

AI Bias: A Practical Guide for Nonprofits



Developed by the She Leads AI Community of Practice
for AI Governance | September 2026

Table of Contents

Note: This document is optimized for mobile viewing.

For the complete interactive experience, view on a desktop computer or within a dedicated PDF reader app. Interactive performance on mobile browsers or preview apps may vary.

- PRE-ASSESSMENT: Where Does Your Organization Stand? 3**
 - How To Use This Guide, By Role 7
- PART 1: What AI Bias Is and Why It Matters 8**
 - How AI Bias Works 9
 - Common Bias Types in Nonprofit Work 10
 - Harmful Patterns to Recognize in Nonprofit Content 11
- PART 2: Where AI Bias Shows Up (and What to Do About It) 12**
 - How to Take Action Now 12
 - Where Algorithmic Bias Shows Up in Your Work 13
 - Real Cases 15
 - Child Welfare Screening 16
 - Tenant Screening 17
 - Dialect Prejudice in Legal Aid 18
 - Benefits Intake 19
 - Immigration Services and Client Trust 20
 - What You Can Do About Algorithmic Bias 21
 - When Tools Make Decisions About People 24
- PART 3: Build Your Organization’s Response 26**
 - Practical Action Steps 26
 - AI Bias Conversation and Planning 27
 - Evaluate Your Tools 28
 - The Bias Section of Your AI Policy 30
 - When Bias Shows Up: Document and Respond 32
 - Keep It Current 33
 - Talking to Your Board 34
- Resource Links 38**



PRE-ASSESSMENT:

Where Does Your Organization Stand?

Before You Begin

Take ten minutes before you read the rest of this guide.

Your answers will tell you where your organization stands and which sections are most relevant to your work.





PART A: What Tools Are You Using?

1. Does your organization use any AI tools, including free tools like ChatGPT, Google AI features, or Canva’s AI, or AI built into platforms you already pay for, such as email systems, scheduling software, or management platforms?	☐ Yes ☐ Not sure ☐ No
2. Do you know which AI tools your staff are currently using, whether or not leadership approved them?	☐ Yes ☐ Somewhat ☐ No
3. Does your organization use any AI-assisted tools in your hiring, volunteer screening, or HR processes?	☐ Yes ☐ Not sure ☐ No
4. Does your organization use a chatbot or AI assistant for client intake, FAQ responses, or community engagement?	☐ Yes ☐ Considering it ☐ No
5. Does anyone on your team use AI to take notes, summarize conversations, or document interactions with the people you serve?	☐ Yes ☐ Sometimes ☐ No

PART A RESULTS

If you answered	Assessment	Recommendation
Not sure to questions 1 or 2:	Before you can address AI bias, you need a clear picture of what tools are actually in use.	- The team conversation in Part Three is your first step. - It maps what your staff is using before you build anything else around it.
Yes or Not sure to question 3:	Your hiring and HR tools carry bias risks documented in the research.	Include Part Three tools in your evaluation.
Yes or Sometimes to questions 4 or 5:	Those tools are in direct contact with the people you serve or their information.	Read Part Two in full before moving to Part Three.



PART B: Where Does AI Affect the People You Serve?

1. Does your team use AI to create public-facing content: newsletters, social media posts, grant narratives, imagery, or website copy?	☐ Yes ☐ Sometimes ☐ No
2. Does your organization use any AI-assisted tool that affects how clients access your services, including scheduling appointments, screening applicants, assessing eligibility, or routing people to programs?	☐ Yes ☐ Sometimes ☐ No
3. When AI generates content representing your community, including images, language, names, or stories, does the output actually reflect the people you serve?	☐ Yes ☐ Sometimes ☐ No
4. Do you use AI to help prepare grant applications, funder reports, or donor communications?	☐ Yes ☐ Sometimes ☐ No
5. Does your organization collect personal or sensitive information about the people you serve using digital tools? Do you know what those tools do with that data and who can access it?	☐ Yes ☐ Somewhat ☐ No/Not sure

PART B RESULTS: Count your “Yes” and “Sometimes” answers

Total:	Assessment	Recommendation
4 or more:	AI is affecting your community in multiple areas.	- In Part Two, prioritize the case studies closest to your work. - In Part Three, start with evaluating your tools, then build your policy.
2 to 3:	AI is present in some of your work.	- In Part Three, start with the policy questions. - Use the tool evaluation to review your highest-exposure areas.
0 to 1:	AI is not yet significantly affecting client-facing work.	Read through the guide to build your foundation before it does.



PART C: What Do You Have in Place?

1. Does your organization have any written guidelines, even informal ones, for how staff should evaluate AI tools and AI-generated outputs for potential bias?	☐ Yes ☐ In progress ☐ No
2. Does a human review AI-generated content before it goes to clients, donors, or the public?	☐ Yes ☐ Sometimes ☐ No
3. Has any staff member received training on what AI bias is and how to recognize it?	☐ Yes ☐ Some staff ☐ No
4. Has your board discussed your organization’s exposure to AI bias and what you have in place to address it?	☐ Yes ☐ It’s come up ☐ No

PART C RESULTS

Take Note:	Assessment	Recommendation
Where did you answer “No” or “In progress”?	These are your gaps.	- Part Three addresses each gap directly. - Keep your Part B and Part C assessment answers in mind as you read through the sections.



How To Use This Guide, By Role

If you are reading this on your own, use the table below to find your highest-priority sections.



Photo: Pexels | Fauxels

Role	START HERE ↓
ED/Leadership	■ <u>Part One</u>: bias types ■ <u>Part Three</u>: policy questions and board section
Program/ Direct Service Staff	■ <u>Part Two</u>: Real Cases, decision-making tools section
Communications/ Fundraising	■ <u>Part One</u>: harmful patterns ■ <u>Part Two</u>: content creation section and CLEAR framework
Finance/HR/ Operations	■ <u>Part One</u>: bias types ■ <u>Part Two</u>: decision-making tools section (hiring and HR)
All Staff	■ <u>Part One</u>: How AI Bias Works ■ <u>Pre-Assessment</u> worked through together



PART 1:

What AI Bias Is and Why It Matters

Bias in AI Tools

Take ten minutes before you read the rest of this guide.

Every AI tool used to create content, make decisions, or process information about people has bias built into it.

Not some tools. **All of them.**

This guide is about where that bias shows up in your organization's work and what you can do about it.





How AI Bias Works

AI tools learn from vast amounts of human-generated content: text, images, decisions, and data.

That content reflects the world as experienced and recorded by the people who created and curated it — a record that skews heavily toward dominant, mainstream cultures and leaves out or distorts the experiences of communities that have been historically excluded from those spaces.

The inequities in the data become part of how the tool operates.

TWO Mechanisms That Drive Most of the Bias That Nonprofits Encounter

1

Training Data Bias

- When the data a tool learned from does not represent all populations equally, its outputs reflect that gap.
- An image generator trained mostly on Western media defaults to certain faces, settings, and bodies.
- A language model trained on mainstream text produces communications that feel foreign or tokenistic to communities outside that mainstream.

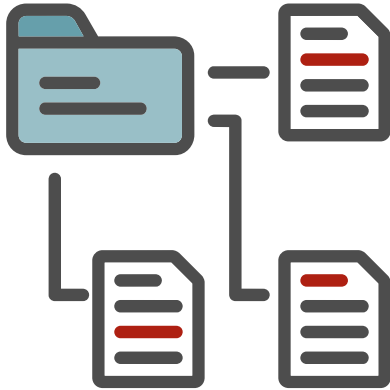
2

Proxy Variables

- Some tools use data points that appear neutral — zip code, name, spending history — but that correlate strongly with race, income, or other characteristics that should not determine outcomes.
- When a tool weights those inputs, it is functionally doing the same thing as directly factoring in race or income.
- The pattern of who lives in which zip code is itself the product of decades of discriminatory housing policy.
- When a tool treats zip code as predictive, it treats the outcome of past discrimination as a basis for future decisions about those same communities.
- This is how a healthcare algorithm can systematically recommend less care for Black patients without any intent to discriminate and without ever being programmed to consider race.



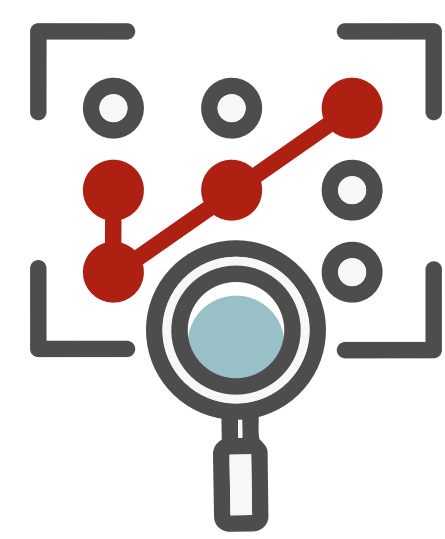
Common Bias Types in Nonprofit Work



Bias Type	What It Looks Like
Representation Bias	■ Certain communities are absent from outputs, imagery, or case examples ■ Conversely they are only present in narrow, stereotyped ways, including body types, ages, and appearances that don't reflect your community
Racial Bias	■ Communities of color are stereotyped or shown only in hardship contexts
Cultural Bias	■ One cultural norm treated as universal ■ There is local context missing
Socioeconomic Bias	■ Poverty framed as identity ■ Structural causes are absent
Disability Bias	■ People defined by disability ■ Individuals are shown as passive or inspirational rather than as full participants.
Automation Bias	■ Staff accept AI output without review because it sounds polished or authoritative. ■ Unlike the other types in this table, this is a human behavior rather than an AI output — but it is how biased outputs get used without being caught.



Harmful Patterns to Recognize in Nonprofit Content



Bias Type	What It Looks Like
Pity Language	- Words like helpless, voiceless, struggling - Removes dignity and agency from the people you serve
Savior Narratives	- Organization as hero, community as backdrop - Erases community leadership and self-determination
Overgeneralization	- Phrases like “African communities” or “the poor” - Erases individuality, specificity, and humanity
Trauma Framing	- Leading with suffering - Reduces people to their worst moments
Bias by Omission:	Whose voices are absent is as important as whose are present



PART 2:

Where AI Bias Shows Up

(and What to Do About It)

How to Take Action Now

You do not need a technical background to recognize AI bias or to take action when you see it.

This section covers where bias shows up in nonprofit work, what it has looked like in documented cases.

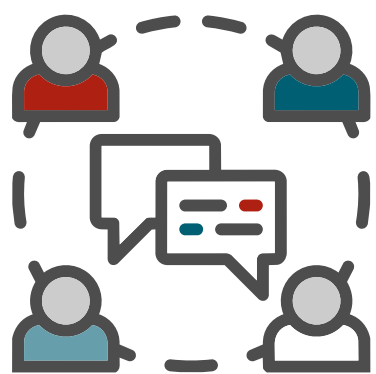
Learn what you can do in the two main contexts nonprofits encounter algorithmic bias:

- creating content
- using tools that make decisions about people





Where Algorithmic Bias Shows Up in Your Work



Content and Communications

Grant narratives, donor appeals, imagery, social media, and website copy

AI defaults to mainstream language patterns and aesthetics. For community-based nonprofits, this produces two specific problems:

- When generating imagery without specific direction, AI tools consistently default to representations that are whiter, wealthier, and more conventionally able-bodied than the communities most nonprofits serve. A prompt for “a successful professional” or “a family in need” will produce biased defaults unless you specify otherwise.
- When generating grant narratives without organization-specific context, AI defaults to the language patterns of larger, better-resourced organizations. When generating grant narratives without organization-specific context, AI tends to produce generic language, which can flatten the distinct voice that community-based organizations bring to their proposals.



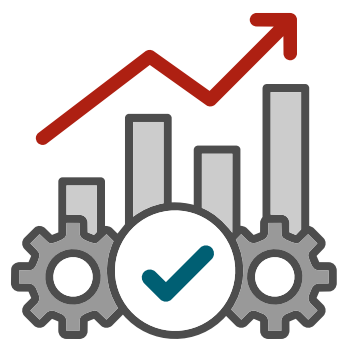
Client-Facing Tools

Intake and eligibility systems, chatbots, scheduling tools, case management platforms

- When a tool affects who gets access to your services, bias at this point directly affects your community.
- Many of these tools use proxy variables, including zip code, prior service history, and language patterns, without disclosing how those variables shape their outputs. This is the highest-risk area. See [Cases 1, 3, and 4](#) for documented examples.



Where Algorithmic Bias Shows Up in Your Work (continued)



Operations

Hiring tools, volunteer screening, HR platforms

- AI-assisted hiring tools carry documented bias in specific, traceable patterns. Tools trained on historical hiring data have penalized resumes that use language associated with women, including references to women's organizations, volunteer care giving roles, and other activities that appear in women's professional histories at higher rates than men's.
- Research consistently shows that identical résumés receive fewer callbacks when the applicant's name reads as Black versus white — a pattern that carries directly into AI-assisted screening because these tools learned from the same historically biased hiring decisions. (MIT Technology Review, 2018; Bertrand & Mullainathan, 2004)
- The tools you use to build your team carry the same risks as the tools you use to serve your community.



Data

What your digital tools collect about the people you serve, where that data goes, and who can access it

- Many AI-connected tools collect, store, and share data beyond what you authorized.
- When data about specific populations, including immigration status, income, or legal history, flows into systems that use it to make predictions or recommendations, that data can be used in ways that produce discriminatory outcomes.
- For organizations serving communities with particular reasons to distrust data collection, this is a mission-critical risk. See [**Case 5: Immigration Services and Client Trust.**](#)



Real Cases

These cases are documented in the sources linked in the following section. Read the examples closest to your organization’s work first.

Each one ends with what could have helped, because understanding what prevents harm is as important as understanding how it happens.



CASE 1: Child Welfare Screening

Child welfare case management at contracted nonprofit organizations



CASE 2: Tenant Screening

Legal aid and housing organizations helping low-income clients find housing



CASE 3: Dialect Prejudice in Legal Aid

Legal aid organizations and public defender offices



CASE 4: Benefits Intake

Direct service nonprofits operating intake and eligibility processes



CASE 5: Immigration Services and Client Trust

Nonprofits providing direct services to immigrant and mixed-status communities

CASE 1:

Child Welfare Screening

Setting: Child welfare case management at contracted nonprofit organizations

New York City’s child services agency used an algorithm to help caseworkers identify families at risk of child abuse.



- The algorithm relied on neighborhood-level data, including borough, zoning area, and community district, as proxies for risk. Families of color were flagged at disproportionately high rates compared to white families.
- Caseworkers were not told how the tool weighted its inputs or how it arrived at a score for any specific case. Families were not told a score had been generated about them. (The Markup, May 2025)

The Harm:	What Could Have Helped:
■ Being flagged triggered a CPS investigation: a formal process that can include home visits and documentation, and in the worst cases, family separation ■ Families of color bore this burden at disproportionate rates based on where they lived, with no knowledge that a score existed and no process to contest it.	■ Training caseworkers on how the algorithm works, what data it uses, and what its known limitations are, before deployment. ■ Establishing a documented override path so caseworkers can easily flag and update scores that contradict their direct family knowledge. ■ Auditing flagging rates regularly by race and income to catch and fix discriminatory patterns before they compound. ■ Ensuring any family affected by an algorithmic recommendation has a meaningful path to request human review of their case, not just notification that a score exists.

For Your Organization:

If you partner with a government agency to deliver services, you may be working downstream of tools you did not choose and were not told about. You have standing to ask your agency partners what tools are in use, how they work, and what oversight exists.

Photo: Pexels - Antonius Ferret

CASE 2:

Tenant Screening

Setting: Legal aid and housing organizations helping low-income clients find housing



- In November 2022 the Consumer Financial Protection Bureau published two reports on the tenant background check industry, drawing on more than **24,000 renter complaints**.
- One of the most common failures was screening companies matching records by name alone, without Social Security numbers or other identifying details. Applicants were denied housing over a record belonging to someone else who shared their name. The CFPB has stated that name-only matching is illegal under the Fair Credit Reporting Act.
- The agency found the industry had prioritized speed and convenience for landlords over accuracy for applicants. (Consumer Financial Protection Bureau, November 2022)

The Harm:	What Could Have Helped:
■ Qualified applicants were denied housing based on factors that correlated with race and income, not their fitness as tenants. ■ With no disclosure of algorithmic involvement, most had no path to appeal.	■ Before referring clients to a property, asking what screening tool the landlord uses and whether performance data across demographic groups is available. ■ Training staff to recognize when a rejection may be algorithm-driven: unexplained denials or identical results across multiple qualified applicants are signals. ■ Tracking denial patterns across clients and properties; a repeated pattern across the same tool is worth escalating to legal advocacy partners.

For Your Organization:

If your organization helps clients navigate application processes, unseen algorithms may create barriers you’re not aware of. Understanding these decision-making tools upfront allows you to better advocate for your clients.

Photo: Pexels - Kampus Production

CASE 3:

Dialect Prejudice in Legal Aid

Setting: Legal aid organizations and public defender offices

Researchers publishing in Nature found that large language models demonstrate dialect prejudice.



- Given identical content written in African American English versus standard American English, the models consistently produced more negative judgments about African American English speakers, including assigning them to lower-prestige jobs and recommending the death penalty more often in hypothetical criminal scenarios.
- The prejudice was covert. The same models, asked directly about race, produced positive and egalitarian statements. It appeared only in what they did, not in what they said. (Hofmann et al., Nature, 2024)

The Harm:	What Could Have Helped:
■ Bias in AI recommendations shapes what a decision-maker sees first, anchoring their thinking before any human review. ■ For clients whose liberty is at stake, that anchoring effect can be decisive	■ Before adopting any AI tool for client-related work, test it using language that reflects how your clients actually communicate, not only formal written English. ■ Establishing a policy that AI-generated recommendations are never used in case decisions without human review and sign-off. ■ Add dialect and language patterns to your bias review process alongside race, income, and geography.

For Your Organization:

Any AI tool that processes written or spoken language from the people you serve is processing the way those people communicate. If your community uses language patterns outside the mainstream, test for that before the tool goes live.

Photo: Pexels - Pavel Danilyuk

CASE 4:

Benefits Intake

Setting: Direct service nonprofits operating intake and eligibility processes

Michigan’s automated unemployment system, known as MiDAS, issued fraud determinations without human review beginning in 2013.



- The state’s Auditor General later reviewed **22,000** of those determinations and found that **93%** did not actually involve fraud.
- Roughly **40,000 residents** were falsely accused. A 2017 settlement required the state to stop making automated determinations without human review, and the litigation concluded with a **\$20 million settlement** in 2024. (Benefits Tech Advocacy Hub)

The Harm:

- For people applying for food assistance, housing support, or other critical resources, a wrongful denial is not a paperwork problem.
- In documented cases, the AI replicated the gatekeeping behavior of the bureaucracies it was meant to replace, and did so faster and with less possibility for the human judgment that sometimes makes exceptions.

What Could Have Helped:

- Building a clear path to human review into the intake process before any AI-assisted tool goes live, not after a problem surfaces.
- Telling people clearly when AI is involved in their intake process and give them a way to request human review of their case.
- Piloting new intake tools before full deployment, testing denial rate patterns by income, race, and primary language.

For Your Organization:

A process that moves more quickly may be turning people away faster. If AI is involved in any part of how community members enter your programs, you need a human in the loop.

Photo: Pexels - Fauxels

CASE 5:

Immigration Services and Client Trust



Setting: *Nonprofits providing direct services to immigrant and mixed-status communities*

- The Urban Institute’s December 2025 Well-Being and Basic Needs Survey found that roughly **9% of adults** in immigrant families avoided safety net programs out of concern that their personal information would be shared with immigration authorities. (Urban Institute, April 2026)
- Reporting in Nonprofit Quarterly describes how agencies traditionally focused on social services, including the IRS, Medicaid, and HUD, have become sources of immigration enforcement data, and how nonprofit offices where immigrants once felt safe are no longer regarded that way. (Nonprofit Quarterly, June 2026)

The Harm:	What Could Have Helped:
■ Clients stopped accessing services they needed. ■ When people disengage from services out of fear, the harm is immediate in the services not received, and long-term in the erosion of trust that makes it harder to reach those communities in the future.	■ Knowing what your data collection tools do with the information that passes through them. Read the privacy terms before signing, and look specifically for language about third-party data sharing and government access. ■ Building an informed consent process that tells clients in plain language what data you collect, where it goes, and what you do and do not control about its use ■ Considering what data you actually need to collect. If you can deliver services with less identifying information, that is a form of protection.

For Your Organization:

The data your organization collects about the people you serve does not always stay with you. Who can access that data is part of your duty of care to your community.

Photo: Pexels - Mikhail Nilov



What You Can Do About Algorithmic Bias

Most of the documented harm above involves tools that make decisions about people — that is where the highest-stakes bias has been recorded.

But bias in content is also real, and it requires a different response. What you can do depends on which situation you are in.



When You're Creating Content

Applies to: *writing assistants, image generators, grant-writing tools, and chatbots used for communications*

The intervention here is in what you put in and what you review before it goes out. This matters most for external-facing content — anything that represents your community to donors, funders, partners, or the public.



What You Put In: the CLEAR Framework

Vague prompts produce vague and biased outputs. When AI lacks context, it fills the gap using its training data defaults.

The **CLEAR** framework structures your prompts to reduce bias before content is generated.

For planned content — grant narratives, donor appeals, campaign materials, annual reports — use the full framework:

Letter	Stands For	To Answer	Example
C	Context	What is the situation, program, or organization?	“We are a community-led organization in [city] serving [specific population] through [type of program].”
L	Lived Experience Awareness	Who is affected and how should they be represented?	“The content should reflect the lived experiences of [community]. Show strength, agency, and specificity.”
E	Ethical Guardrails	What must the tool avoid?	“Avoid pity language, savior narratives, overgeneralizations, and trauma framing.”
A	Audience Sensitivity	Who will read or see this content?	“This will be shared with [community members/funders/partners].”
R	Response Style	What tone, format, and length is needed?	“Use dignified, community-centered language. No more than [word count]. [Tone].”

CLEAR framework adapted from AI Colab Studio, 2026



For everyday content, such as quick emails, routine communications, you do not need the full framework each time.

At minimum, tell the tool who your community is, what they should look like, and what to avoid.

- “Write a social media caption for a food pantry in [city] serving [community]. Use dignified language. No pity framing.”

That is enough to move the output away from defaults.

For Images Specifically:



- Name who is in the image and how they are shown. Skin tone, age, body type, ability, family structure, housing type, and setting.
- “Diverse” produces vague results. A person in a wheelchair who looks passive is not neutral representation.
- Specify what you actually want.



What You Review Before Your Content Goes Out

Every person on your team who creates content using AI should know how to review it for bias. This is not one person’s job. These are the questions to ask before anything goes to clients, donors, or the public.

- Does this reflect the actual people we serve, or a mainstream default version of them?
- Does it make assumptions about income, family structure, housing, language, or ability?
- Would the community described feel accurately and respectfully represented?
- Does it use any of the harmful patterns listed in Part One?

If the answer to any of these is **no** or **not sure**: re prompt with more specific direction. Note what was off. A pattern of corrections on the same tool tells you something about that tool’s bias baseline.



When Tools Make Decisions About People

Applies to: eligibility systems, intake tools, hiring platforms, case management software, screening tools, scheduling tools that affect service access

The intervention here is different because you often cannot directly change how the tool works. What you can do is control what happens around it.



Before a Tool Goes Live:

Start by learning how it works. Enter test scenarios — fictional cases that represent the range of people you serve — and observe what it produces.

- Ask the company, in plain language, what data the tool uses to generate its recommendations and how it weights those inputs.
- Use the [evaluation questions in Part Three](#) to guide that conversation. You cannot catch a biased output if you do not understand what the tool is doing.
- Require a human review step for every recommendation that affects a specific person. A staff member should see the output and have the authority to override it before anyone is turned away, scored, or assigned to a program.
- Build this step before deployment.



Once A Tool Is In Use:

Establish a clear process for staff to flag when an output looks wrong.

- Name who receives that concern.
- Document it.
- One flag is a data point.
- A pattern of flags on the same tool is grounds to escalate, audit, or stop using it.

Track outcomes by demographic group where possible.

- If a tool is producing different results across race, income, or primary language, that pattern is worth investigating.
- When a tool is producing biased results and you cannot change the system, raise it — internally to leadership, and externally to legal and advocacy partners if clients are being harmed.



What Staff Can Do:

- Trust your judgment. If a recommendation does not match what you know about a person, flag it.
- You are not required to implement a recommendation that contradicts your direct knowledge of someone's situation.



PART 3:

Build Your Organization's Response

Practical Action Steps

The following section provides guidance on how to:

- facilitate an internal AI bias discussion
- evaluate decision-making software
- draft a lean AI policy

It equips you with concrete evaluation frameworks, incident response protocols, and board-level talking points to ensure mission-aligned technology adoption.





AI Bias Conversation and Planning

You just need **one hour and this guide**.
(No facilitator required.)

The One-Hour Conversation

MINUTES 0:–0:15 What Tools Are We Using?

- Go around the room. Ask everyone to name any AI tool they have used for work in the past month, including tools built into the platforms they use daily.
- List and identify the tools that affect the people you serve.

MINUTES 0:15–0:35 What Is It Producing?

- Read aloud one case from Part Two that is closest to your organization's work.
- Then: where do we see something similar in how we work?
- Are there outputs we have accepted without reviewing them?

MINUTES 0:35–0:50 What Do We Put In Place?

- Walk through the policy questions in the next section.
- Which can your organization answer right now?
- Which have no answer yet?
- Identify one thing you can change in the next 30 days and name who owns it.

MINUTES 0:50–0:60 How Do We Flag A Concern?

- Name the person at your organization who receives bias concerns.
- Make sure everyone in the room knows who the point person is before you conclude the discussion.



After the Conversation

Assign someone to start your policy using the free tools in the following policy section.

- Schedule a **30-day check-in**
- Add a **six-month policy review** to the calendar



Evaluate Your Tools

Use these questions when adopting any new tool that will be used for client-facing work, hiring, or content representing your community.

If the tool is already in use, these questions still apply. Ask them now and use the answers to decide whether to continue, add safeguards, or stop.



For Tools That Make Decisions About People

(eligibility, hiring, intake, screening, case management)

These questions apply to tools that process information about specific people and produce scores, recommendations, or decisions. They do not apply to general conversational tools like ChatGPT.

- How does this tool arrive at a score or recommendation for a specific person? Ask the company to explain the mechanism in plain language. If they cannot, that is significant.
- When the tool influences a decision about one of our clients or applicants, will our staff be told that the tool was involved?
- What is the process for a client or applicant to appeal a decision, or for staff to override a recommendation?
- Has the tool been tested to see how its results differ across race, income, primary language, and geography? Ask for documentation.
- To find independent reviews, search the **Algorithmic Justice League (ajl.org)** or **AI Now Institute (ainowinstitute.org)** for any published assessments of the tool. A company that cannot provide any information about how their tool performs across different populations does not know their own tool's limitations.



- What data about the people we serve will this tool collect, store, or share? How could that data be used to inform future decisions about those same people? This is a bias question, not only a privacy one: data about specific populations stored in AI systems can be used in ways that reinforce the same disparities you are trying to address.



For Tools Used to Create Content

(writing assistants, image generators, grant-writing tools)

- Ask the tool for work-related content without specifying demographics. Its unprompted output reveals its bias baseline.
- Test it using language and names that reflect the people you serve, including dialects and communication styles outside standard written English.
- Once you understand the tool's defaults, use the CLEAR framework and bias review checklist in Part Two to reduce bias in what you put in and what you publish.

Chatbots

Chatbots can serve either function depending on how your organization uses them.

- If your chatbot handles communications — FAQ responses, general information, community engagement — it is a content tool. Test its responses using language and names that reflect your community and apply the bias review checklist in Part Two.
- If your chatbot handles intake, scheduling, eligibility screening, or service routing, it is making decisions that affect access. Apply the decision-making tool questions above and build a human override process before it goes live.
- Many organizations use the same chatbot for both functions. Evaluate each function separately.



The Bias Section of Your AI Policy

Your organization's AI policy may cover many areas. These are the questions it needs to answer specifically about bias.

A policy does not have to be long.

- One to two pages that your team can find and use is more valuable than a comprehensive document no one reads.
- Build in a review date: AI changes fast, and a policy from a year ago may already be outdated.

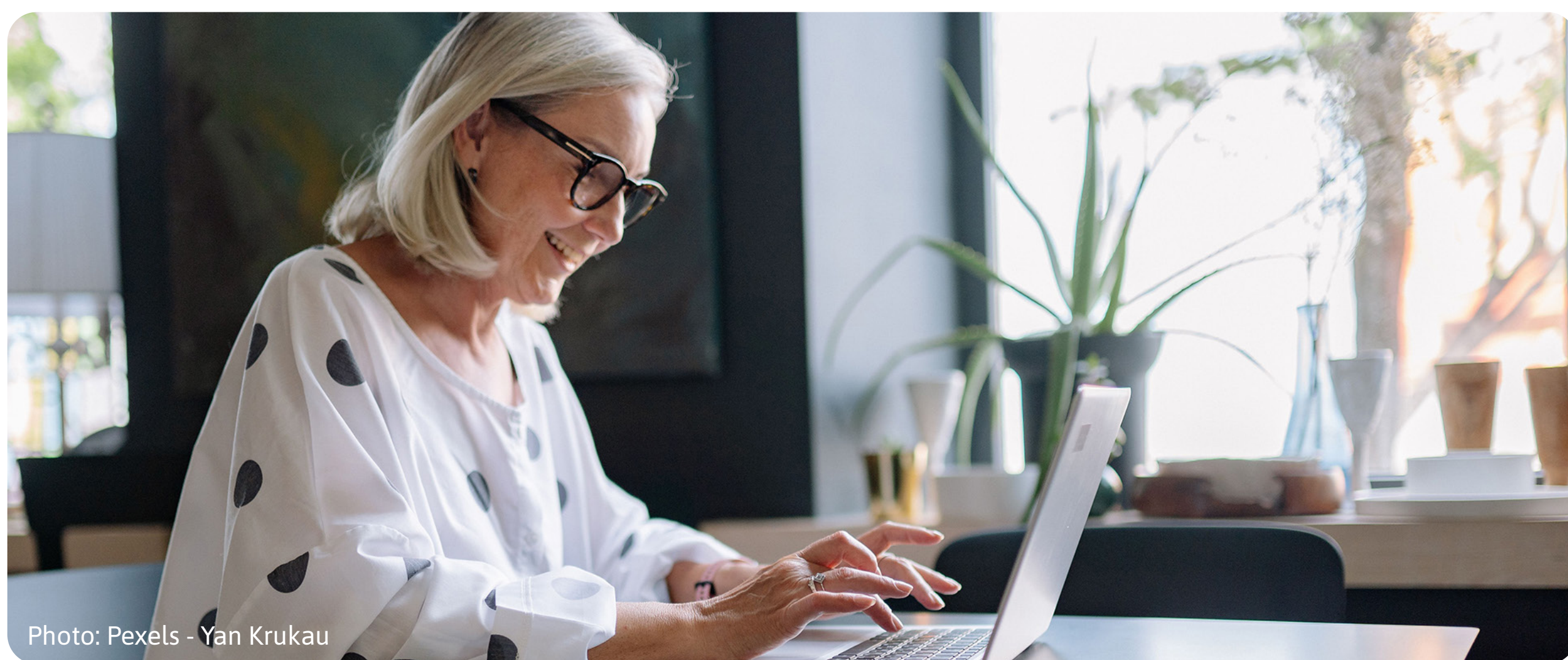


Photo: Pexels - Yan Krukau

Questions Your AI Bias Policy Needs To Answer:

1. What counts as AI use at our organization?

AI is built into tools you already use: Google Workspace, Canva, Mailchimp, scheduling and management software. Your policy should cover these embedded tools, not only standalone tools like ChatGPT.

2. Who reviews AI-generated content or recommendations before they affect a person?

Name the role. Whether it is a newsletter, a grant narrative, an eligibility recommendation, or an image: who is responsible for reviewing it before it goes out or is acted on?

3. What can staff use AI for in work that involves the people we serve, and what requires additional review?

A workable starting point: AI is appropriate for internal communications with human review; client-facing content requires designated review before use; AI tools should not be used to make final decisions about individual clients without a human override process.



4. What information about the people we serve can and cannot go into AI tools?

Client data stored in AI systems can be used in ways that produce discriminatory outcomes, including being used to train future models or inform decisions about people in similar situations.

Name the specific categories that are off limits: case details, financial information, legal history, housing status, immigration status, health information, and any other sensitive details relevant to your community.

This is a bias protection, not only a privacy measure.

5. What happens when a staff member identifies a bias concern?

Name who receives the concern. Describe what happens next.

A staff member who sees something biased needs to know where to take it and that something will be done with it.

6. When do we revisit this policy?

Name a review date and the person who owns it.

Every six months is the current standard.



Free Tools to Build Your Policy:

Fast Forward Nonprofit AI Policy Builder:

- Free; built for nonprofits
- Questionnaire-based, generates a customized policy
- No account required
- Budget 20 minutes

Recommended as it is faster, requires no personal information, and is designed for non-technical staff

Policy4good:

- Free; built for nonprofits
- More comprehensive, multiple export formats
- Requires your information to use
- Review their privacy policy before entering your organizational details



When Bias Shows Up: Document and Respond

A policy that names a person to receive bias concerns is only as useful as the process that follows.

Before something goes wrong, your organization needs to know what happens next. Document it in writing so it can be followed without being invented in the moment.



Photo: Pexels - Tiger Lily

What Triggers a Formal Review

- A single staff flag is a data point. A pattern of flags on the same tool, or a flagged output that has already reached clients or the public, triggers a formal review.
- Define the threshold before you need it.

What the Designated Person Does When They Receive A Concern

- Name the steps: acknowledge the flag, document the incident, review the output in question, determine whether other outputs from the same tool need review.
- Decide within a defined timeframe whether the tool continues in use as is, continues with added safeguards, or is paused.

How the Incident Gets Documented

- At minimum: tool name, date, what the output was, who flagged it, what was decided, and what changed as a result.
- A simple shared document or form is enough. The goal is a record your organization can learn from and refer back to.



What Happens to Outputs That Already Went Out

If biased content reached clients, donors, or the public, who decides what correction or follow-up is needed? Name the role.

Whether Affected Community Members Are Informed

If an AI-assisted decision affected how someone accessed your services, your policy should take a position on whether that person is told and given a path to request human review.

Keep It Current



Review Cadence

Your policy should have a review date every **six months**. When that date arrives, the review should take about **30 minutes** and cover:

- Has anything changed in how your AI tools work or what data they use?
- Have there been any bias flags or incidents since the last review?
- Are the people named in your policy — the reviewer, the person who receives concerns — still in those roles?
- Have you adopted any new tools since the last review that are not yet covered by your policy?
- Does your policy still reflect how your organization is actually using AI?

If the answer to any of these is yes, update the policy before the next review date.



New Staff

- When new staff join, point them to **Part One** and **Part Two** of this guide as part of onboarding.
- Make sure they know who receives bias concerns at your organization before they start working with any AI tool.



Talking to Your Board

For the ED to use in a board meeting, committee discussion, or direct conversation.



Photo: Pexels - Vlada Karpovich

Why We Are Having This Conversation

92% of nonprofits are now using AI in some form, but nearly half have no written policy governing how it is used. (2026 Nonprofit AI Adoption Report; Virtuous and Fundraising.AI, 2026)

- AI tools are already in use at this organization.
- The board's role is not to approve every tool.
- It is to ensure we have oversight in place so that AI use aligns with our mission.

What AI Bias Is and Why It Matters For Our Mission

- Every AI tool has bias built into it, a byproduct of what these systems learned from: human-generated content skewed toward dominant, mainstream cultures.
- The communities we serve are the communities most likely to be underrepresented in AI training data, and most likely to bear the consequences when AI produces biased results.



Documented examples from the sector:

- Tenant screening tools screening out Black and Latino applicants with applicants rarely told an algorithm was involved (Consumer Financial Protection Bureau, November 2022)
- Michigan's automated unemployment system issuing fraud determinations without human review, with **93% of audited determinations** found not to involve fraud (Benefits Tech Advocacy Hub)
- Roughly **9% of adults in immigrant families** avoiding safety net programs out of concern their information would reach immigration authorities (Urban Institute, April 2026)

When AI reinforces the inequities your organization exists to address, that is a mission alignment problem. A 2026 analysis from Forvis Mazars explicitly frames AI governance as part of the board's fiduciary responsibility. (Forvis Mazars, February 2026)



Photo: Pexels - Tiger Lily

What We Are Putting In Place

- A written AI bias policy addressing who reviews AI outputs before they affect people, what information cannot enter AI tools, and what happens when a bias concern is identified
- A process for staff to flag bias concerns, with a named person to receive them
- A review of our policy and our tools every six months



What We Are Asking From the Board

- **Awareness and endorsement:** Be aware of how AI is being used at this organization and formally endorse having a written policy that addresses bias risk.
- **A board owner:** one board member to take ownership of AI governance as part of their portfolio, not as a technical expert, but to keep it on the board's agenda
- **A position on transparency:** how should we communicate our AI use to clients, donors, and grant funders? We need board guidance on this.



Photo: Pexels - Yan Krukau

The Short Version:

We are already using AI. The question is whether we are using it in ways that align with our mission and protect the people we serve.



A Final Note

Bias in AI tools is not a problem that gets solved once. The tools change, the bias shifts, and new patterns emerge.

What you are building with this guide — **the habit of reviewing outputs, a policy your staff can use, a person who knows where to take a concern** — is what makes the difference for the communities you serve.



Resource Links

For organizations that want to do further research



Documented Bias in Nonprofit Settings

American Civil Liberties Union. (2022, October 3).

Algorithms are making decisions about health care, which may only worsen medical racism.

<https://www.aclu.org/news/privacy-technology/algorithms-in-health-care-may-worsen-medical-racism>

Benefits Tech Advocacy Hub. (n.d.).

Michigan unemployment insurance false fraud determinations.

National Health Law Program, TechTonic Justice, & Upturn.

<https://www.btah.org/case-study/michigan-unemployment-insurance-false-fraud-determinations.html>

Brennan, K, Kak, A., & Myers West, S. (2025, June 3).

Artificial power: 2025 landscape report. AI Now Institute.

<https://ainowinstitute.org/publications/research/ai-now-2025-landscape-report>

Consumer Financial Protection Bureau. (2022, November 15).

CFPB reports highlight problems with tenant background checks.

<https://www.consumerfinance.gov/about-us/newsroom/cfpb-reports-highlight-problems-with-tenant-background-checks/>

Costa, M. C. (2026, June 2).

As surveillance of immigrant communities expands, how can nonprofits respond?

Nonprofit Quarterly.

<https://nonprofitquarterly.org/as-surveillance-of-immigrant-communities-expands-how-can-nonprofits-respond/>

Gonzalez, D., Haley, J. M., Bernstein, H., Kenney, G. M., & Karpman, M. (2026, April 8).

Immigrant families disengaged from public life and essential services because of immigration concerns in 2025. Urban Institute.

<https://www.urban.org/research/publication/immigrant-families-disengaged-public-life-and-essential-services-because>



Hofmann, V., Kalluri, P. R., Jurafsky, D., & King, S. (2024).

AI generates covertly racist decisions about people based on their dialect.
Nature, 633(8028), 147–154.

<https://ora.ox.ac.uk/objects/uuid:4e0fe173-52f2-433d-8042-0ff941919401>

Lecher, C. (2025, May 20).

The NYC algorithm deciding which families are under watch for child abuse.
The Markup.

<https://themarkup.org/investigations/2025/05/20/the-nyc-algorithm-deciding-which-families-are-under-watch-for-child-abuse>

Hiring And Employment

Bertrand, M., & Mullainathan, S. (2004).

Are Emily and Greg more employable than Lakisha and Jamal? A field experiment on labor market discrimination. *American Economic Review*, 94(4), 991–1013.

<https://www.aeaweb.org/articles?id=10.1257/0002828042002561>

MIT Technology Review. (2018, October 10).

Amazon ditched AI recruitment software because it was biased against women.

<https://www.technologyreview.com/2018/10/10/139858/amazon-ditched-ai-recruitment-software-because-it-was-biased-against-women/>

For Your Policy

Algorithmic Justice League. (n.d.).

<https://www.ajl.org>

AI Now Institute. (n.d.).

<https://ainowinstitute.org>

Bodach, K., & Prater, D. (2026, February 13).

AI governance for nonprofit boards. Forvis Mazars.

<https://www.forvismazars.us/forsights/2026/02/ai-governance-for-nonprofit-boards>

Fast Forward. (n.d.).

Nonprofit AI policy builder.

<https://www.ffwd.org/nonprofit-ai-policy-builder>

Policy4good. (n.d.).

<https://policy4good.ai/>



Sector Context

Das, M. (2025, November 6).

What AI equity for nonprofits really looks like. Candid.

<https://candid.org/blogs/what-ai-equity-for-nonprofits-means-looks-like-in-practice/>

Hanna, H., & Nowack, D. (2025, December 1).

Bridging the AI investment divide. Stanford Social Innovation Review.

<https://ssir.org/articles/entry/ai-investment-divide>

Lomastro, J. A. (2026, April 1).

AI in the nonprofit sector is a question of governance, not just technology. Nonprofit Quarterly.

<https://nonprofitquarterly.org/ai-in-the-nonprofit-sector-is-a-question-of-governance-not-just-technology/>

NetHope. (2026, April 2).

AI governance and the nonprofit sector: Mapping the missing middle.

<https://nethope.org/toolkits/ai-governance-and-the-nonprofit-sector-mapping-the-missing-middle/>

NonProfit PRO. (2026, February 17).

Nonprofit AI adoption hits 92%, but only 7% see major impact.

<https://www.nonprofitpro.com/article/nonprofit-ai-adoption-hits-92-but-only-7-see-major-impact/>

Developed by Kim Olpin, Executive Director, Jericho Road Pasadena with Tomar Beh (AI Colab Studio) for She Leads AI AI Governance Community of Practice, Pod 2.

The CLEAR prompting framework in Part Three is adapted from AI Colab Studio (2026).

Foundational research synthesized by Audrey Schaub, COO, Vativ.ai, for Pod 2.

Pod 2: Audrey Schaub (Pod Lead) · Carmela Pereira (Communications and Creative Direction) · Jasmine Schwarz (Artifact Lead) · Tomar Beh (Tech Steward) · Marcia Narine Weldon · Kim Olpin.

All sources verified August 2026. Last updated: September 2026.