#)

2022/05/18

**QX. “Did you establish mechanisms to inform data subjects on the reasons and criteria behind the AI system’s outcomes?”**

Score	Responses
3	Yes
0	No

**Intent**

This question asks if the system adequately communicates its decision-making rationale to the data subject.

**Rationale**

Since data subjects have a right to understand how their data is used, it is important that AI systems provide data subjects with a rationale behind outcomes or decisions. Without this, AI systems may not be explainable and interpretable by data subjects. Understanding the rationale for outcomes or decisions made by AI systems is crucial as it enables data subjects to trust an AI system and its ability to make fair and effective decisions. A lack of, or poor AI system decision outcome transparency can lead data subjects to distrust the AI system’s ability to make fair and effective decisions. Improving this transparency can be done without giving data subjects a manual on how to “game” the system to negative effects. If governments or regulatory bodies are among the end-users, explainability and interpretability also helps ensure that their decisions are well-informed.

The Canadian Office of the Superintendent of Financial Institutions (OSFI) highlights soundness, explainability, and accountability as “core

[Click to return to Scheme Document Sample](#)

2022/05/18

## EXPLAINABILITY AND INTERPRETABILITY: COMMUNICATION ABOUT THE OUTCOME

principles to manage heightened risks associated with advanced analytics including AI and ML (OSFI, 2020). Transparency is covered under NIST's *AI Risk Management Framework* (NIST, 2022) as an important means of evening out the information balance between AI system operators and consumers. OECD includes transparency and accountability as key values of responsible AI (OECD, 2019: 2022). The Government of Canada's *Directive on Automated Decision-Making* Section 6.2.3 lists providing a "meaningful explanation to affected individuals of how and why the decision was made" as a key component of transparency (Treasury Board Secretariat of Canada, 2019). The Council of Europe's report on AI systems recommends that transparency levels are maximized and "proportionate to the severity of adverse human rights impacts, including ethics labels or seals for algorithmic systems to enable users to navigate between systems" (Committee of Ministers, 2020). IEEE's *Ethically Aligned Design* guidance includes "achieving transparency" for users, creators, accident investigators, those in the legal process, to "build confidence in the technology" among other reasons (IEEE, 2017). The *EU Ethics Guidelines for Trustworthy AI* (High-Level Expert Group on Artificial Intelligence, 2019) points out that "for a system to be trustworthy, it is necessary to be able to understand why it had a given behaviour and why it has provided a given interpretation." These guidelines highlight that this is important for gaining trust with the developer or the user and to effectively deploy "reliable AI systems."

The Montreal Declaration also includes the importance of transparency and guaranteeing "access to fundamental resources, knowledge and digital tools" for all (Université de Montréal, 2017). Guidance from the EU notes the need for respect for, and enhancement of, human autonomy, for which transparency, explainability, and interpretability are needed. This question is also informed by the EU's transparency assessment list, which asks if usage scenarios have been "specified and clearly

[Click to return to Scheme Document Sample](#)

2022/05/18

**EXPLAINABILITY AND INTERPRETABILITY: COMMUNICATION ABOUT THE OUTCOME**

communicated” and if measures are established to inform the “reasons/ criteria behind outcomes of the product” (High-Level Expert Group on Artificial Intelligence, 2019). The assessment list for this value includes questions about if the user is aware of algorithmic decisions and if users can interrogate algorithmic decisions to fully understand “purpose, provenance, the data relied on, etc.”

With HR-related use cases, informing applicants of the rationale and criteria behind how they are matched to, or screened for jobs by an AI system is critical to ensuring equitable outcomes and trust in the organizations’ job process as well as mitigating harm that results from uncertainty in the process (OECD, 2019; Schumann et al, 2020; Engler, 2021). For example, if a job application process includes a digital exercise and allows disabled applicants to request accommodations, they may be faced with the uncertain choice between revealing their disability and requesting the accommodations they need and potentially being marked as less fit for the job or foregoing accommodations and potentially taking longer and scoring lower on the exercise. To mitigate this problem, ensure fairness across the process, and remain compliant with Equal Employment Opportunities regulation (U.S. Equal Employment Opportunities Commission, 1992) and the Americans with Disabilities Act (U.S. Americans with Disabilities Act, 1990), online skills tests should be tested for accessibility so job seekers with disabilities are not disadvantaged compared to others by seeking accommodations. Additionally, the AI system should establish mechanisms to inform data subjects on the reasons and criteria behind the AI system’s outcomes so that, in this example, a job seeker could have clarity on how their data is used in the evaluation process.

[Click to return to  
Scheme Document  
Sample](#)

2022/05/18

## Required Documentation

Brief description from the respondent to support the response provided, including, but not limited to, an explanation of the AI system's mechanisms to inform data subjects of the outcomes' reasons and criteria and the certification-seeking organization's process of developing these mechanisms, if any.

**EXPLAINABILITY AND INTERPRETABILITY: COMMUNICATION ABOUT THE OUTCOME**

[Click to return to  
Scheme Document  
Sample](#)



2022/05/18

## EXPLAINABILITY AND INTERPRETABILITY

## 2.3 Recourse

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

[Click to return to  
Scheme Document  
Sample](#)

2022/05/18

**QX. “Which recourse mechanisms are available for AI system errors? Select all that apply.”****EXPLAINABILITY AND INTERPRETABILITY: RECOURSE**

Score	Responses
5	Update will be made to the process.
5	If an issue with the system has been identified that could cause harm to an individual or community, it is immediately taken off-line until it is remediated.
5	Users can report any experienced adverse effects.
5	Users are notified.
5	Data subjects are notified.
0	No recourse mechanisms exist.

**Intent**

This question determines the depth of resources provided to redress system errors and reduce harm or negative impacts.

**Rationale**

While system operations should be monitored for reported issues and failures, an organization should also provide capabilities for users or other external parties to report adverse impacts. Recourse mechanisms are important for real or suspected AI system errors or design flaws. Ideally, there should be multiple recourse mechanisms, like notifications, course correction, or user/subject ability to report issues and request resolution. This last recourse mechanism is similar to vehicle recall models, in which the public is notified of a faulty part so they can adjust their behavior in response to the risk and bring their vehicle in to get fixed as soon as possible.

[Click to return to  
Scheme Document  
Sample](#)

2022/05/18

## EXPLAINABILITY AND INTERPRETABILITY: RECOURSE

OSFI includes explainability, accountability, and soundness as their three core principles to manage AI and ML risks (2020). RAI argues that recourse mechanisms are a key component to all three OSFI AI and ML risk management principles.

IBM's ethics recourse notification asks AI teams to provide users a feedback mechanism to proactively solicit input on issues (IBM, 2019). For example, Woebot, a talk therapy chatbot founded by a team of Stanford psychologists and AI experts, prompts users to "Let me know what you think" after it suggests a link, inviting feedback and error reporting (Woebot, n.d.). The EU's guidelines state that good AI governance includes appropriate accountability mechanisms as well as an explanation depending on context (High-Level Expert Group on Artificial Intelligence, 2019).

ISO notes that the AI system should consider how users can contact the appropriate help to report issues (ISO, 2015). It also states that in the case of a nonconformity, the organization should "take action to control and correct it," "make changes to the AI management system, if necessary," and document evidence of action taken and their results. The Montreal Declaration recommends that errors and flaws discovered in systems should be "publicly shared, on a global scale" (Université de Montréal, 2017). UNESCO's principles state that harms should be mitigated throughout the life cycle of an AI system and that "effective remedy should be available against discrimination" (UNESCO, 2021). The EU *Ethics Guidelines for Trustworthy AI* (2019) notes the importance of an AI system offering an "ability to redress." The Council of Europe's Report on AI systems notes the need for "effective remedies" (Committee of Ministers, 2020).

[Click to return to  
Scheme Document  
Sample](#)

2022/05/18

**EXPLAINABILITY AND INTERPRETABILITY: RECOURSE**

In this question, RAI scores each recourse mechanism equally. Answer choices include updating the process, taking the system offline to correct issues, providing user and data subject notifications, and allowing users to report adverse effects, which are all mentioned in AI risk management literature as effective recourse mechanisms (NIST, 2022).

For example, if a data subject suspects that an AI résumé-processing system used their zip code as a proxy for race in their job application process, an effective recourse mechanism could include giving the data subject the ability to report this concern to the organization and request a response within a specified time period. If a serious issue is identified, it could also mean that the AI system team offlines the processor until the issue is resolved and then notifies users and data subjects of the errors.

**Required Documentation**

Documentation of recourse mechanisms to verify that each mechanism sufficiently meets the criteria of a meaningful remedy. For example, notifications should be clear and distributed widely, reporting mechanisms should be user-friendly, and there should be mechanisms to immediately offline a system if needed.

[Click to return to  
Scheme Document  
Sample](#)

### 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.

[Click to return to  
Scheme Document  
Sample](#)

## 3.1 Organizational Governance

The organization's documentation requirements for various AI system changes, oversight processes, and implementation methods.

ACCOUNTABILITY

2022/05/18

[Click to return to  
Scheme Document  
Sample](#)

2022/05/18

## ACCOUNTABILITY: ORGANIZATIONAL GOVERNANCE

**QX. “Have you reviewed the risks associated with your AI system’s models and implemented design changes/ additional oversight processes accordingly?”**

Score	Responses
5	Yes
0	No

**Intent**

This question determines if the AI team has catalogued and responded to their AI system’s risk to minimizing harm.

**Rationale**

AI systems require a thoughtful approach to risk management. Some risks can only be partially mitigated and pose design and implementation trade-offs. Risk identification and mitigation planning should be conducted early and interventions made at appropriate life cycle stages. For example, design changes can often mitigate certain harms that may be harder to mitigate in future stages.

The United Kingdom (UK)’s Centre for Data Ethics and Innovation (CDEI) emphasizes the importance of risk assurance, including impact assessment, bias audit, and impact evaluation (Ahamat et al., 2021). Canada’s Office of the Comptroller of the Currency (OCC) notes that “proper model risk management” and validating the model is key to well-informed board and management decisions (Office of the Comptroller of the Currency, 2021). Per the Global Partnership on Artificial Intelligence (GPAI), “good data governance should be risk-based as the need for data governance measures increases with the potential impact an activity

[Click to return to Scheme Document Sample](#)

2022/05/18

## ACCOUNTABILITY: ORGANIZATIONAL GOVERNANCE

may have on others, including on society and economy at large” (GPAI, 2020).

Google’s *Responsible AI Practices* note the importance of building on quality software engineering best practices to ensure the system is trustworthy works as intended by conducting rigorous unit tests, integration tests, data drift detection tests, building in quality checks, and updating the test set regularly “in line with changing users and use case” to mitigate risk (Google, n.d.). Google notes the importance of staying up-to-date on research advances and defense techniques. The ISO/IEC 27001 standard underscores the need to “maintain and continually improve an AI management system” (ISO, 2013). The EU’s High-Level Group on AI’s recommended guidance (2019) also mentions pruning away biases that inevitably exist in the dataset to have high quality data and minimal-risk outputs.

IBM’s *Everyday Ethics for AI* (2019) points out that real-time analysis of AI sheds light on bias and that the AI team’s responsibility is to “schedule team reviews,” investigate, and mitigate that bias. ISO 9001 and ISO/IEC DIS 23894 (currently under development) recommend that AI policy include a commitment to “continual improvement of the AI management system,” identifying risks and their potential consequences, and building a plan to address risks in its system processes (ISO, 2015: 2022). ISO also recommends corrective action when nonconformities occur and to prevent others from occurring. IEEE recommends that AI teams should cultivate a “safety mindset,” that is vigilant of potential harms and committed to harm prevention (IEEE, 2017). Furthermore, because AI teams cannot fully anticipate the future, IEEE recommends that teams establish “multiple additional strategies to mitigate the chance and magnitude of harm” (IEEE, 2017). UNESCO specifies that “thorough monitoring by the relevant stakeholders as appropriate” of an AI system

[Click to return to Scheme Document Sample](#)



2022/05/18

**ACCOUNTABILITY: ORGANIZATIONAL GOVERNANCE**

is “an essential requirement for trustworthiness” (UNESCO, 2021). Building on the frameworks, “yes” response receives a full 5 points (as opposed to 3 points) because this question is essential to managing an AI system’s risk .

In HR-related use cases, vigilance and risk mitigation are critical to minimizing potential harms. Without this vigilance, potential applicants may be adversely affected in accessing employment, healthcare, dignity, self-determination, and more. For example, if an employer uses automated background checks, it is important to identify and manage potential risks like privacy violations, via data harvesting or consent given only under pressure, and replication of data bias via thoughtful risk management throughout the lifecycle, from designing to decommissioning, (Eaglin, 2017; Martinez, 2020; Nelson, 2019; Smith, 2020; U.S. FTC, 2016).

**Required documentation**

Documentation of the AI system model, evidence that an AI team has documented risks and implemented mitigation measures.

[Click to return to  
Scheme Document  
Sample](#)

## 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.

[Click to return to  
Scheme Document  
Sample](#)

2022/05/18

CONSUMER PROTECTION

## 4.1 Transparency to the User and Data Subject

The degree to which AI system users are informed that AI is assisting with decisions.

[Click to return to  
Scheme Document  
Sample](#)

2022/05/18

**QX. “Are clear data use policies communicated to the data subject (e.g. data subject informed that their data may be used to train AI systems)?”**

Score	Responses
3	Yes
0	No

**Intent**

This question identifies whether policies exist to communicate with data subjects about how their data is used within the AI system.

**Rationale**

A data subject's right to understand how their data is being used concerns autonomy and self-determination with regard to one's data. The EU's principle of “respect for human autonomy” - which extends to data use - suggests providing users with the necessary information to enable them to make a decision in full-determination (High-Level Expert Group on Artificial Intelligence, 2019).

The expectation of clear data use policies is set out in best practices and regulation regarding data usage. The EU's *General Data Protection Regulation* (GDPR) requires that any information about the processing of data belonging to the data subject is described in a “concise, transparent, intelligible, and easily accessible form, using clear and plain language” (European Commission, 2016). The UK's Information Commissioner's Office (ICO) notes that this entails being “clear, open and honest from the start with people on how you will use their data” (ICO, n.d.). The proposed *U.S. Information Transparency & Personal Data Control Act* requires

CONSUMER PROTECTION: TRANSPARENCY TO THE USER AND DATA SUBJECT

[Click to return to Scheme Document Sample](#)

2022/05/18

**CONSUMER PROTECTION: TRANSPARENCY TO THE USER AND DATA SUBJECT**

the controller to provide the user with an up-to-date transparent data use policy regarding sensitive personal information (H.R. 2013 - 116th Congress, 2019). Transparency is a value expressed across a wide range of relevant standards as it helps build trust in an AI system (OECD, 2019; Committee of Ministers, 2020; FDA, 2021; Wachter et al., 2016). IEEE's *Ethically Aligned Design* (IEEE, 2017) guidance states the importance of transparency as it relates to traceability, non deception and honest design, verifiability, and intelligibility. It also suggests ensuring public awareness of (mis)use through measures like data privacy warnings on websites. The British Standards Institution (BSI) discusses the need to develop standards "around protecting consumers and ensuring privacy in usage of AI" (BSI, 2020).

The EU *Transparency Assessment List* asks if the usage scenarios for the product are clearly specified and communicated, whether users are given the ability to seek information about, and revoke, valid consent, and whether the manner in which privacy violation concerns should be raised is clearly communicated (High-Level Expert Group on Artificial Intelligence, 2019). The US. Federal Trade Commission (FTC) recommends practices that "tell the truth about how you use data," through methods such as opening data or source data to outside inspection, publishing the results of independent audits, and following transparency frameworks and standards (Jillson, 2021).

In HR use cases, data subjects often provide extremely sensitive data. Organizations should therefore be clear about how data is used and processed and how a data subject can opt-out. For example, during a job application process, if a person is prompted to consent to their data being used by the platform and affiliated businesses, they might believe that providing this data is required to apply for the job. So, they may agree to provide sensitive information to affiliated businesses

[Click to return to Scheme Document Sample](#)

2022/05/18

**CONSUMER PROTECTION: TRANSPARENCY TO THE USER AND DATA SUBJECT**

without carefully reading through the agreement. In this case, applicants should have an easy way to opt out of their data being shared with affiliated businesses and this option should be clearly and concisely explained if necessary. This is in line with IEEE's *Ethically Aligned Design's* recommendation to "design the terms of service (ToS) as negotiable to consumers (IEEE, 2017).

### Required Documentation

Documentation of data use information provided to data subjects.

[Click to return to  
Scheme Document  
Sample](#)

## 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.

[Click to return to  
Scheme Document  
Sample](#)

2022/05/18

**BIAS AND FAIRNESS**

## 5.1 Bias Impacts

The degree to which the organization has put mitigation processes in place to combat unintended bias and similar issues.

[Click to return to  
Scheme Document  
Sample](#)



2022/05/18

**QX. “Does a member of the team have expertise in employment discrimination and related laws, policies, or compliance as they apply to your system?”****BIAS AND FAIRNESS: BIAS IMPACTS**

Score	Responses
3	Yes
0	No

**Intent**

This question seeks to gauge the extent of diversity awareness and competency on the system team.

**Rationale**

Research speaks to the efficacy of including social justice/discrimination expertise within the design team of an AI system. Several notable AI documents require that AI team members be adequately trained and competent based on education, training, or experience (Committee of Ministers, 2020; ISO, 2015). For HR use cases, this means proper training and expertise to handle the intricacies of employment discrimination. The Montreal Declaration describes the importance of designing and training AI systems to not “create, reinforce, or reproduce discrimination” and to “eliminate relationships of domination between groups and people based on differences of power, wealth, or knowledge” (Université de Montréal, 2017). The EU High-Level Expert Group on AI mentions the importance of paying attention to how AI systems can lead to adverse impacts due to asymmetry specifically in the relationships between employers and employees (2019).

[Click to return to Scheme Document Sample](#)

2022/05/18

**BIAS AND FAIRNESS: BIAS IMPACTS**

In HR use cases, an AI team member should have expertise in employment discrimination as specialized knowledge can inform data and design choices in a way that mitigates discrimination risks. Consider the example of a job matching platform that heavily relies on data collected from online activities such as social platforms like LinkedIn or Facebook. Since veterans or formerly incarcerated applicants may lack social media presences, they may not be matched fairly with jobs. A team member with expertise in employment discrimination could help identify issues like this early and suggest ways to address them.

**Required documentation**

An AI team member's profile outlining their relevant employment discrimination education, training, or experience.

[Click to return to  
Scheme Document  
Sample](#)

## 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.

[Click to return to  
Scheme Document  
Sample](#)

## 6.2 System Acceptance Test is Performed

The extent to which the AI system has been exposed to and tested across several edge cases.

ROBUSTNESS

2022/05/18

[Click to return to  
Scheme Document  
Sample](#)

2022/05/18

**QX. “What safeguards does your AI system have to handle edge cases and extreme scenarios? Select all that apply.”**

**ROBUSTNESS: SYSTEM ACCEPTANCE TEST IS PERFORMED**

Score	Responses
3	Preventive and precautionary measures have been taken.
3	Ongoing research is being conducted to ensure the latest tools are being applied.
3	Third party adversarial testing of the AI system was completed.
3	The AI system has been tested against adversarial attacks.
3	A rigorous threat model to understand all possible attack vectors has been implemented.
3	The unintended consequences resulting in a mistake from the AI system were assessed and mitigated to the best extent possible.
3	Have ensured that a mitigation plan exists for any individual, group, or organization who has an incentive to make the AI system misbehave.
0	AI system does not have any safeguards.

### Intent

This question explores the depth of an AI system’s ability to handle undesirable outcomes when faced with edge cases and extreme scenarios.

[Click to return to Scheme Document Sample](#)

2022/05/18

**ROBUSTNESS: SYSTEM ACCEPTANCE TEST IS PERFORMED**

## Rationale

Research shows the importance of planning for contingencies, even unlikely ones (Pryor & Collins, 1996). Statistically, unlikely events could still be deeply harmful to one person or group of people. Appropriately protecting against edge cases and extreme scenarios is therefore important.

The EU assessment list gauges an AI system's vulnerability to attack and asks what systems exist to ensure data security and integrity, and notes the importance of determining thresholds for unacceptable impact and defining fall-back plans. ISO's 30111 information technology standard recommends evaluating if threats such as adversarial attacks, data poisoning or model stealing can be handled by existing security measures. If not, the system should update its measures to detect and handle these threats (ISO, 2015). Canada's Comptroller's Handbook states that model risk management frameworks should be more "extensive and rigorous" in cases where model failure would have a particularly harmful impact (Office of the Comptroller of the Currency, 2021). The GPAI recommends implementing "appropriate safeguards" when using data with a high degree of sensitivity and output data with a high potential for harm (GPAI, 2020).

In the HR use case, inadequate system performance in the face of edge cases or extreme scenarios could potentially result in the violation of discrimination protections.

[Click to return to Scheme Document Sample](#)

2022/05/18

## Required documentation

Documentation of the system model showing analysis and implementation of safeguards indicating a threat model that shows rigor and depth.

**ROBUSTNESS: SYSTEM ACCEPTANCE TEST IS PERFORMED**

[Click to return to  
Scheme Document  
Sample](#)

# References

---

Ahamat, G., Chang, M., & Thomas, C. (2021). Types of assurance in AI and the role of standards. Centre for Data Ethics and Innovation Blog. <https://cdei.blog.gov.uk/2021/04/17/134/>

Berendt, B. & Preibusch, S. (2017). Toward Accountable Discrimination-Aware Data Mining: The Importance of Keeping the Human in the Loop—and Under the Looking Glass (pp. 135-152). Big Data Vol. 5, No. 2. <http://doi.org/10.1089/big.2016.0055>

Bogen, M. & Rieke, A. (2018). HELP WANTED. An Examination of Hiring Algorithms, Equity, and Bias (pp. 1-72). Upturn. <https://www.upturn.org/static/reports/2018/hiring-algorithms/files/Upturn%20--%20Help%20Wanted%20-%20An%20Exploration%20of%20Hiring%20Algorithms,%20Equity%20and%20Bias.pdf>

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](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>

Demartini, G., Difallah E., Gadiraju U., & Catasta, M. (2017). An Introduction to Hybrid Human-Machine Information Systems. Foundations and Trends in Web Science (pp. 1–87). <https://nyuscholars.nyu.edu/en/publications/an-introduction-to-hybrid-human-machine-information-systems>



Eaglin, J. M. (2017). Constructing Recidivism Risk. Emory Law Journal, 67(59), 64. <https://scholarlycommons.law.emory.edu/elj/vol67/iss1/2/>

Engler, A. (2021). Auditing employment algorithms for discrimination. Brookings Institution. <https://www.brookings.edu/research/auditing-employment-algorithms-for-discrimination/>

European Commission. (2016). Regulation (EU) 2016/679 - General Data Protection Regulation. <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A02016R0679-20160504&qid=1532348683434>

European Commission. (2021). Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) And Amending Certain Union Legislative Acts. European Commission. <https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:52021PC0206&from=EN>

Expert Group on Liability and New Technologies – New Technologies Formation. (2019). Liability for Artificial Intelligence and other emerging digital technologies. (pp. 1-65). European Commission. [https://www.europarl.europa.eu/meetdocs/2014\\_2019/plmrep/COMMITTEES/JURI/DV/2020/01-09/AI-report\\_EN.pdf](https://www.europarl.europa.eu/meetdocs/2014_2019/plmrep/COMMITTEES/JURI/DV/2020/01-09/AI-report_EN.pdf)

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>

Google. (n.d.). Responsible AI practices. Google AI. <https://ai.google/responsibilities/responsible-ai-practices/>

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>

H.R. 2013 - 116th Congress. (2019). Information Transparency & Personal Data Control Act. <http://www.congress.gov/>

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](https://iaf.nu/iaf_system/uploads/documents/IAF_MD_25_Criteria_for_the_Evaluation_of_CAS_07012022.pdf)

IBM. (2019). Everyday Ethics for Artificial Intelligence. IBM. <https://www.ibm.com/watson/assets/duo/pdf/everydayethics.pdf>

ICO. (n.d.). Principle (a): Lawfulness, fairness and transparency. Information Commissioner's Office. <https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/principles/lawfulness-fairness-and-transparency/>

ICO. (n.d.). Rights related to automated decision making including profiling. Information Commissioner's Office. <https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/individual-rights/rights-related-to-automated-decision-making-including-profiling/>

ISO. (2013). ISO/IEC 27001 - Information security management. International Organization for Standardization. <https://www.iso.org/standard/69725.html>

ISO. (2015). ISO/IEC 30111:2019 - Information technology – Security techniques – Vulnerability handling processes (pp. 1-32). International Organization for Standardization. <https://www.iso.org/standard/69725.html>

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](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. International Organization for Standardization. <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>

Jotterand, F. & Bosco, C. (2020). Keeping the “Human in the Loop” in the Age of Artificial Intelligence (pp. 2455-2460). Sci Eng Ethics 26. <https://doi.org/10.1007/s11948-020-00241-1>

Kessler, E.H., Bierly, P.E., & Gopalakrishnan, S. (2000). Internal vs. external learning in new product development: Effects on speed, costs and competitive advantage (pp. 213-224). R&D Management. <https://onlinelibrary.wiley.com/doi/abs/10.1111/1467-9310.00172>

Martinez, A. (2020). Considering AI in Hiring? As Its Use Grows, So Do the Legal Implications for Employers. Hire Right Blog. <https://www.hireright.com/blog/compliance/considering-ai-in-hiring-as-its-use-grows-so-do-the-legal-implications-for-employers>

- Nelson, A. (2019). Broken Records Redux: How Errors by Criminal Background Check Companies Continue to Harm Consumers Seeking Jobs and Housing. National Consumer Law Center. <https://www.nclc.org/issues/rpt-broken-records-redux.html>
- 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. (2019). Transparency and explainability (Principle 1.3). OECD.AI Policy Observative. <https://oecd.ai/en/dashboards/ai-principles/P7>
- 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](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>
- Pryor, L. & Collins, G. (1996). Planning for Contingencies: A Decision-based Approach. The Institute for the Learning Sciences, (pp. 288-339). Northwestern University. <https://arxiv.org/pdf/cs/9605106.pdf>

Raghavan, M., Barocas, S., Kleinberg, J., & Levy, K.. (2020). Mitigating bias in algorithmic hiring: evaluating claims and practices (pp. 469-481). FAT\* '20: Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency. <https://doi.org/10.1145/3351095.3372828>

Schumann, C., Foster, J., Mattei, N., & Dickerson, J. P. (2020). We Need Fairness and Explainability in Algorithmic Hiring. AAMAS. <https://dl.acm.org/doi/abs/10.5555/3398761.3398960>

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>

Sonderling, K. (2021). No Bots need apply: Microtargeting employment ads in the age of AI. HR Dive. <https://www.hrdive.com/news/no-bots-need-apply-microtargeting-employment-ads-in-the-age-of-ai/601502/>

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>

United States. (1990). Americans With Disabilities Act of 1990, 42 U.S.C. § 12101 et seq. <https://www.ada.gov/pubs/adastatute08.htm>

United States. (1992). Equal Employment Opportunity Commission. <https://www.eeoc.gov>

Université de Montréal. (2017). Montreal Declaration for a Responsible Development of Artificial Intelligence. Université de Montréal. <https://recherche.umontreal.ca/english/strategic-initiatives/montreal-declaration-for-a-responsible-ai/>

U.S. FTC. (2016). Big Data: A Tool for Inclusion or Exclusion? Understanding the Issues (pp. 1-33). U.S. Federal Trade Commission. <https://www.ftc.gov/reports/big-data-tool-inclusion-or-exclusion-understanding-issues-ftc-report>

Wachter, S., Mittelstadt, B., & Floridi, L. (2016). Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation. International Data Privacy Law, 2017. <https://ssrn.com/abstract=2903469>

Woebot. (n.d.). Woebot Health. Woebot. [woebot.io](https://woebot.io)

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/>

2022/05/18

Responsible AI Institute

11501 Century Oaks Terr, Suite 1325

Austin TX 78758

United States

[admin@responsible.ai](mailto:admin@responsible.ai)