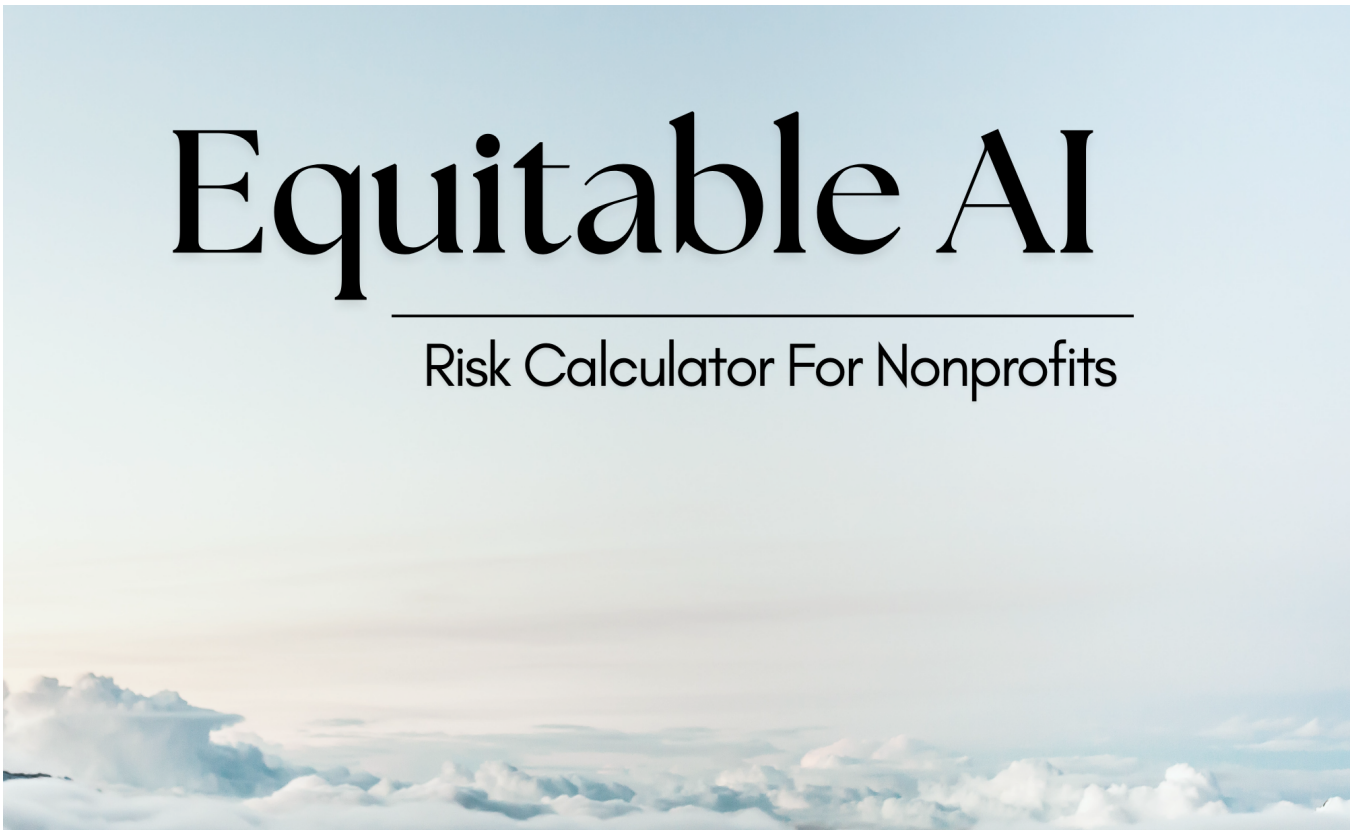
EQUITABLE AI FOR NONPROFITS



A practical toolkit for nonprofits to assess and mitigate AI-related equity risks.

INTRODUCTION 3
UNDERSTANDING AI EQUITY RISK 4
MAPPING AI USE AND RISK 6
THE AI EQUITY RISK MATRIX 10
HOW TO USE THIS FRAMEWORK 12
COMMON AI RELATED EQUITY RISKS 15

INTRODUCTION

In a world increasingly shaped by algorithms, AI is no longer just a tool-it is a decision-maker, a gatekeeper, and sometimes an unseen force influencing who gets access, who gets overlooked, and who thrives. For mission-driven organisations, this power brings unprecedented opportunity-but also unprecedented risk.

The AI Equity Risk Framework is a practical guide for nonprofits and social impact organisations to navigate this landscape responsibly. It is not a theoretical exercise in ethics; it is a hands-on toolkit designed to identify, assess, and mitigate AI-related equity risks before they become harm. From client intake to fundraising, from case management to communications, this framework centres the communities most affected by decisions made by automated systems.

Here, you will find not only a matrix to score risks but also concrete guidance to act, ensuring that AI amplifies your mission rather than undermines it. Each scenario, each checklist, and each mitigation strategy is rooted in real-world examples, offering clarity in a domain that can often feel opaque and technical.

This book is for those who refuse to accept that "innovation" alone is enough. It is for leaders who understand that accountability, fairness, and equity are not optional-they are mission-critical. Whether you are just beginning to explore AI in your organisation or seeking a more rigorous way to safeguard your communities, this framework provides a roadmap to act deliberately, ethically, and effectively.

The journey begins here.

UNDERSTANDING AI EQUITY RISK

Artificial Intelligence is often presented as neutral, objective, and infallible. The reality is quite the opposite. AI systems are built on human decisions about data, design, assumptions, and priorities. These decisions shape who benefits, who is harmed, and who is left invisible.

Every automated decision, from client intake to grant scoring, carries the potential to reinforce or disrupt these structures. Equity risk is not an abstract concept; it is about **real communities being affected by real decisions** every day.

For nonprofits, the stakes are particularly high: marginalised communities, already facing structural inequities, are often the first to feel the unintended consequences of automation.

Equity risk in AI is not a theoretical construct. It manifests in everyday organisational processes:

- A chatbot that fails to understand non-standard English or code-switching may discourage immigrant clients from accessing services.
- A grant review algorithm that favours well-resourced applicants can reinforce historical funding inequities.
- Donor targeting tools that optimise only for wealth may ignore or even alienate the very communities your mission seeks to serve.
- Predictive case management systems may amplify biases present in historical data, systematically deprioritising those most in need.

These are not hypothetical problems—they are practical examples of **how bias and inequity are encoded in AI workflows.** Understanding these risks requires a shift in perspective. It is not enough to ask, **"Does this AI work?"** We must ask, **"Who does this AI work for, and who does it leave behind?"**

Unfortunately equity risk compounds overtime and look different in different sectors. Below I talk about what equity risk means in practice, why it matters for nonprofits, and how to begin identifying it in your own organisation. The tools and strategies that follow are designed to make this understanding actionable.

This chapter lays the foundation for the AI Equity Risk Framework. It introduces key concepts such as:

- **Likelihood vs. Impact:** How often a risk may occur and how severe its consequences might be.
- **Structural Vulnerabilities:** Patterns of inequity embedded in society and historical data.
- **Community-Centric Assessment:** Centring the voices of those directly affected to validate and contextualise risk.

In any AI system, not all risks are created equal. Likelihood measures how often a particular harm or inequity might occur. Think of it as the frequency of exposure to a risk. Impact measures the severity of the consequences when that risk becomes a reality, ranging from minor inconvenience to systemic harm. Considering both together allows organisations to prioritise interventions, ensuring scarce resources target the most consequential threats.

For nonprofits, this means understanding not just whether an AI tool can misclassify someone, but how damaging that misclassification could be to the individual and the community they serve.

AI is a reflection the systems and histories in which it is built. These are structural vulnerabilities, I.e., patterns of inequity embedded in societal institutions, historical data, and operational norms. These include racial, economic, gender, or geographic biases that exist before any AI is deployed. Recognising structural vulnerabilities is essential, because AI can amplify them—even if unintentional. A risk assessment that ignores these patterns is fundamentally incomplete. We're past the stage of measuring symptoms, and must shift focus to the root causes.

While there's much talk about equity in AI, AI safety, alignment and such. Much work is being done on those fronts. Unfortunately, technical assessment alone cannot capture the lived realities of those affected. Centring the voices of communities ensures that risk evaluations reflect actual harm and context. This approach validates assumptions about likelihood and impact, uncovers risks that may be invisible to internal teams, and surfaces intersectional dynamics: how multiple marginalised identities compound potential harm. Engaging the community is not a box-check; it is an ethical imperative and a practical necessity to make risk scoring meaningful.

It's important to appreciate that identifying equity risks is not a one-time exercise. These risks evolve as your programs, communities, and AI tools change. The next step is to systematically map your AI uses, understanding where technology touches decisions that affect real people.

At this stage, focus on two complementary perspectives:

- Operational Lens: Where and how AI is deployed within your organisation (intake, case management, communications, fundraising, or grants.)
- Equity Lens: Who is impacted by these processes, and how might outcomes differ across communities with intersecting vulnerabilities?

By combining these lenses, you create a living map of potential points of harm or exclusion. This map becomes the backbone for the structured scoring, mitigation, and validation exercises that follow in the framework, ensuring that every assessment is both methodical and grounded in the lived realities of the communities you serve.

MAPPING AI USE AND RISK

With the foundation in place, the next step is to turn abstract concepts into actionable assessment. Before scoring risks or deciding mitigations, you need a clear picture of where AI touches your organisation and the communities you serve.

The first stage is identifying AI use across operational processes. Each touchpoint carries distinct equity risks, and the same tool can have different implications depending on context, user population, or deployment method.

Once AI uses are mapped, the framework layers two dimensions on each: likelihood (how often a risk may manifest) and impact (how severe the consequences could be). These dimensions are informed not only by historical data and operational realities but also by the voices of the communities affected.

By systematically cataloguing AI uses and associating each with potential risks, you create a risk map that guides prioritisation. High-likelihood, high-impact risks demand immediate attention, while low-likelihood, low-impact scenarios may be monitored.

By systematically cataloguing AI uses and associating each with potential risks, you map risks that guides prioritisation. High-likelihood, high-impact risks demand immediate attention, while low-likelihood, low-impact scenarios may be monitored. That's why I created the AI Equity Risk Framework:

1. Start by Identifying AI Use: Map where AI influences decisions, communications, and service delivery.
2. Define Specific Equity Risks: Describe how each AI use could disadvantage communities or reinforce inequities.
3. Score Each Risk: Assign likelihood and impact levels, informed by data and community insights.
4. Determine Mitigation Actions: Decide on monitoring, redesign, or immediate intervention.
5. Validate with Community: Test assumptions, severity ratings, and proposed actions with the affected populations.

With this structured approach, nonprofits can move from theory to practice, ensuring that AI is not just deployed, but deployed equitably and responsibly.

Before diving deep into the framework, answer the following three questions to see if a framework is ever required for your operations:

1. Does your organisation use AI tools? (Chatbots, grant screening, donor targeting, case management, automated communications, generative AI, etc.)
2. Do you serve marginalised or vulnerable communities? (Low-income families, immigrants, disabled individuals, BIPOC communities, LGBTQ+ youth, etc.)
3. Could your AI use create barriers or disadvantage certain groups?

If you answered YES to any of these → Keep reading.
If NO to all → You may not need this framework yet, but bookmark it for future reference.

This framework helps mission-driven organisations identify, evaluate, and mitigate equity risks in AI systems, centring the needs of marginalised communities.
Most AI risk frameworks focus on enterprise compliance, cybersecurity, or technical safety. This framework asks a different question: "Could this AI harm, exclude, or disadvantage the communities we exist to serve?"

Step 1: Identify Your AI Uses

List all AI tools you currently use or plan to use:

☐ Chatbots or intake forms
☐ Grant review or scoring systems
☐ Fundraising or donor targeting
☐ Case management or client prioritisation
☐ Communications or translation tools
☐ Other: _____

Step 2: Check for Red Flags

For each AI use, check any that apply:

CRITICAL RED FLAGS (Stop immediately if ANY of these apply):

☐ AI makes decisions about who gets life-impacting services (housing, healthcare, legal aid)
☐ AI automatically denies services without human review
☐ AI has never been tested with the communities we serve
☐ We don't know how the AI makes decisions
☐ The AI uses historical data that may reflect past discrimination

HIGH-RISK INDICATORS (Requires urgent attention):

☐ AI communication doesn't work for non-native English speakers
☐ AI screens or prioritises people based on past "success" data

☐ AI targets only wealthy donors, ignoring our community
☐ AI isn't accessible to disabled users
☐ AI uses client stories without explicit consent

MEDIUM-RISK INDICATORS (Plan to address within 60-90 days):
☐ AI sometimes produces stereotypical or insensitive content
☐ AI favours polished/professional writing over substance
☐ AI optimisation might dilute our equity message
☐ We haven't involved affected communities in AI design

Step 3: Take Immediate Action

If you checked ANY Critical Red Flags:

1. PAUSE deployment immediately
2. Alert leadership/board within 48 hours
3. Read the full framework below
4. Involve affected communities before proceeding

If you checked High-Risk Indicators:

1. Schedule mitigation planning within 30 days
2. Read relevant use cases in this framework
3. Add human oversight immediately
4. Begin community testing

If you checked Medium-Risk Indicators:

1. Document in your risk register
2. Review quarterly
3. Plan remediation within 60-90 days

Step 4: Calculate Risk Level

Find where Likelihood and Impact intersect on the matrix

Step 5: Take Action Based on Risk Level

Each action depends on the job function and type of risk. Feel free to refer to the ***Use cases*** to tailor to your organisation's mission.

Step 6: Validate with Community

Test your risk assessments with people from the communities you serve, especially for CRITICAL and HIGH risks. For what questions to ask and how to validate the work, please refer to the ***HOW TO USE THIS FRAMEWORK*** below**.**

THE AI EQUITY RISK MATRIX

Risk Levels Guide

Risk Level	Timeline	What to Do
Low	Quarterly review	Monitor and document; add light guardrails
Medium	60-90 days	Plan remediation; adjust design or process; add human review
High	30 days	Prioritize mitigations; formal checks, sign-offs, training; increase oversight
Critical	Immediate (2 weeks)	Pause deployment; escalate to leadership/board; involve communities in redesign or decision to abandon

Impact Scale Definitions

Level	Definition	Examples
1 – Minimal	Mild, reversible disadvantage for a small group	Awkward phrasing; minor misgendering; easily corrected errors
2 – Moderate	Measurable disadvantage for specific communities; creates barriers to access or quality	Campaign underrepresentation; communication gaps; slower response times
3 – Major	Reinforces existing structural inequities for marginalized groups	Systematic deprioritization; biased screening; inequitable resource allocation
4 – Severe	Systematically harms or excludes marginalized groups; mission-level breach	Service denial; rights violations; contradicts organizational equity commitments

The Matrix

	1 – Minimal Impact	2 – Moderate Impact	3 – Major Impact	4 – Severe Impact
4 - Frequent Observed in current operations	**MEDIUM** Everyday AI use subtly erodes inclusive language if unchecked	**HIGH** Most campaigns shaped by AI skew away from underserved communities	**CRITICAL** AI systems routinely reinforce inequitable access or burden marginalized clients	**CRITICAL** AI use frequently contradicts equity commitments; may constitute rights violations
3 - Likely Expected under current controls	**MEDIUM** AI routinely centres dominant narratives; equity lens is diluted	**HIGH** AI-generated communications regularly under-represent key communities	**CRITICAL** Case prioritization AI systematically gives slower support to certain groups	**CRITICAL** Program decisions driven by AI worsen existing inequities in outcomes
2 - Possible Plausible in pilots or scale-up	**LOW** AI sometimes misgenders or mislabels groups; staff usually catches it	**MEDIUM** Fundraising targeting creates barriers to access for specific communities	**HIGH** Screening AI occasionally deprioritizes applications from marginalized groups	**CRITICAL** AI-assisted decisions could deny life-impacting services to whole communities
1 - Rare Requires multiple control failures	**LOW** Minor wording bias in drafts; easily corrected by staff	**LOW** Occasional insensitive phrasing to small subgroup; quick fix	**MEDIUM** Rare but serious misclassification of vulnerable client	**HIGH** Rare but catastrophic AI-driven exclusion from critical service

HOW TO USE THIS FRAMEWORK

Before You Start

Identify the communities you serve and their specific vulnerabilities:

- Immigrants and refugees
- Low-income families
- Disabled individuals (physical, cognitive, sensory)
- LGBTQ+ youth and adults
- Racial and ethnic minorities
- Elderly populations
- People experiencing homelessness
- Survivors of trauma or violence
- Others: _____

Consider intersecting identities: A disabled, low-income immigrant faces compounded barriers. Always score based on the most severely impacted subgroup.

Step 1: List Your AI Uses

Create an inventory of all AI tools currently in use or under consideration:

AI Tool/System	Purpose	Department	Launch Date
Example: ChatGPT	Draft grant proposals	Development	Jan 2024
Example: Intake chatbot	Client screening	Programs	Pilot phase

Step 2: Define Specific Equity Risks

For each AI use, ask:

1. Could this disadvantage certain communities?
2. Could this reinforce existing inequities?
3. Does this affect some groups more than others?

Example:

- AI Use: Intake chatbot for housing services
- Communities Served: Low-income immigrants
- Specific Risk: "Chatbot doesn't recognise non-standard English or cultural communication styles, causing clients to abandon intake process"

Not specific enough: "Chatbot might have bias"
Specific and actionable: "Chatbot trained only on standard American English may misunderstand Caribbean English Creole syntax, causing Caribbean immigrant applicants to drop off"

Step 3: Score Each Risk

Likelihood (1-4): How likely is this to happen?

Score	Definition	When to Use
1 - Rare	Requires multiple control failures	We've tested extensively; multiple safeguards in place
2 - Possible	Plausible in pilots or scale-up	We're piloting; some controls exist but untested with target population
3 - Likely	Expected under current controls	We haven't tested with affected communities; known gaps exist
4 - Frequent	Observed in current operations	We're already seeing this happen; no effective controls in place

Impact (1-4): What's the equity impact on communities you serve?
Use the Impact Scale Definitions above. Always score based on the most severely impacted subgroup.

Step 4: Calculate Risk Level

Find where Likelihood and Impact intersect on the matrix:
Example:

- Likelihood: 3 (Likely) – Language barriers are common; chatbot hasn't been tested with this population
- Impact: 3 (Major) – Reinforces existing access barriers for already marginalised group
- Risk Level: CRITICAL (Likely + Major)

Step 5: Take Action Based on Risk Level

Risk Level	Timeline	Required Actions
Critical	Immediate (2 weeks)	• Pause deployment immediately • Escalate to leadership/board • Involve affected communities in redesign • Either fix or abandon
High	30 days	• Prioritize mitigations • Add formal checks and sign-offs • Increase oversight • Provide training

Medium	60-90 days	• Plan remediation • Adjust design or process • Add human review • Document in risk register
Low	Quarterly	• Monitor ongoing • Document in risk register • Add light guardrails

Step 6: Validate with Community

CRITICAL: Test your risk assessments with people from the communities you serve.

Questions to ask:

- Does this risk feel real to you?
- Is the severity rating accurate?
- What are we missing?
- What would make this AI tool work better for you?

How to do this:

- Focus groups with current clients
- Advisory panels with community members
- One-on-one interviews
- Community surveys
- Partnerships with community-based organisations

Don't skip this step. Those closest to potential harm are best positioned to identify and evaluate risks.

COMMON AI RELATED EQUITY RISKS

Before scoring your AI uses, review these common scenarios. Check any that apply to your context:

Client Intake & Screening

- Language/accent barriers in chatbots or forms
- Assumes literacy or digital fluency
- Not accessible to disabled users (screen readers, timing, navigation)
- Requires documentation marginalised groups lack (SSN, photo ID, fixed address, legal name)

Grant Review & Scoring

- Privileges polished writing over substance
- Favours large/professionalized organisations
- Uses location (ZIP code) as proxy for capacity or need
- Weights indicators that correlate with race and class

Fundraising & Communications

- Targets only wealthy donors, ignores community
- Uses client stories without consent
- Generates stereotypical imagery or language
- Optimises for engagement over message integrity

Case Management

- Priorities based on biased historical data
- Flags poverty/trauma responses as "risky"
- Automates denials without human review
- Increases surveillance of marginalised clients

Communications & Outreach

- Machine translation misses cultural context
- AI optimisation dilutes equity message
- Platform choices exclude digitally marginalised communities

USE CASE 1: Client Intake & Screening [Disclaimer: The scenarios under this use case are currently being tested]

Scenario 1.1: Language/Accent Barrier

ID: NP-INTAKE-001

Description: AI chatbot or intake form doesn't understand non-standard English, accents, or multilingual code-switching, causing clients to abandon process

Affected Communities: Immigrants, refugees, non-native English speakers, Deaf/HoH using ASL syntax

Typical Likelihood: 3-4 (Likely to Frequent) for organisations serving multilingual populations

Typical Impact: 3 (Major) - reinforces language-based access barriers

Real-World Example

Housing assistance chatbot trained on standard English couldn't parse Caribbean English Creole syntax, resulting in 40% drop-off rate for Caribbean immigrant applicants

Mitigation Strategies:

- Test with actual clients from target communities before deployment
- Offer human intake alternative prominently
- Train AI on diverse language patterns
- Add "switch to human" button on every screen

Scenario 1.2: Literacy/Digital Literacy Assumption

ID: NP-INTAKE-002

Description: AI intake assumes high literacy or digital fluency; complex language, jargon, or navigation requirements exclude low-literacy populations

Affected Communities: Low-literacy adults, elderly, people with cognitive disabilities, people experiencing homelessness

Typical Likelihood: 3 (Likely) if not tested with target population

Typical Impact: 3-4 (Major to Severe) depending on service criticality

Real-World Example:

Benefits enrolment AI used 10th-grade reading level and multi-step navigation; excluded 35% of eligible elderly clients with limited digital experience

Mitigation Strategies:

- Plain language testing (6th-grade level max)
- Voice/audio option
- Visual aids and icons
- Progress indicators and easy back navigation
- In-person assistance option

Scenario 1.3: Disability Accessibility Failure

ID: NP-INTAKE-003

Description: AI interface not compatible with screen readers, lacks alt text, requires mouse/touch, has timing constraints, or uses inaccessible CAPTCHA

Affected Communities: Blind/low-vision, Deaf/HoH, motor disabilities, cognitive disabilities

Typical Likelihood: 4 (Frequent) unless specifically designed for accessibility

Typical Impact: 4 (Severe) - systematically excludes disabled clients; may violate ADA

Real-World Example:

Mental health intake chatbot required clicking small buttons within 30 seconds; excluded clients with motor disabilities and anxiety

Mitigation Strategies:

- WCAG 2.1 AA compliance minimum
- Test with screen readers (JAWS, NVDA)
- No timing requirements or clear extensions
- Keyboard-only navigation
- Simple, accessible CAPTCHA alternatives

Scenario 1.4: Identity Documentation Bias

ID: NP-INTAKE-004

Description: AI requires or prioritizes documentation that marginalized groups lack (SSN, photo ID, fixed address, legal name matching gender identity)

Affected Communities: Undocumented immigrants, homeless individuals, trans/non-binary people, people leaving domestic violence, formerly incarcerated

Typical Likelihood: 3 (Likely) if using standard identity verification

Typical Impact: 4 (Severe) - excludes those who most need services

Real-World Example: Food bank AI required government ID and proof of address; excluded 60% of unhoused clients and undocumented immigrants

Mitigation Strategies:

- Accept multiple forms of identity verification
- Allow self-attestation with follow-up verification
- Support chosen names and pronouns
- Alternative address options (shelter, PO box, general delivery)

Disclaimer: All content on this site is original and belongs to Ms. Bhardwaj unless otherwise specified. The use of Artificial Intelligence (AI) is limited to grammar or spell check, optimising productivity, listing resources, data scraping or visualisations. Any use, reproduction, or distribution of this material without prior written consent is strictly prohibited.

The use cases and examples are based on real-world incidents, sector research, and practitioner feedback, and are pending studies with non-profits.

Developed by Bhavyatta Bhardwaj at Danuma

BHAVYATTA