



Human Rights AI Impact Assessment

November 2024



LAW COMMISSION OF ONTARIO
COMMISSION DU DROIT DE L'ONTARIO



Ontario Human
Rights Commission
Commission ontarienne des
droits de la personne



1.

Introduction

The Law Commission of Ontario and Ontario Human Rights Commission have joined together to create an AI impact assessment tool to provide organizations a method to assess AI systems for compliance with human rights obligations. The purpose of this human rights AI impact assessment (“HRIA” or “the tool”) is to assist developers and administrators of AI systems to identify, assess, minimize or avoid discrimination and uphold human rights obligations throughout the lifecycle of an AI system.

The HRIA is based on the following principles:

Human rights must be forefront in AI development

Human rights recognize the dignity and worth of every person and provide for equal rights and opportunities without discrimination.

Bias and discrimination in AI are real and complex. Bias and discrimination can be easy to overlook or ignore. Left unchecked, AI can cause deep and longstanding harm to individuals, communities and organizations. Bias and discrimination can also present economic, legal and public relations consequences for organizations.

Assessing for bias and discrimination is not a simple task. As such, it should not be an afterthought or minor consideration but be integrated into every stage of design, development and implementation of AI.

Ontario’s and Canada’s Human Rights Laws apply to AI systems

This tool is expected to help designers, developers, operators and owners of AI systems to identify and reduce bias and discrimination. The tool is a guide that, when applied carefully and thoughtfully, should help organizations build better AI and help in understanding human rights obligations.

Assessment of human rights in AI is a multi-faceted process that requires integrated expertise

This tool encourages designers of AI systems to involve human rights experts and engage with a diversity of communities throughout the lifecycle of an AI system, including its design, development and operation.

The HRIA is one piece of AI governance

There are many important human rights and legal issues that can arise with AI. Removing or reducing bias does not necessarily resolve other issues such as surveillance, privacy, data accuracy, fairness etc. This tool should be seen as complementary and to be used in conjunction with other AI assessment tools that address, for example, procedural fairness or privacy. Organisations and industry groups are encouraged to adopt and adapt this tool into existing assessments.

2.

Introduction to the HRIA

a. Purpose and Limitations of the HRIA

Existing and emerging AI regulations are increasingly requiring organizations to conduct AI impact assessments and/or comply with human rights laws prior to deploying AI systems. Not all AI impact assessments are the same. For example, privacy, while an element of human rights, is best addressed through specific privacy impact assessment methodologies. This HRIA is intended to provide organizations with a comprehensive tool to assess bias and discrimination.

Purpose

The HRIA is a practical, understandable framework to help private and public organizations assess and mitigate the human rights impact of AI systems in a broad range of applications.

The HRIA is intended to:

- Strengthen knowledge and understanding of human rights impacts;
- Provide practical guidance on specific human rights impacts, particularly in relation to non-discrimination and equality of treatment; and
- Identify practical mitigation strategies and remedies to address bias and discrimination from AI systems.

Limitations

The HRIA does not constitute legal advice and does not provide a definitive legal answer regarding any adverse human rights impacts, including violations of federal or provincial human rights law or other relevant legislation. Organizations and individuals should seek independent legal advice if they have concerns regarding their compliance with applicable legislation and their legal obligations.

While the use of the HRIA is to help identify, address and remedy potential adverse human rights impacts, an organization or individual will not be protected from liability for adverse human rights impacts, including unlawful discrimination, if they claim they complied with or relied on the HRIA.





b. Form of the HRIA

This HRIA is a series of questions with explanations. It is split into two parts.

Part A is an assessment of the AI system for human rights implications. In this section, organizations are asked questions about the purpose of the AI, the significance of the AI system, and the treatment of individuals and communities.

Part A includes questions to help organizations assess an AI system to determine whether it is “high risk” for human rights issues; whether the AI system is demonstrating differential treatment on protected grounds; whether differential treatment is justified; and whether the system is accommodating different needs.

Part B is about mitigation. Once the AI system has been categorized, Part B provides a series of questions to assist organizations to minimize the identified human rights issues in the given AI system.

c. Human Rights Law as the Benchmark

Human rights are protected by the Ontario *Human Rights Code* (the *Code*), the *Canadian Human Rights Act* (the *Act*), and the *Canadian Charter of Rights and Freedoms* (the *Charter*). The *Code* focuses on equality and freedom from discrimination. The *Charter* is broader in scope. In addition to the right to equality and freedom from discrimination, it includes rights to liberty, to a fair trial, freedom of expression, freedom of mobility, and other rights.

The *Charter*, the *Act*, and the *Code* protect individuals and groups from discrimination based on enumerated grounds such as disability, race, sex, age, and religion. The *Code* applies to everyone in Ontario, including public and private entities, and focuses on social areas such as jobs, housing, services, unions and contracts. The *Act* protects against discrimination in federal jurisdiction, including discrimination in federal government/programs and the federally-regulated private sector, including banking, railways and airlines. The *Charter* applies only to government action.

3.

Conducting an HRIA

The team responsible for conducting an HRIA should include people with a variety of socio-technical expertise – technological, legal, human rights, business strategy and community advocacy.

AI assessments are iterative and should be considered circular, not linear. Parties should be assessing an AI system for human rights issues, taking steps to mitigate issues, and then returning to assess the system again.

Organizations often conduct a number of risk assessments. The intention with the HRIA is to provide a tool specific to AI and human rights discrimination that can be tailored for various applications and uses, such as adding this tool to other existing AI assessment procedures.

a. Who Should Use the HRIA?

This tool is relevant and applicable to any organization, public or private, intending to design, implement, or rely on an AI system.

This tool is designed with a focus on the laws in Ontario. However, it could be useful to any organization or individual in Canada.

This tool is designed to apply broadly to any algorithm, automated decision-making system or artificial intelligence system.¹

b. When to Use the HRIA

This assessment should be completed:

- 1) when the idea for the AI system is explored and developed;
- 2) before the AI system is made available to external parties (eg: before a vendor makes a model or application available to purchasers, or service providers deploy an AI technology for customer service);
- 3) within ninety days of a material change in the system;
- 4) yearly as part of regular maintenance and reviews.

The frequency required for human rights assessments will vary depending on the type and use of the AI system. Large language models and complex neural networks that are fed new data regularly will require recurring monitoring and assessment.

¹ For consistency, the term “AI system” is used throughout the tool.

4.

Structure of the HRIA

The HRIA has two parts:

Part A: Impact and Discrimination

In **Section 1**, parties are asked to describe the purpose of the AI system, what it is intended to do, and the reason for implementing the system.

Section 2 is an assessment of the significance of the impact of the AI system on individuals and communities. Human rights law applies in all circumstances and to all communities. This section is not to determine whether an organization is “in the clear” with regards to human rights, but rather, whether the system is at high risk for potential human rights violations. This section assesses whether an AI system is at high risk of human rights violations because of the context in which it operates and/or because of the population that is expected to be impacted by the AI system.

If your AI system is considered to be at high risk, then you should continue to fill out sections 3 and 4. If not, you do not have to continue to fill out the assessment, but should continue to monitor for human rights issues.

Section 3 is an assessment of whether the AI system disproportionately impacts individuals or communities on protected grounds, and whether the disproportionate treatment can be justified. Justifications for differential treatment are rare and must be very specific. We encourage organizations to seek legal advice from a human rights expert if you suspect that the AI system might fit into a statutory or legal exception or justification.

Section 4 is an assessment of whether the AI system accommodates people with disabilities and children.

Section 5 provides guidance and additional steps based on the outcome of sections 1, 2, 3 and 4.

Part B: Response and Mitigation

Part B is about mitigation. After completing Part A, the AI system will be categorized into one of several potential risk levels. Part B provides guidance about steps that can be taken to minimize human rights issues including transparency, explainability, disclosure, data accuracy, and audit, for example. The questions in this section are intended to both inquire and suggest what mitigation steps need to take place.



Part A
Assessing Purpose and Impact

SECTION 1

The purpose of the AI system

Questions 1-4 are intended to identify the purpose of an AI system, why it is needed and the objective it hopes to achieve. These questions are important to human rights analysis because they:

- Promote transparency and understanding about why an AI system is being built or implemented.
- Encourage developers to consider whether there are other means to achieve the AI system's stated purpose.
- Provide a base level against which to measure the success of the AI system once it is in operation.
- Help developers assess the proportionality of the AI system, i.e. whether or how the objectives of the AI system balance against the real or potential risks to human rights.

1.

What is the general function of the AI system?

Provide a general description of the function of the AI system. Possible functions could include triaging, assessing eligibility for services, calculating the quantity of benefits or services, assessing risk, assessing performance, providing information or guidance, making recommendations, making decisions, or scanning for anomalies (such as detecting for fraud).

2.

What is the intended purpose of the AI system? What are the main and secondary objectives? If there is more than one objective, they should be ranked.

State the objective as concretely and specifically as possible. Potential objectives include: to improve efficiency, accuracy, or fairness; to lower costs; to better target resources or services; to assess performance; or to provide a service. If the AI system is intended to solve a problem, state what the problem is.

3.

Who is the AI system designed to benefit? Who could be harmed by the AI system?

After describing the function of the AI system (question 1) and the intended purpose of the system (question 2), consider the people who will be affected by the system. Who will benefit from the system and how? Who could be harmed if the system fails or makes unintended or undetected errors?

4.

What are the alternatives for meeting these objectives? Why is an AI system needed or preferred?

Why is an AI system the preferred option to meet the objectives set out in question 2? Are there other options?



Part A
Assessing Purpose and Impact

SECTION 2

Is the AI system at high risk for human rights violations?

Questions 5-13 assess whether the AI system is at high risk for human rights violations. This analysis focuses on two potential risks:

First, questions 5-9 ask if the AI system is being used in a context or situation where human rights are likely to be affected.

Second, questions 10-13 ask if the AI system is likely to capture or impact a historically marginalized or disadvantaged population.

These questions should be considered early in the AI design process.

Questions 5-9: Use and Context

There are no circumstances where human rights law is not relevant or applicable. However, there are certain situations where human rights are more significant and/or where discrimination is likely to cause greater harm.

Question 5 asks developers and deployers to assess whether AI was a factor in a decision or decision-making process. AI systems that make or contribute to a decision are more likely to attract human rights concerns.

Question 6 asks organizations to consider whether the AI system will be used in a context or situation where human rights issues are likely to be present. These areas include housing, employment, education, government services, health care, or other services.

Questions 7 and 8 ask developers and deployers to evaluate the AI system's impact on an individual's body and behaviour – deeply personal attributes that go to the core of the human experience.

Question 9 asks about the potential reach of the AI system. An AI system that impacts more people has a greater chance of raising human rights issues.

Questions 10-13: Population Affected

Questions 10-13 ask developers to consider whom the AI system is likely to impact. All humans have a right to equality and a right to live their life free from discrimination. However, treating people the same does not necessarily result in substantive equality. In some cases, populations that have experienced current and historical barriers to full and fair participation in society may continue to face discrimination unless it is detected and addressed. If the AI system targets, affects or overlooks a historically disadvantaged group, the AI system may be at high risk.

The purpose of these questions is to determine whether the AI system is likely to impact a historically disadvantaged population.

Questions 5-9 and 10-13 identify AI systems that could be at high risk for human rights violations because of their use or context, or because they may affect a historically disadvantaged population. AI systems that fall into either category (or both) need to be assessed and monitored for human rights issues frequently and carefully.

Two important points to keep in mind when answering these questions:

- Organizations should have a multidisciplinary team to assess human rights issues in AI systems. This model promotes thoughtful and comprehensive analysis.
- Many of the questions in Sections 2, 3 and 4 request parties answer with YES, NO or I DON'T KNOW. We encourage parties to consider the perspectives of as many different stakeholders as possible when answering, particularly from those who may be impacted by the AI system.

If the AI system is being deployed in the following sectors or areas, the AI system is likely operating in an area covered by human rights law:

- a. **Rights and freedoms of the individual or community affected** (this includes the use of AI in the justice system, court services and administrative tribunals, adjudication, policing, law enforcement, sentencing, corrections, probation and parole).
- b. **Government action or government services such as social services, employment insurance, regulatory bodies, housing** (including, but not limited to eligibility, fraud detection, access to services, application screening, facial recognition technology, credit scores).
- c. **Health, safety, or well-being of the individual or community affected** (including but not limited to screening, selection, health or medical advice, access to services, cost and provision of services, resource allocation, and prioritization).
- d. **Public education including university and colleges** (including but not limited to student risk or academic assessments, determinations for suspension or expulsion, resources allocation, teacher performance reviews).
- e. **Employment** (including but not limited to hiring, referral, job screening, performance management, training).
- f. **Contracts** (including rental accommodation).
- g. **Membership in unions, trade or professional associations** (including but not limited to access to membership, access to benefits, performance reviews).
- h. **Goods, services and facilities** (including but not limited to screening of applicants, selection of applicants, access to services, provision of services, professional advice, cost of services, prioritization. This includes provision of professional services such as banking, insurance, law, rental housing).
- i. **Children and youth** (including but not limited to child welfare administration, such as determinations of opening child welfare files or child apprehension, allocation of child benefits, and influencing children's learning, development, and online interactions with schools, and learning and social communities).

5.

Does the AI system make a decision, or provide information or a score that may influence a decision?

If the AI system is being used to make a decision, recommend a decision, provide information, or provide a score that will impact or influence a decision, then a human rights assessment is crucial.

A decision can be one small step in a series of decisions; it does not need to be a final decision to attract human rights scrutiny. A decision could include a referral, recommendation, prioritization, or assessment. It can be a decision about an individual or a community. AI systems designed to assist with casual everyday tasks, such as text editors, digital assistants, e-payments and navigation tools do not fall in this category.

☐ YES ☐ NO ☐ DON'T KNOW

6.

Does the AI system make or aid decisions in an area covered by human rights law?

The categories listed in the information box on this page are areas that are either protected by domestic human rights legislation and/or are included in international human rights' treaties. These areas attract greater human rights scrutiny because they are significant aspects of the human experience. The further away an AI system is from impacting a community or individual, the lower the human rights impact is likely to be.

☐ YES ☐ NO ☐ DON'T KNOW

7.

Does the AI system employ biometric tools (eg: facial recognition technology, fingerprints, voice prints, gait analysis or iris scans)?

Biometric AI tools assess human characteristics that align with protected grounds such as the colour of a person's skin or the shape of their face. An AI system that monitors, assesses or relies on biological factors raises the potential for human rights issues.

In addition to issues with discrimination, biometric AI tools introduce significant concerns about surveillance, privacy, free speech and freedom of association.

☐ YES ☐ NO ☐ DON'T KNOW

8.

Does the AI system track behaviour? For example, does the AI system analyze keystroke patterns, purchasing habits, patterns of device use, or use affect recognition?

Behaviour is deeply personal and can be a proxy for a protected ground. An AI system that monitors, assesses, or relies on human behaviour patterns could affect human rights if the system makes distinctions that categorize or make recommendations based on those patterns.

In addition to issues with discrimination, behaviour tracking AI tools introduce significant additional human rights concerns such as surveillance, privacy, and control.

☐ YES ☐ NO ☐ DON'T KNOW

9.

Does the AI system have the ability to influence, elicit, or predict human behaviour, expression, and emotion on a large scale?

This could include social media sites, chatbots, and search engines. These systems raise human rights concerns primarily because of the size of the audience they reach. They have the potential to cause harm on a large scale.

☐ YES ☐ NO ☐ DON'T KNOW



RESULTS

Question 5

☐ YES ☐ NO ☐ DON'T KNOW

Question 6

☐ YES ☐ NO ☐ DON'T KNOW

Question 7

☐ YES ☐ NO ☐ DON'T KNOW

Question 8

☐ YES ☐ NO ☐ DON'T KNOW

Question 9

☐ YES ☐ NO ☐ DON'T KNOW

If you answered “yes” to question 5 and “yes” to any other question in this section (questions 6-9), the AI system may be at high risk for human rights issues. A high risk finding requires ongoing human rights review, consideration and mitigation. Please continue the assessment and complete the rest of section 2 and sections 3 and 4.

If you answered “yes” to question 5 and “no” to every other question in this section (questions 6-9), the AI system is not at high risk for human rights issues. Please complete questions 10-13.

If you answered “no” to question 5 and “yes” to any other question in this section (questions 6-9), human rights issues could arise but the AI system is not at high risk at the present time. We encourage you to continue the rest of the assessment and to monitor the AI system for drift or change.

If you answered “no” to question 5 and “no” to every other question in this section (questions 6-9), the AI system is not at high risk for human rights issues. Please complete questions 10-13.

If you do not know the answer to one or more of these questions, you should seek input from colleagues and experts who do. After this section is completed, you should continue the assessment.

10.

Is the AI system operating in an area where there have been concerns raised about bias and discrimination in the past?

A human rights expert will be able to advise on what areas of society and activities have historically had concerns with bias and discrimination. Examples include: age discrimination in job applications and terminations; racial discrimination in housing; and postal codes as proxies for ethnic origin in banking or credit scoring.

☐ YES ☐ NO ☐ DON'T KNOW

11.

Who is subjected/exposed to the AI system? Be specific.

Who will be affected by the AI system? Persons or communities potentially affected by an AI system could be a broad or narrow group, potentially including customers; job, benefit, or service applicants; residents of a specific geographic area; etc.

12.

What are the demographics of the people who are subjected/exposed to the AI system? Be specific.

Take the answer from question 11 and consider the demographics of the people affected by the AI system. In answering this question consider the individuals':

- a. Socio-economic status
- b. Geographic location
- c. Demographic information
- d. Whether they fit into a protected category such as age, race, sex, religion, disability, etc.
- e. Whether they are in a vulnerable circumstance (eg: unwell, unemployed or unhoused)
- f. Other, please specify



13.

Does the AI system have the potential to impact a historically disadvantaged group?

Consider your answers to questions 11 and 12.
Are historically disadvantaged groups included among the people who are likely to be impacted or overlooked by the AI system?

☐ YES ☐ NO ☐ DON'T KNOW

RESULTS

Question 10

☐ YES ☐ NO ☐ DON'T KNOW

Question 13

☐ YES ☐ NO ☐ DON'T KNOW

If the answer to questions 10 or 13 is “yes”, the AI system is at high risk for human rights issues. Please continue the assessment and complete sections 3 and 4.

If the answer to questions 10 or 13 is “no”, the AI system is not at high risk on the basis of who is affected by the AI system. The system may still be at high risk, however, if it is a high risk use or context (questions 5-9).

If you do not know the answer to questions 10 or 13, you should consult with experts or colleagues. This question should be answered before deploying an AI system.

If you are not at high risk after answering questions 5-13, you do not need to complete the rest of the assessment. We recommend that you revisit the human rights assessment within 90 days of a material change in the system and as part of annual maintenance.



Part A

Assessing Purpose and Impact

SECTION 3A

Does the AI system show differential treatment?

Questions 1-13 assessed whether an AI system is at high risk for human rights issues.

Section 3 helps determine if the AI system is discriminatory under human rights law. More specifically, questions 14-22 ask detailed questions about two legal issues that help determine whether a system is discriminatory: 1) Which communities are impacted by an AI system 2) Whether the AI system demonstrates differential treatment among or within a community.

Questions 14-22 should be asked after the system has been developed but before it is deployed.

Questions 14, 15, 16 and 17 assess whether output from the AI system demonstrates differential treatment on a protected ground under human rights law.

Questions 18 and 19 are relevant only if you are unable to answer questions 14-17. Questions 18 and 19 ask about gaps in the data and request that you clarify what you can and cannot assess in the AI system.

If the answers to questions 14-17 suggest that the AI system demonstrates differential treatment, or if you do not know whether the AI system has differential treatment, questions 20, 21 and 22 help determine whether the differential treatment is acceptable.

If after answering the questions in this section, you find that your AI system is shown to be discriminatory on protected grounds, continue to sections 4 and 5.

14.

What are the demographic characteristics of the people flagged by the AI system, or for whom it recommends or makes a decision? Be specific.

What does the testing/auditing of the AI results tell you?

15.

Does the AI system produce results that differentiate based on one or more protected grounds?

Are there specific communities that are over- or under-represented in the results of the AI system?

Does the AI system target, directly or indirectly, people based on one or more of the following protected grounds of discrimination?

The text box on page 19 discusses protected grounds in more detail.

- a. Age
- b. Ancestry, colour, race
- c. Citizenship
- d. Ethnic origin
- e. Genetic characteristics
- f. Place of origin
- g. Creed (including religion)
- h. Disability
- i. Family status
- j. Marital status
- k. Gender identity, gender expression
- l. Sex
- m. Sexual orientation
- n. Receipt of public assistance (in housing)
- o. Record of offences (in employment)

☐ **YES** ☐ **NO** ☐ **DON'T KNOW**

PROTECTED GROUNDS

Protected grounds are characteristics listed under human rights law with which people identify (eg: race, colour, gender, age, religion, etc.). It is against the law to treat someone in a negative way based on any of these characteristics. Even if the negative treatment, or negative effect, is unintentional, it can still be considered discrimination under human rights protections. All individuals in Ontario have the legal right to be treated equally.

Why are they protected?

Certain individuals and communities have experienced current and historical barriers to full and fair participation in society as a result of discrimination based on these personal characteristics. This can have serious and long-term negative effects on social and economic stability, as well as lasting trauma and injury to human dignity that has repercussions over generations. Prohibiting discrimination based on these grounds fosters a society that values diversity and ensures that everyone has opportunities to participate and contribute equally.

Many people identify with more than one of these grounds and may experience multiple forms of discrimination at the same time.

Protected Grounds

The Ontario *Human Rights Code* recognizes the following grounds:

- Age
- Ancestry, colour, race
- Citizenship
- Ethnic origin
- Place of origin
- Creed
- Disability

Disability refers to a medical condition that a person has. This can be a temporary or long-term condition. It includes physical, mental, cognitive and learning disabilities, mental disorders, hearing or vision disabilities, epilepsy, drug and alcohol dependencies, environmental sensitivities, and other conditions.

- Family status
- Marital status (including single status)

- Gender identity, gender expression
- Receipt of public assistance (in housing only)
- Record of offences (“criminal record” – in employment only)
- Sex (including pregnancy and breast feeding)
- Sexual orientation

The *Charter* lists the following characteristics, known as “enumerated grounds”:

- Race
- National or ethnic origin
- Colour
- Religion
- Sex
- Age
- Mental or physical disability

Under the *Charter*, courts can also recognize additional unlisted characteristics known as “analogous grounds”. Analogous grounds are personal characteristics that, like enumerated grounds, are “immutable, difficult to change, or changeable only at unacceptable personal cost.” Once a court recognizes an analogous ground, it functions in the same way as any of the enumerated grounds and can form the basis of future equality claims.

To date, the Supreme Court of Canada has recognized four analogous grounds:

- Citizenship
- Sexual orientation
- Marital status
- Aboriginality-residence (discrimination against First Nations people on the basis that they live off-reserve)

While “economic status” such as poverty or homelessness are not a recognized ground, many people who are low-income or living in poverty experience disadvantage, which often intersects with protected grounds.

INDIRECT IMPACT:

Where a system, law or policy is equal or neutral on its face (i.e. not intended to be discriminatory), but in practice has a discriminatory effect, the system can violate human rights laws.

INTERSECTIONAL DISCRIMINATION:

- A person identified by multiple grounds may experience disadvantage that is compounded by the presence of each of the grounds.
- Based on their unique combination of identities, people may be exposed to particular forms of discrimination and may experience significant personal pain and social harm that come from such acts of discrimination. For example, a Jewish lesbian with a child and same-sex spouse can be seen as a “mother of a child” or a “Jewish woman” and would be protected under the grounds of marital status, family status, creed and sexual orientation. As a lesbian, this woman and her spouse may be exposed to forms of discrimination that other Jewish women with children are not.

16.

Have you tested or validated the AI system to see what factors it relies on? Does it rely on factors that correlate with a protected ground?

The first step of this question is to determine what factors the AI system relies on and how the AI system weighs those factors. The second step is to determine whether any of the factors or combination of factors correlate with a protected ground. A human rights expert can help advise on how to determine which factors may link to a protected ground.

This is often spoken about as “proxies”. Proxies that have been discovered in AI systems include postal code as a proxy for race where neighbourhoods have high concentrations of a particular racial group; gaps in work experience as a proxy for women because women are more likely to take time off for childcare; and playing competitive sports as a proxy for men because men are more likely to have played a competitive sport.

☐ YES ☐ NO ☐ DON'T KNOW

17.

Does the AI system assign characteristics to individuals based on proxies and other available data? Does the system produce outputs based on personal characteristics of individuals that are assumed, and not explicitly available in data? Does the technical system rely on a statistical model of human behaviour or personal characteristic?

Question 16 asks you to test/validate the AI system for the factors it is relying on and to consider whether these factors are proxies for protected grounds. This question asks you to consider whether the AI is attaching/assigning information to the individuals that is not explicitly in the data.

☐ YES ☐ NO ☐ DON'T KNOW

HOW TO GAUGE DISCRIMINATION: THERE IS NO CLEAR MEASUREMENT

- There is no universal measure for what level of statistical disparity is necessary to demonstrate disproportionate impact – the pattern must be significant and not just the result of chance.
- In high risk AI systems, there should be no discrimination or bias.
- It is not necessary for discrimination to affect all members of a protected group in the same way. For example, discriminating against pregnant women is discrimination against women even though not all women are pregnant.

RESULTS

Question 15

☐ YES ☐ NO ☐ DON'T KNOW

Question 16

☐ YES ☐ NO ☐ DON'T KNOW

Question 17

☐ YES ☐ NO ☐ DON'T KNOW

If you answer “yes” to questions 15, 16 or 17, your AI system displays disparate treatment on protected grounds. In this case, proceed to section 3B to determine whether the disparate treatment is legal discrimination.

If you are unable to answer questions 14, 15, 16 or 17, continue to questions 18-22.

If you answered “no” to questions 15, 16 and 17, your AI system is not displaying disparate treatment. In this case, proceed to section 4.

Assessing gaps in the system

18.

Are there gaps or limitations in your ability to meaningfully answer questions 14-17? If you are not able to test the AI system for differential treatment on protected grounds, make a record of the gaps and limitations in the data and go to question 19.

Be as clear and candid as possible which protected grounds you are and are not capable of testing. Note the quality and accuracy of your testing.

Can you test for intersectional discrimination? Can you assess whether individuals who identify on multiple protected grounds have a different outcome or treatment?

☐

YES

☐

NO

☐

DON'T KNOW

19.

What is the cause of those limitations? Are they surmountable?

In some cases, organizations cannot assess an AI system for discrimination based on race, sex, religion, age, etc. because the data does not include that information. If that is the case, consider solutions:

- Are there ways to assess the data without having the direct information? How reliable and accurate is this assessment?
- Can you start collecting the necessary data now and overcome this limitation in the future?

RESULTS

▶ **If your AI system is at high risk for human rights issues and you are unable to determine if its results are discriminatory, you are in a precarious position.**

Organizations in Ontario have an obligation to ensure that the products and services they provide do not violate human rights law.

▶ **In this case, organizations should consider taking the same steps they would if the AI system was demonstrating disparate treatment.**

▶ **If you cannot fully assess your AI system because you do not have the proper data to do so,** consider collecting the data going forward so you can correct this issue in the future.

▶ **Continue to Section 3B.**

PARAMOUNTCY OF HUMAN RIGHTS LAW

- Human rights law is constitutional (the *Charter*) or quasi-constitutional (Ontario *Human Rights Code*, *Canadian Human Rights Act*). This means human rights take priority over all other laws in Canada. If there is a conflict between human rights law and other laws, human rights take priority.
- Canadians have a positive obligation to ensure they are not providing services or products that violate human rights law. In other words, an organization is not shielded from human rights liability because an AI system's developers or administrators did not know the system was discriminatory.
- It is not a defence to a human rights violation to state that privacy laws prevented a government or private organization from knowing that it was violating human rights.



Part A Assessing Purpose and Impact

SECTION 3B

Is the Differential Treatment Permissible?

If your AI system is showing differential treatment on protected grounds (or if you cannot fully assess the system), questions 20-22 will help you determine whether the differential treatment is permissible.

Under Canadian law, there is no statistical or percentage amount of discrimination that is acceptable. Rather, there are certain contexts in which it is recognized that discrimination may be necessary (such as affirmative action) or (in much rarer cases) tolerable because it is necessary to achieve a greater goal or impossible to avoid.

Determining whether discrimination is justifiable can be complex. Questions 20, 21 and 22 consider whether the different treatment is permissible under human rights law. We recommend organizations seek guidance from human rights and legal experts when answering these questions.

20.

Is the purpose of the AI system, directly or indirectly, to advance a historically disadvantaged group?

Human rights legislation allows discrimination when it is created to correct the cumulative impacts of historical discrimination by advancing a historically disadvantaged group. This is sometimes called affirmative action or special programs. See the text box on this page for more information about affirmative action and special programs.

☐ YES ☐ NO ☐ DON'T KNOW

RESULTS

If the answer to question 20 is “yes”, the differential treatment is likely not discrimination.

If you do not know or are unsure of the answer to question 20, we recommend you consult a human rights expert who is knowledgeable about affirmative action or special programs. Exceptions to discrimination are rare; do not assume you fit into this exception unless it is very clear.

AFFIRMATIVE ACTION/ SPECIAL PROGRAMS

Affirmative action/special programs are designed to address the historic disadvantage that identifiable groups (racialized persons, women) have experienced by increasing their representation in employment and/or higher education.

See [Why are special programs protected? | Ontario Human Rights Commission \(ohrc.on.ca\)](https://www.ohrc.on.ca/en/why-are-special-programs-protected)

AMELIORATIVE EFFORTS

Human rights laws permit the development of policies and practices designed to address the discrimination, economic hardship, or disadvantage that groups may face based on protected grounds.

Under the *Code*, organizations and employers are permitted to create “special programs” to address these concerns. The Ontario Human Rights Commission encourages the development of special programs as an effective way to help reduce discrimination and address historical disadvantage. Similarly, section 15(2) of the *Charter* enables governments to proactively combat discrimination and assist disadvantaged groups by permitting programs that have an ameliorative or remedial purpose targeted at a disadvantaged group on an enumerated or analogous ground.

Examples of ameliorative or special programs include:

- A program designed to promote the hiring and advancement of women in tech professions.
- A social service organization that provides life-skills and counselling programs exclusively to its members who are refugees to Canada and have experienced trauma and abuse.

21.

If an individual or community is excluded from the results of the AI system, will it have a negative or adverse impact in their life? Or will being included in the results of the AI system have a negative or adverse impact on the individuals affected?

“Adverse impact” is an important but complicated principle in Canadian human rights law. The text box on this page explains adverse impact in more detail. Organizations are encouraged to seek legal advice to answer this question.

☐ YES ☐ NO ☐ DON'T KNOW

ADVERSE IMPACT

“Adverse effect discrimination” is an important concept in human rights law. It involves situations where a policy, rule or practice that seems to treat everyone equally has the opposite effect on a protected group under the *Charter* or human rights legislation, such as the *Ontario Human Rights Code* or the *Canadian Human Rights Act*. For example, a work schedule requiring work on Friday evenings for all employees might have a negative effect on employees with religious observations at that time. This type of unintentional discrimination is also called “constructive” or “indirect” discrimination. The workplace policy, rule or practice has the effect of unintentionally singling out particular people and results in unequal, differential or negative treatment on the basis of a protected ground.

RESULTS

If the answer to question 21 is “no”, the differential treatment is likely not discrimination.

If you do not know or are unsure of the answer, we recommend you consult a human rights expert who is knowledgeable about adverse impact. Exceptions to discrimination are rare; do not assume you fit into this exception unless it is very clear.

22.

Is there a justifiable reason for why the system is showing differential treatment?

“Justifiable reason” is an important but complicated principle in Canadian human rights law. The text box on this page explains justifiable reasons in more detail. Organizations are encouraged to seek legal advice to answer this question.

☐ YES ☐ NO ☐ DON'T KNOW

RESULTS

► If the answer to question 22 is “no”, the differential treatment is likely not discrimination.

If you do not know or are unsure of the answer, we recommend you consult a human rights expert who is knowledgeable about justifiable reasons for discrimination. Exceptions to discrimination are rare; do not assume you fit into this exception unless it is very clear.

WHAT IS A “JUSTIFIABLE REASON” for using an AI system that appears to be discriminatory?

It is only permissible to use a system that appears to be discriminatory in very rare circumstances. Those circumstances must be based on a “justifiable reason” that is supported by evidence that using an alternative system, or not using the system at all, would cause an organization undue hardship based on factors related to health, safety or cost. It is important to remember that some hardship is acceptable, and the size, resources, nature, and structure of an organization are factors in determining whether the threshold has been met. Legal counsel should be consulted for guidance on determining a “justifiable reason”.

Questions to ask in trying to determine whether the use of the AI system can be justified are:

- Is the AI system meeting its intended objective(s)?
- Is the need/requirement to use the AI system pressing enough to outweigh the negative impact of the AI system?
- What steps have been taken to minimize or address the human rights harms? Have you done everything possible?
 - Have alternative non-AI systems been considered?
 - Could a different AI system or a non-AI solution achieve the objective?
 - Was the system designed to minimize the harm on those it will impact?
 - Were accommodations sought?
 - What evidence exists that the organization would face hardship if it adopted alternative means or accommodated those who are harmed?



Part A
Assessing Purpose and Impact

SECTION 4

Does the AI system consider accommodation?

Questions 22-26 help determine whether the AI system considers specific communities and makes necessary accommodations.

23.

Is the AI system equally available, accessible, and relevant to all parties? Have the rights or needs of communities represented under all protected grounds been considered in the creation of the AI system?

There is a legal obligation to provide accommodation to individuals or communities who require it.

The text box on this page discusses human rights accommodation in more detail.

ACCOMMODATION

- Under the Ontario *Human Rights Code*, people identified by protected grounds are entitled to the same opportunities and benefits as everybody else. In some cases, they may need special arrangements or “accommodations” to take part equally in the social areas the *Code* covers, such as employment, housing and education.
- For example, where an AI-powered app is used to triage clients based on urgency of need, it may be necessary to provide an alternative method for those unable to use the app for reasons related to a protected ground of discrimination, such as physical disability, cultural or religious reasons.
- Employers, service providers and other duty holders have a legal obligation to accommodate *Code*-identified needs unless they can prove it would cause them undue hardship. Undue hardship is based on cost, outside sources of funding and health and safety factors.

24.

Have you tested the accessibility and availability of the AI system with diverse populations to ensure that it is accessible to all parties?

AI testing should include a wide variety of disabilities including hearing, sight, and mobility. Testing should also consider cultural, linguistic, religious, racial and gender differences.

26.

Have you put in place processes to test and monitor the AI system during development, deployment and use phases to uncover potential harm to children?

25.

Does the AI system respect the rights of children and take their best interests into account?

RESULTS

- ▶ If you have not considered how different populations might need to be accommodated, and addressed accommodation needs, you may be in violation of human rights obligations.
- ▶ If your AI system will impact children or be used by children, you will need to consider potential harms and mitigation measures.



Part A

Assessing Purpose and Impact

SECTION 5

Results

Below are six categories. After answering questions 1-26 your AI system will fit into one of the six categories. Follow the guidelines to see which category your AI system fits into.

Category I has little to no issues with human rights. This category requires the least amount of mitigation. As the categories go higher in numbers, the concern with human rights rise correspondingly.

Category VI is the highest concern. It requires the most intervention and may require organizations to reconsider whether an AI system is suitable in the context.

We recommend reviewing your answers with a human rights expert to ensure that you have categorized your AI system accurately.

I. Not High Risk

To be in this category, you answered:

- question 5 “yes” or “no”, and “no” to all of questions 6, 7, 8, 9, 10 and 13

Your AI system is not being used in a context where human rights are likely to be an issue. As such, it is unlikely that your AI system will have human rights concerns. We encourage you to revisit the human rights assessment if there is a meaningful change to the AI system, and annually to address any potential AI drift.

Question 5

☐ YES ☐ NO ☐ DON'T KNOW

Question 6

☐ YES ☐ NO ☐ DON'T KNOW

Question 7

☐ YES ☐ NO ☐ DON'T KNOW

Question 8

☐ YES ☐ NO ☐ DON'T KNOW

Question 9

☐ YES ☐ NO ☐ DON'T KNOW

Question 10

☐ YES ☐ NO ☐ DON'T KNOW

Question 13

☐ YES ☐ NO ☐ DON'T KNOW

II. High Risk and low stakes

To be in this category, you answered:

- question 5 “no”, and “yes” to one or more of questions 6, 7, 8, 9, or 13.

Your AI system is being used in a context where human rights could be an issue, but the AI system is not being used in a significant way. We encourage the following steps:

- Monitor the AI system for drift or change;
- Monitor the AI system for its subsequent or ongoing use to ensure it continues to not be used in a significant way; and
- Re-assess the AI system if there is a material change, and annually.

Question 5

☐ YES ☐ NO ☐ DON'T KNOW

Question 6

☐ YES ☐ NO ☐ DON'T KNOW

Question 7

☐ YES ☐ NO ☐ DON'T KNOW

Question 8

☐ YES ☐ NO ☐ DON'T KNOW

Question 9

☐ YES ☐ NO ☐ DON'T KNOW

Question 13

☐ YES ☐ NO ☐ DON'T KNOW

III. High Risk and no differential treatment

To be in this category, you answered:

- question 5 “yes” and “yes” to one or more of questions 6, 7, 8, 9, 10 or 13; and
- questions 15-17 “no”; and
- questions 20, 21 and 22 were not applicable

Your AI system is being used in a context where human rights issues may arise. However, currently it is not displaying differential treatment. We encourage the following steps:

- Ensure that your auditing/testing and validating is thorough, accurate and reliable;
- Continue to test your system regularly for differential treatment; and
- Implement mitigation strategies listed in Part B, including: internal procedure for assessing human rights; transparency and data quality; consultations, metrics testing, and de-biasing.

Question 5

☐ YES ☐ NO ☐ DON'T KNOW

Question 6

☐ YES ☐ NO ☐ DON'T KNOW

Question 7

☐ YES ☐ NO ☐ DON'T KNOW

Question 8

☐ YES ☐ NO ☐ DON'T KNOW

Question 9

☐ YES ☐ NO ☐ DON'T KNOW

Question 10

☐ YES ☐ NO ☐ DON'T KNOW

Question 13

☐ YES ☐ NO ☐ DON'T KNOW

Question 15

☐ YES ☐ NO ☐ DON'T KNOW

Question 16

☐ YES ☐ NO ☐ DON'T KNOW

Question 17

☐ YES ☐ NO ☐ DON'T KNOW

Question 20

☐ YES ☐ NO ☐ DON'T KNOW

Question 21

☐ YES ☐ NO ☐ DON'T KNOW

Question 22

☐ YES ☐ NO ☐ DON'T KNOW

IV. High Risk and cannot assess whether there is differential treatment

To be in this category, you answered:

- question 5 “yes” and “yes” to one or more of questions 6, 7, 8, 9, 10 or 13; and
- question 18 “yes”: you are unable to properly assess differential treatment in your AI system.

Your AI system is operating in a context where human rights issues may arise and you currently cannot assess whether it has differential treatment. Under the Ontario *Human Rights Code*, you are obligated to ensure that the products and services you provide do not violate human rights law; since you cannot do that, we encourage the following steps:

- Follow the steps outlined in category IV as if there were differential treatment (unless you answered “yes” to one of questions 20, 21 or 22, in which case go to category V); and
- Take steps to try to correct and improve your assessments.

Question 5

☐ YES ☐ NO ☐ DON'T KNOW

Question 6

☐ YES ☐ NO ☐ DON'T KNOW

Question 7

☐ YES ☐ NO ☐ DON'T KNOW

Question 8

☐ YES ☐ NO ☐ DON'T KNOW

Question 9

☐ YES ☐ NO ☐ DON'T KNOW

Question 10

☐ YES ☐ NO ☐ DON'T KNOW

Question 13

☐ YES ☐ NO ☐ DON'T KNOW

Question 18

☐ YES ☐ NO ☐ DON'T KNOW

Question 20

☐ YES ☐ NO ☐ DON'T KNOW

Question 21

☐ YES ☐ NO ☐ DON'T KNOW

Question 22

☐ YES ☐ NO ☐ DON'T KNOW

V. High Risk and differential treatment that fits into an exception *(fitting into an exception is rare, and you should have legal advice and human rights experts involved in assessing this issue)*

To be in this category, you answered:

- question 5 “yes” and one or more of questions 6, 7, 8, 9, 10 or 13 “yes”; and
- questions 15-17 – one or more “yes”; and
- questions 18-19 were not applicable; and
- “yes” to one of questions 20, 21, 22

Your AI system is operating in a context where human rights issues exist and your system is shown to have differential treatment on protected grounds. However, the purpose of your system fits into an exception to discrimination. We encourage the following steps:

- Continue to audit/validate/test the AI system routinely;
- Monitor the AI system for drift or change; and
- Follow the mitigation steps in Part B, especially:
 - Section 1 – internal procedure for assessing human rights
 - Section 2 – explainability and data quality
 - Section 3 – consultations

Question 5

☐ YES ☐ NO ☐ DON'T KNOW

Question 6

☐ YES ☐ NO ☐ DON'T KNOW

Question 7

☐ YES ☐ NO ☐ DON'T KNOW

Question 8

☐ YES ☐ NO ☐ DON'T KNOW

Question 9

☐ YES ☐ NO ☐ DON'T KNOW

Question 10

☐ YES ☐ NO ☐ DON'T KNOW

Question 13

☐ YES ☐ NO ☐ DON'T KNOW

Question 18

☐ YES ☐ NO ☐ DON'T KNOW

Question 20

☐ YES ☐ NO ☐ DON'T KNOW

Question 21

☐ YES ☐ NO ☐ DON'T KNOW

Question 22

☐ YES ☐ NO ☐ DON'T KNOW

VI. High Risk and differential treatment that is not justified or acceptable or you are questioning whether it is justifiable

To be in this category, you answered:

- question 5 “yes” and one or more of questions 6, 7, 8, 9, 10 or 13 “yes”; and
- questions 15-17 one or more “yes”; and
- questions 18-19 were not applicable; and
- questions 20, 21 and 22 all “no”.

Your AI system is being used in an area where human rights are a concern, it is showing differential treatment and there is no justification to suggest that the differential treatment is acceptable.

This is the most concerning category. We encourage the following:

- Consider limiting the use of this AI system or abandoning its use altogether;
- Closely implement all steps in Part B – Mitigation, questions 1-35;
- Continue to assess and review the AI system routinely; and
- Test and assess the AI system to gauge whether and how well it is achieving its objective(s).

Question 5

☐ YES ☐ NO ☐ DON'T KNOW

Question 6

☐ YES ☐ NO ☐ DON'T KNOW

Question 7

☐ YES ☐ NO ☐ DON'T KNOW

Question 8

☐ YES ☐ NO ☐ DON'T KNOW

Question 9

☐ YES ☐ NO ☐ DON'T KNOW

Question 10

☐ YES ☐ NO ☐ DON'T KNOW

Question 13

☐ YES ☐ NO ☐ DON'T KNOW

Question 15

☐ YES ☐ NO ☐ DON'T KNOW

Question 16

☐ YES ☐ NO ☐ DON'T KNOW

Question 17

☐ YES ☐ NO ☐ DON'T KNOW

Question 18

☐ YES ☐ NO ☐ DON'T KNOW

Question 20

☐ YES ☐ NO ☐ DON'T KNOW

Question 21

☐ YES ☐ NO ☐ DON'T KNOW

Question 22

☐ YES ☐ NO ☐ DON'T KNOW



PART B:

Mitigation Strategies

Part A is an assessment of whether an AI system presents human rights implications, including whether a system is at high risk for human rights harms.

Part B of this assessment is about minimizing human rights risks. Part B is divided into four elements:

- Internal procedures for assessing human rights
- Explainability, disclosure and data quality
- Consultations
- Testing and review

Each element is an important component of human rights risk mitigation.



Part B
Mitigation Strategies

SECTION 1: Internal Procedure for Assessing Human Rights

This section encourages organizations to develop an internal human rights review system. An AI system should be assessed for human rights issues throughout its life cycle. This section further encourages organizations, especially large ones, to establish lines of communication between different departments and areas of expertise.

Questions 1-6 encourage organizations to establish multidisciplinary internal procedures to assess AI-related human rights concerns. If this assessment demonstrates that your AI system is at high risk for human rights violations, your internal procedures should pay constant and close attention to human rights issues.

1.

Have you created a process to review and assess human rights regularly throughout the lifecycle of the AI system?

2.

What stage of the AI lifecycle are you currently in?

- a. Early concept/brainstorming
- b. Early design
- c. Data collection
- d. Data review
- e. Testing/auditing
- f. Deployment
- g. Maintenance
- h. Other

Human rights assessments should occur regularly and throughout an AI system's life cycle.

3.

i) How often will your team meet to review and assess human rights for this AI system?

The frequency of a human rights assessment depends on how the system is used and how it is designed. If your AI system is at high risk for human rights harms and shows disparate treatment, human rights assessments should occur frequently. Further, large language models and complex neural networks that are fed new data regularly will require recurring monitoring and assessment.

ii) Who is included in the AI team?

Human rights AI assessment should be multidisciplinary. The list below includes examples and suggestions of who should be included in discussions and reviews of human rights issues. This list is not exhaustive, nor is it necessary that every person attend every meeting. Organizations need to determine their own internal process to address the human rights concerns in their AI systems. If an AI system is at high risk for human rights issues and is shown to have disparate treatment, the organization will want to include as many perspectives as possible from the list below, particularly from groups that may be impacted by the AI system.

- a. Data scientists
- b. Lawyers
- c. Human rights experts
- d. Community members
- e. Procurement officers
- f. Front-end staff
- g. Policymakers
- h. Senior management
- i. Other

4.

If a human rights issue is flagged during the assessment, who should be informed? Who has the knowledge and authority to assess, address, and mitigate the issue?

This question is to encourage organizations to have established lines of communication, authority and responsibility.

- a. Senior management
- b. Client who has purchased or is adopting the tool
- c. Data scientists developing the tool
- d. Front-end staff deploying the tool
- e. Human rights experts or lawyers
- f. Other

5.

Who is overseeing the completion of the human rights assessment to ensure it is handled thoughtfully and completely?

Organizations should establish accountabilities for AI human rights issues and assessments.

6.

Are individuals encouraged to flag human rights issues without concern for repercussions?

- a. Do you provide a safe space for individuals or groups to raise issues about the AI system?
- b. Do you have protections for whistleblowers?



Part B
Mitigation Strategies

SECTION 2:

Disclosure, Data Quality and Explainability

Questions 7-9 ask about transparency and disclosure of AI systems. Transparency and disclosure about the existence of an AI system, what it is being used for, and how it works are crucial elements of human rights protection and legal accountability.

Questions 10-19 ask about AI outputs, data accuracy, and data reliability. These questions encourage developers to consider the link between data and the communities who may be affected by an AI system.

Finally, question 20 asks about explainability models. Explainability — the process of making an AI system or decision comprehensible to humans — can assist in assessing the justification or reasonableness of a specific AI outcome. Understanding how or why an AI system produced its results may, in some cases, be necessary to meet legal obligations under human rights or administrative law.

Disclosure

7.

What steps have you taken to inform the party (individual/community) impacted by the AI system about the AI system's use?

- a. Where is the information about the existence of the AI system available?
- b. How is information about the AI system communicated to the impacted community?
- c. What steps have you taken to ensure that the community impacted by the AI system is receiving and understanding the information about its use?

8.

What information about the AI system have you made available to the public?

- a. Algorithm, source code, software? Are these available in plain language?
- b. Data set details including source of data, purpose for collection of data, timeline of collection of data, updates to data, whether or why you use synthetic data, information about training data.
- c. List of factors that the AI system uses and how they are weighted.
- d. Explainability models and understanding of how the AI system is operating.
- e. Thresholds and data used to determine labels for scoring.

9.

What part of your human rights AI accountability measures have you made public?

- a. A summary or the completed answers to this AI human rights assessment.
- b. Whether or not your AI system was identified as a high risk system.
- c. Your internal process/measures taken to oversee and address human rights issues.
- d. Steps taken to validate, test, and monitor the AI system, validation criteria and results.
- e. Consultation efforts, metrics testing, de-biasing.

Data

10.

Have you reviewed, audited, or validated the data for accuracy, completeness, and relevance?

- a. Does the data include information about communities impacted by the AI system? (Eg: if older adults will be impacted by the AI system, is the data representative of older adults?)
- b. Is the data being relied on a fair representation of the parties who will be impacted by the AI system? (Eg: if a company is using AI to flag resumes to assist in deciding who to interview, if the AI system relies on data from existing or previously successful candidates and employees, and the corporation is dominated by white males, then the AI system will train itself to find candidates who resemble the existing employees.)
- c. If you are using synthetic data, are you able to validate it thoroughly?

11.

Is your dataset representative of everyone impacted by or intended to be served by the AI system? If training data is used, is the training data representative for the context in which the AI system will be used? If synthetic data is used, is the real-world data upon which it relies representative of everyone impacted or intended to be served by the AI system?

Review the dataset to ensure that it is fair, equitable, and most importantly, representative.

12.

Is the quality and reliability of the data used in the development (eg: training, testing and validation) sufficient for the intended application of the AI system?

13.

Are there biases or assumptions embedded in the data that increase the likelihood of discriminatory outcomes? What are they?

14.

If you are using synthetic data in an effort to correct for biases in real-world data, what methods have you deployed to ensure diversity, variety and fair representation in the synthetic data?

- a. Is synthetic data appropriate for this particular AI system?
- b. Have you evaluated the quality of the synthetic dataset?
- c. Are you able to validate the synthetic dataset to ensure accuracy and fairness?
- d. Have you identified any biases in the synthetic data?
- e. Have you measured the synthetic datasets against real-world datasets?

15.

Have you considered where, how and by whom the data was collected? Was it for a purpose different than to be used in the AI system under review?

16.

Is the data updated regularly to account for changes to the community the AI serves? If you are using synthetic data, are you monitoring the real world data it relies upon?

17.

Have risks associated with changing data quality and potential data drift been identified?

18.

Do you have a privacy compliance officer? Or have you consulted with a privacy expert?

19.

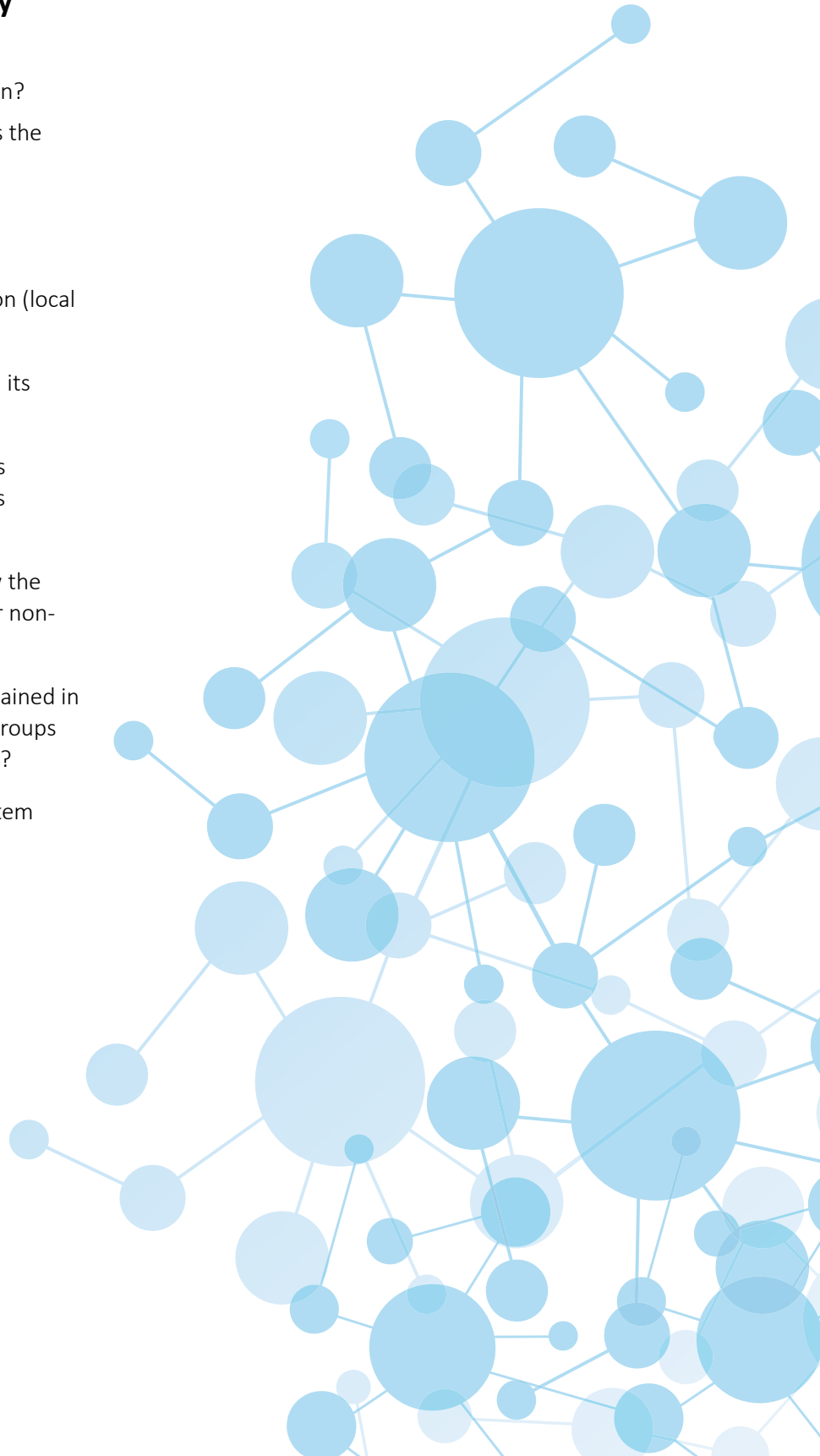
Have you complied with all privacy requirements and legislation?

Explainability

20.

Have you created an explainability model for this system?

- a. What does the explainability model explain?
 - i. Whether the AI system is performing as the designers intended
 - ii. Rank factors in level of significance
 - iii. Other
- b. Does it explain a specific AI system decision (local explainability)?
- c. Does it explain the functioning of the AI in its entirety (general explainability)?
- d. Is the AI system interpretable? Can parties understand how an AI system produces its prediction or recommendation?
- e. Can the creator/designer understand how the AI system works and explain it in a way for non-technologists to understand?
- f. Can the operation of the AI system be explained in a sufficiently understandable manner for groups impacted by the outcome of the AI system?
- g. Does it keep track/records of how the system works?





Part B
Mitigation Strategies

SECTION 3:

Consultations

Engaging people who are likely to be impacted by an AI system in the design, purpose and end use of the AI system can help address unexpected and unintended consequences and applications of an AI system.

Questions 21-26 ask you to consider consultations. For consultations to be effective and meaningful, they must include sufficient outreach, time and resources to respond, disclosure about the AI, and education on AI literacy.

21.

Did you consult with a diverse cross-section of parties and of expertise on this tool prior to and during development and design?

24.

Did parties have sufficient time and opportunity to engage with the information and provide meaningful feedback?

22.

Did you consult with communities likely to be impacted by this tool? Did your consultations include members of the communities that may be adversely or disproportionately impacted by the tool?

25.

Did you have a process in place for recording, reporting, and implementing feedback received during the consultations?

23.

Did your consultation process include educational opportunities about AI generally and provide sufficient information about the specific AI system for parties to meaningfully engage with the AI system?

26.

What did you do with the feedback you received?



Part B
Mitigation Strategies

SECTION 4: Testing and Review

Questions 27-33 consider how the AI system is tested and reviewed. The form and frequency of testing and review of an AI system will depend on the type and use of the system. Any AI system that has been identified as a high risk system by this assessment should be tested and reviewed frequently.

If the AI system is adaptive, it should be assessed for human rights issues frequently. An adaptive system evolves constantly, and its potential impact on human rights is an ongoing concern.

This section addresses three important components of AI testing and review: AI auditing, metrics testing, and de-biasing.

Audit/Review

27.

Have you audited the AI system for discrimination against people based on protected grounds?

28.

Was the audit conducted by an independent third party?

29.

Have the results of the audit been reviewed and considered?

30.

Is there a plan in place for regular testing and auditing of the AI system for unintended consequences?

Metrics Testing

Metrics testing is an important part of assessing and understanding the results produced by an AI system. However, metrics testing is not a replacement for analyzing discrimination under human rights law. Organizations should consider the results of the metrics testing as separate and in addition to human rights analysis.

31.

i) What metric of fairness did you apply when measuring the outcome of the AI system?

- a. Demographic parity
- b. Equal opportunity
- c. Equal odds
- d. PPV – positive predictive value
- e. FPR parity – false positive rate
- f. NPV parity – negative predicted value
- g. Other

ii) Why did you choose this metric of fairness and why do you think it is appropriate?

iii) What were the results of the fairness assessment?

iv) Was there any reason why a fairness assessment might be compromised, incomplete or inaccurate? If so, why?

De-biasing

32.

Have you employed any de-biasing techniques?

33.

Have you tried red-teaming the AI system for bias and discrimination?



LAW COMMISSION OF ONTARIO
COMMISSION DU DROIT DE L'ONTARIO

Law Commission of Ontario

2032 Ignat Kaneff Building
Osgoode Hall Law School, York University
4700 Keele Street
Toronto, Ontario, Canada M3J 1P3



**Ontario Human
Rights Commission**
**Commission ontarienne des
droits de la personne**

Ontario Human Rights Commission

180 Dundas Street West, 9th Floor
Toronto, Ontario, Canada M7A 2G5