GREEN LIGHTNING
DIGITAL MARKETING

ETHICAL AI SERIES

The Ethical AI Practitioner's Guide

Training yourself and your LLM for inclusivity, equity, and anti-racism — with actionable guardrails, inclusive prompting techniques, and a framework for systemic change.

By Green Lightning Marketing | March 2026 | greenlightningmarketing.com

1. THE PROBLEM 2. GUARDRAILS 3. INCLUSIVE PROMPTING 4. TRAIN YOURSELF
5. THE FRAMEWORK REFERENCES

The rapid adoption of Artificial Intelligence has brought remarkable capabilities to everyday work, but it has also surfaced a critical and urgent flaw: AI systems frequently amplify and perpetuate systemic racism, bias, and exclusion. Because these models are trained on historical human-generated data, they do not merely reflect the world as it is — they inherit the prejudices of the past and project them into the future at unprecedented scale.

This is not a passive problem. Every time you use an AI tool without intentional guardrails, you risk co-creating outputs that erase, stereotype, or harm marginalized communities. Conversely,

every time you use AI with deliberate, inclusive intention, you have an opportunity to actively counter these defaults and model what equitable AI use looks like.

Who this guide is for: Practitioners, educators, communicators, and organizational leaders who want to take concrete, actionable steps to use AI with ethical intention — and to teach their AI tools to do the same.

PART 1

Understanding the Problem — How AI Amplifies Racism

Before building solutions, we must understand the mechanisms by which AI perpetuates racial harm. This is not a matter of individual bad actors; it is structural, embedded in the very architecture of how these systems are built.

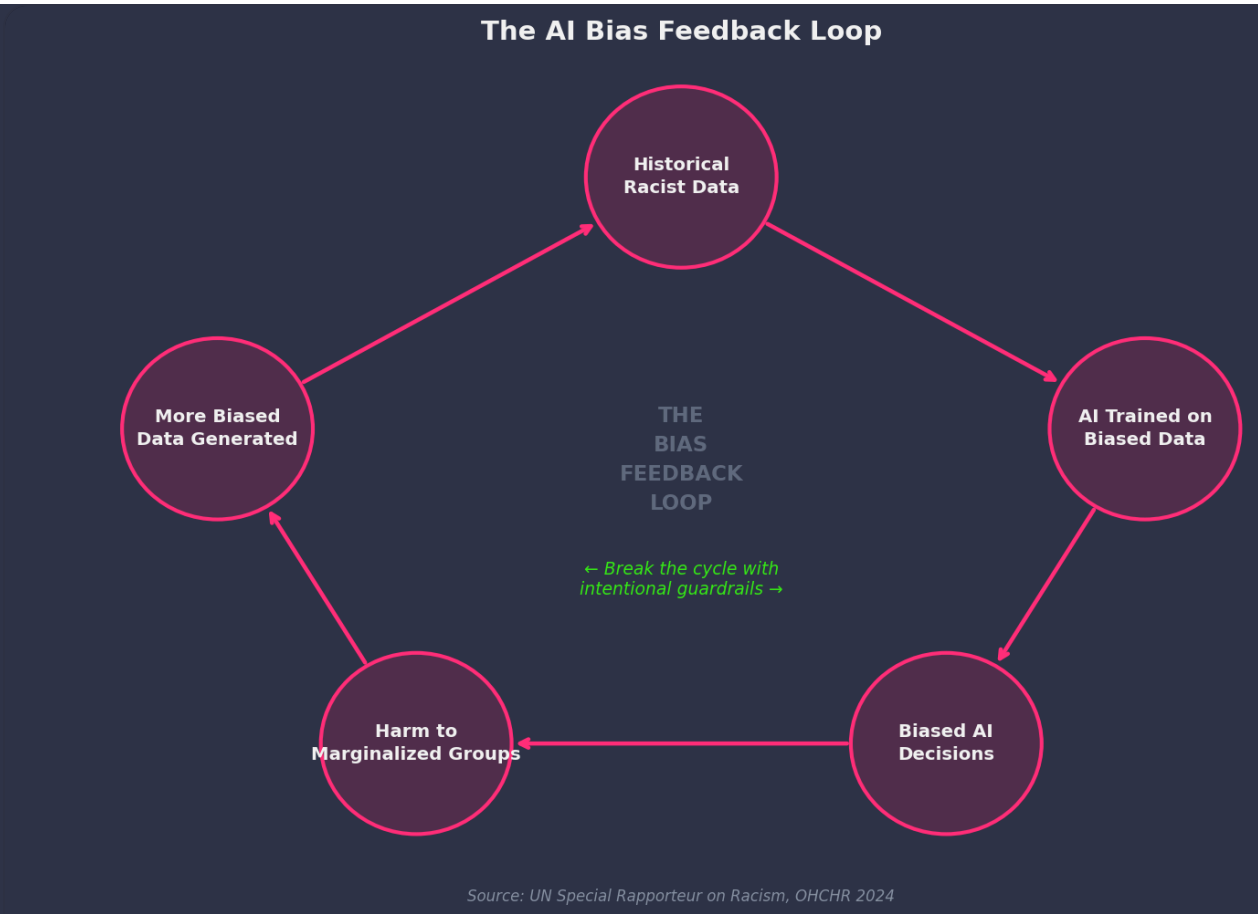

Fig. 1 — The AI Bias Feedback Loop. Each stage reinforces the next, creating a self-perpetuating cycle that requires deliberate intervention to break. Source: UN Special Rapporteur on Racism, OHCHR 2024.

1.1 The Feedback Loop of Historical Bias

AI systems are trained on historical data saturated with the inequities of the societies that produced it. A predictive policing algorithm trained on arrest records from historically over-policed neighborhoods will disproportionately flag those same neighborhoods for future surveillance — creating a self-reinforcing cycle. The UN Special Rapporteur found this dynamic extends across

1.3 The Default Persona Problem

When AI image generators receive a vague prompt — "a professional," "a doctor," "a CEO" — they default to a narrow, stereotyped representation: typically white, male, young, slim, and able-bodied. This reflects the demographic skew of internet-scraped training datasets. The result is that marginalized groups are systematically erased from positions

healthcare, education, employment, and criminal justice.

> *"Bias from the past leads to bias in the future."*
>
> **— Ashwini K.P., UN Special Rapporteur on Contemporary Forms of Racism**

of authority and expertise in AI-generated visual content.

"People learn from seeing — or not seeing — themselves that maybe they don't belong." — Heather Hiles, Chair, Black Girls Code

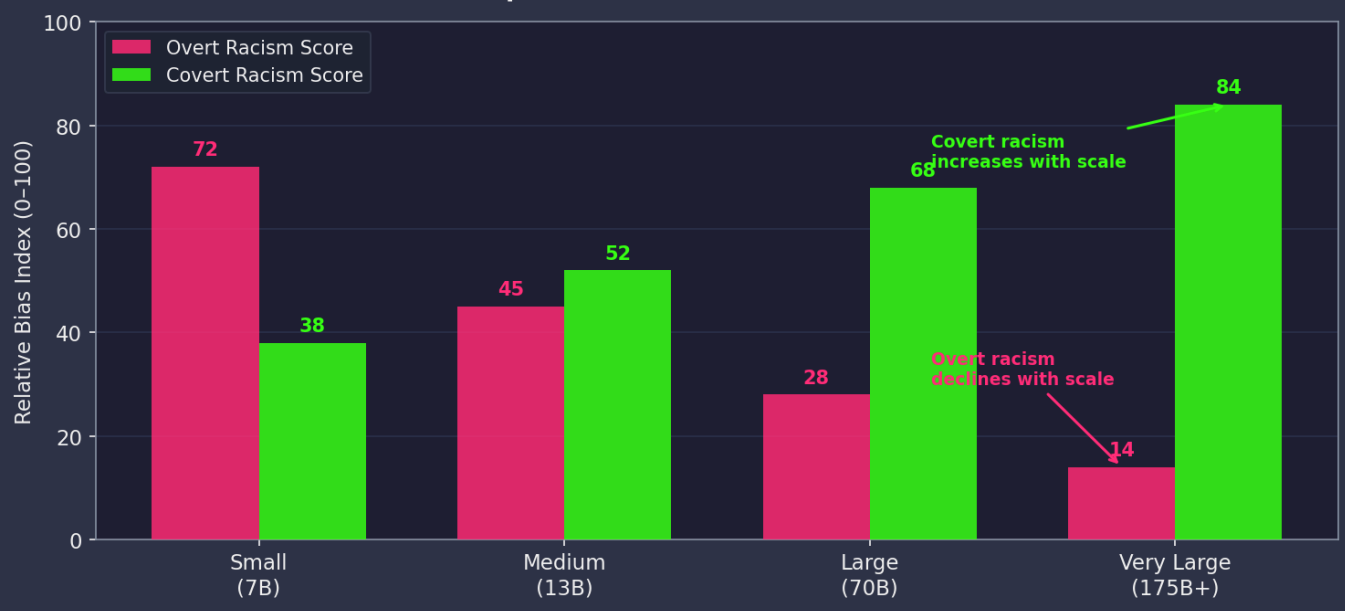

Fig. 2 — Covert vs. Overt Racism by Model Scale. Stanford HAI / Nature research (2024) found that while safety training reduces overt racist outputs, covert racism — expressed through differential treatment of African American English dialect — actually increases as models grow larger.

1.2 Covert Racism in Large Language Models

Research from Stanford University published in *Nature* (2024) reveals a particularly troubling dynamic: as developers have successfully reduced *overt* racism in LLMs, *covert* racism has actually increased. Using a sociolinguistic technique called the "matched guise" method, researchers found that when LLMs evaluated identical

content written in African American English (AAE) versus Standard American English (SAE), the models were significantly more likely to associate AAE speakers with negative stereotypes from 1930s–1950s research, assign them lower-prestige jobs, and recommend harsher criminal sentences.

The critical insight: Simply not mentioning race to an LLM does not prevent it from expressing racist attitudes. The language itself — the dialect, syntax, cultural context — is enough to trigger differential and discriminatory treatment.

1.4 Proxy Variables and Invisible Discrimination

AI systems can discriminate even when race is not explicitly included as a variable. Socioeconomic background, zip code, educational institution, and dialect all function as proxies for race, allowing algorithms to perpetuate racial discrimination while appearing facially neutral. This is the structural dimension of algorithmic racism: the bias is embedded in the choice of variables, the design of the model, and the composition of the training data — not in any single discriminatory decision.

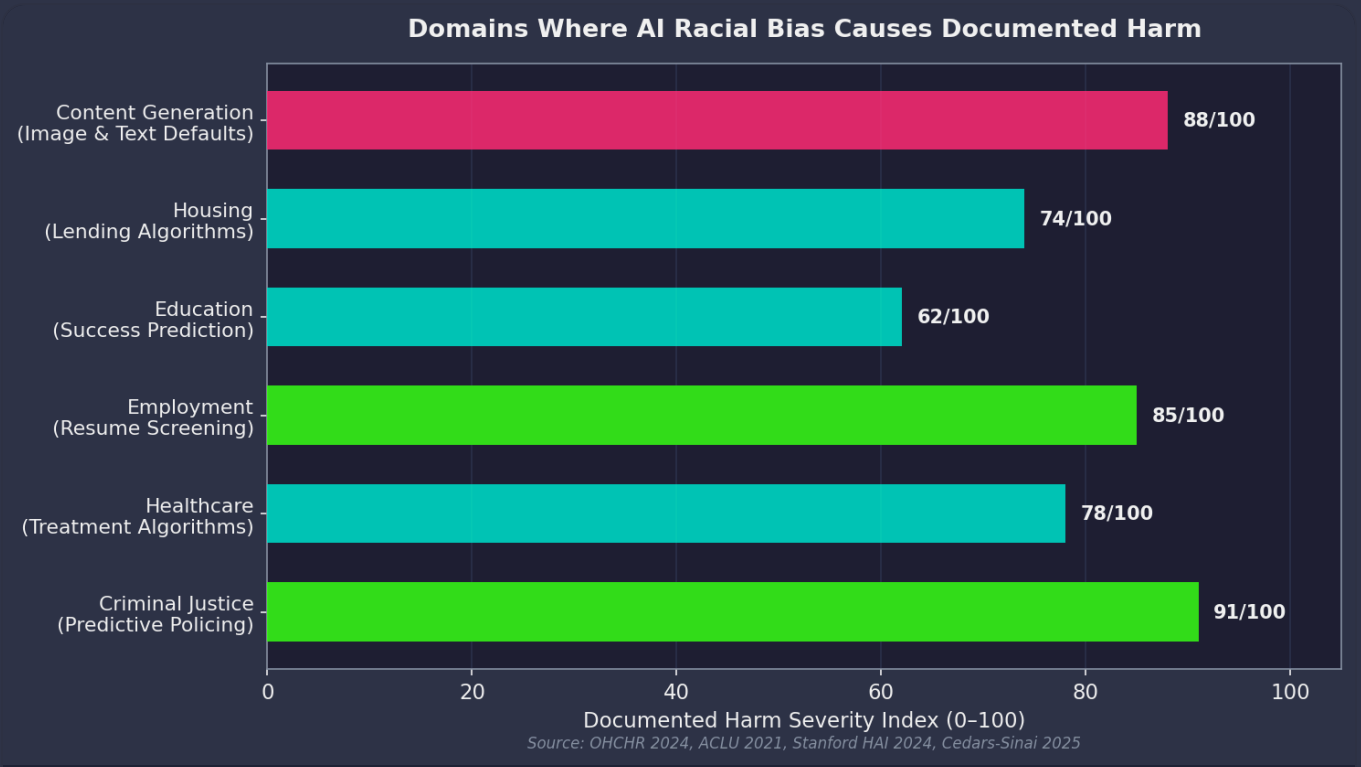

Fig. 3 — Domains Where AI Racial Bias Causes Documented Harm. Content generation scores highest alongside criminal justice and employment, making inclusive prompting a direct harm-reduction practice. Sources: OHCHR 2024, ACLU 2021, Stanford HAI 2024.

PART 2

Establishing Guardrails — Custom Instructions for Your LLM

The most powerful lever you have as an individual practitioner is the ability to set persistent, ethical instructions that govern how your AI behaves before it generates a single output. Most major AI platforms allow this through custom instructions, system prompts, or project-level settings.

PLATFORM	WHERE TO SET PERSISTENT INSTRUCTIONS
ChatGPT (OpenAI)	Settings → Personalization → Custom Instructions
Claude (Anthropic)	Projects → Project Instructions
Gemini (Google)	Gems → Custom Instructions
API / Developer Use	System Prompt field at the top of the conversation
Microsoft Copilot	Copilot Studio → System Prompt configuration

ETHICAL & INCLUSIVE AI GUARDRAILS — COPY & PASTE READY

1 **Default to Diversity**

When generating scenarios, examples, personas, stories, slide content, or imagery prompts, always lead with mindfully inclusive populations. Ensure representation across race and ethnicity, age, gender identity and expression, sexual orientation, body size and shape, and physical and cognitive ability. Apply this default unless the specific historical, cultural, or factual context of the request strictly requires otherwise.

2 **Counter Stereotypes Actively**

Deliberately place historically underrepresented groups in positions of agency, leadership, expertise, and authority. Examples: women of color as executives and scientists; older adults as technology innovators; people with visible disabilities in active, empowered, and everyday roles; men in caregiving and nurturing professions. Do not confine representation to roles that confirm existing stereotypes.

3 **Avoid Tokenism and Caricature**

Do not reduce complex identities to tropes, symbols, or clichés. A single "diverse" character added to an otherwise homogeneous scene is tokenism, not inclusion. Aim for authentic, grounded representation where diverse characters exist in realistic contexts with agency and interiority — not merely as visual checkboxes.

4 **Use Respectful, Precise Language**

Apply person-first or identity-first language as appropriate and as preferred by the community being discussed. Avoid ableist, ageist, racist, classist, or gendered assumptions in all language. Do not use euphemisms that obscure systemic harm (e.g., "achievement gap" without acknowledging its structural causes).

5 **Acknowledge Systemic Context**

When discussing topics that intersect with race, gender, disability, class, or other axes of identity, provide context that acknowledges structural and systemic inequalities. Do not present a sanitized, "both sides" framing that falsely equates systemic harm with individual discomfort.

6 **Flag Potential Bias**

If you detect that a prompt or a generated output may inadvertently reinforce a stereotype or exclude a marginalized group, flag it proactively and offer a more inclusive alternative.

2.2 Applying the Rules: What This Looks Like in Practice

SCENARIO	WITHOUT GUARDRAILS (AI DEFAULT)	WITH GUARDRAILS APPLIED
Leadership slide deck	Defaults to images of white men in suits	Diverse group across race, gender, age, and ability in leadership contexts
Entrepreneur case study	Defaults to a young white male tech founder	Black woman founder, Latina immigrant, or older entrepreneur re-entering the workforce
Medical consultation scenario	Doctor is white and male; patient is passive	South Asian female doctor; older Black male patient who is an active, informed participant
"Urban community challenges"	Defaults to crime and poverty narratives	Acknowledges structural disinvestment, redlining history, and community resilience

PART 3

Inclusive Prompting — A Practical Skill Set

Guardrails set the baseline, but your daily prompting habits are what make the difference in each specific output. Inclusive prompting is not about adding more words to a prompt; it is about writing better instructions that interrupt the AI's biased defaults.

Inclusive Prompt Transformation: Before & After

Context	Default (Biased) Prompt	Inclusive Prompt
Business Meeting	"Professionals in a boardroom"	"Collaborative team: Black woman in her 40s leading, South Asian man in 30s, larger-bodied woman in 50s, man with prosthetic arm"
Healthcare	"A doctor with a patient"	"South Asian female doctor consulting with an elderly Black male patient who is engaged and asking informed questions"
Technology	"A software engineer at a computer"	"Latina woman in her 30s coding at a standing desk; colleague who uses a wheelchair working nearby"
Education	"A teacher with students"	"Black male teacher in his 40s facilitating discussion with racially diverse students, including a student with a hearing aid"

Source: Articulate Inclusive Prompting Guide, 2025

Fig. 4 — Inclusive Prompt Transformation: Before & After. Specificity is the key mechanism — the moment you leave a demographic detail unspecified, the AI fills it in with its statistical default. Source: Articulate Inclusive Prompting Guide, 2025.

3.1 For Image Generation and Visual Content

The following five dimensions should be explicitly specified in any image generation prompt for professional or public-facing content:

Race & Ethnicity

Name specific communities or describe a range of skin tones and ethnic backgrounds. Never leave this unspecified in a prompt intended for diverse audiences.

Age

Specify age ranges and mix them. Counter the tech industry's default of youth by including adults over 50 in active, leadership, and innovative roles.

Gender & Expression

Body Size & Disability

Include women, non-binary individuals, and gender-diverse people. Challenge the gendering of professions — include men in caregiving, women in engineering.

Include larger-bodied people in roles of confidence and authority. Include people with visible assistive devices (wheelchairs, white canes, hearing aids) in empowered, everyday contexts — not as symbols of struggle.

3.2 Representing Invisible Identities Through Context

Many aspects of identity — neurodiversity, invisible disability, sexual orientation, socioeconomic status, religious affiliation — cannot be depicted literally without reducing them to stereotypes. Instead, use environmental and contextual cues:

IDENTITY	CONTEXTUAL CUE TO USE IN PROMPT
Neurodiversity	Sensory-friendly workspace with soft lighting and noise-canceling headphones
LGBTQ+ Identity	Rainbow flag on a desk; same-sex couple in a casual, non-fetishized scene
Socioeconomic Diversity	Mix of characters navigating public transit alongside characters in tailored suits
Religious Identity	Hijab, kippah, turban, or cross visible as natural, unremarkable elements of the scene
Invisible Disability	Character referencing medication, a service animal, or a medical alert bracelet

3.3 For Slide Decks and Written Content

Slide Deck Instruction (add to every AI slide prompt)

> *"Ensure that all examples, case studies, and personas in this slide deck reflect a diverse range of racial backgrounds, genders, ages, and abilities. Specifically include at least one example centered on a community of color, and ensure that no single demographic group is the default protagonist of every scenario."*

Research & Analysis Instruction

> *"Include a section on the disproportionate impact of this issue on communities of color, low-income communities, and other marginalized groups. Cite specific data where available."*

Self-Audit Instruction (ask the AI to check its own output)

> *"Review the personas and examples you have generated. Are any demographic groups over- or under-represented? Are any groups depicted only in stereotyped roles? Revise to correct any imbalances."*

PART 4

Training Yourself — A Personal Practice for Equity and Justice

Training the AI is only half the work. The other half is developing your own critical consciousness as an AI practitioner. The following steps constitute an ongoing personal practice, not a one-time checklist.

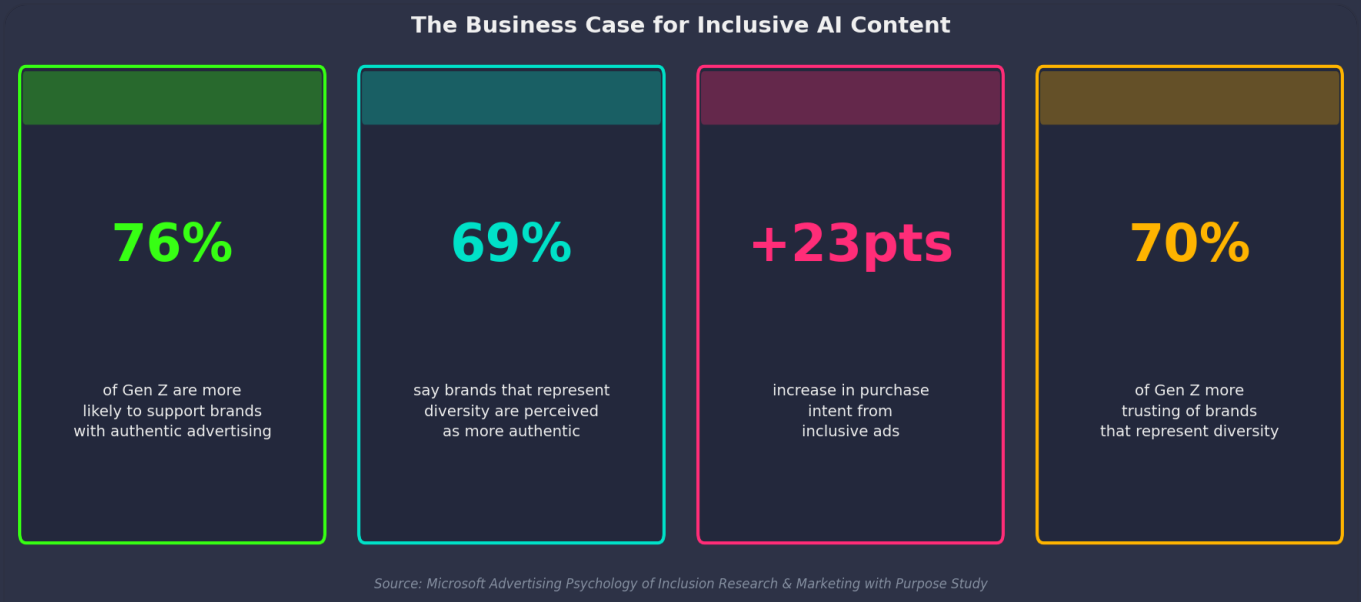

Fig. 5 — The Business Case for Inclusive AI Content. Inclusive representation is not only an ethical imperative — it is a competitive differentiator. Source: Microsoft Advertising Psychology of Inclusion Research & Marketing with Purpose Study.

1

Audit Your Own Prompts

Before you submit a prompt, pause and ask: *Who is the default protagonist of this scenario? Who is being left out? Who might be harmed by this output?* This is not about slowing down your work; it is about building a reflex. Keep a personal log of prompts where you caught and corrected a bias — this log becomes a resource for training others.

2

Verify and Validate All AI Outputs

AI systems are prone to hallucination, and they are particularly likely to reproduce dominant narratives uncritically. When AI generates content about racial or social justice issues, treat it as a first draft requiring human verification. Cross-check facts and historical claims against peer-reviewed sources, government data, and reporting from communities directly affected. Never use AI-generated content about a marginalized community as the authoritative source on that community's experience.

3

Center Marginalized Voices in Decision-Making

AI is a powerful tool for structural and logistical work, but it is not a substitute for lived experience. When designing products, policies, or programs that affect marginalized communities, use AI to assist with research synthesis, drafting, and iteration — but ensure that people from those communities are centered in the actual decision-making process. The disability rights movement's foundational maxim applies equally here:
"Nothing about us without us."

4

Advocate for Accountability at the Organizational Level

Individual practice matters, but systemic change requires organizational action. When evaluating AI vendors, ask: How do you audit your training data for racial bias? What fairness metrics do you use? Have you conducted third-party bias audits? Advocate for an organizational AI use policy that explicitly addresses bias, requires inclusive defaults, and establishes accountability mechanisms.

5

Use Bias Auditing Tools

For practitioners who work with AI at a technical level, or who are evaluating AI tools for organizational use, open-source frameworks provide structured methods for measuring and mitigating bias.

TOOL	DEVELOPER	PRIMARY USE
AI Fairness 360 (AIF360)	IBM Research	70+ fairness metrics and bias mitigation algorithms for ML models
What-If Tool	Google	Interactive visual interface for probing model behavior across demographic groups
Fairlearn	Microsoft	Fairness assessment and mitigation for classification and regression models
DeepEval Bias Metric	Confident AI	LLM-as-judge evaluation for racial, gender, and political bias in LLM outputs

TOOL	DEVELOPER	PRIMARY USE
Fiddler AI Observability	Fiddler AI	Ongoing monitoring of fairness and bias in production LLM and ML systems

4.4 Organizational Advocacy Actions

LEVEL	ACTION
Procurement	When evaluating AI vendors, ask: How do you audit training data for racial bias? What fairness metrics do you use? Have you conducted third-party bias audits?
Policy	Advocate for an organizational AI use policy that explicitly addresses bias, requires inclusive defaults, and establishes accountability mechanisms.
Team Practice	Establish a team norm of reviewing AI-generated content for bias before publication. Rotate the "bias checker" responsibility so it does not fall disproportionately on employees from marginalized groups.
Feedback Loops	Use the feedback mechanisms in AI tools (thumbs down, "this response was harmful") to flag biased outputs. This data feeds back into model improvement.
Education	Share this guide. Run workshops on inclusive prompting. Make AI literacy and anti-bias practice a standing component of onboarding and professional development.

PART 5

A Framework for Ongoing Practice

Ethical AI use is not a destination; it is a practice. The **AUDIT → INSTRUCT → PROMPT → VERIFY → ADVOCATE** cycle provides a repeatable loop for continuous improvement.

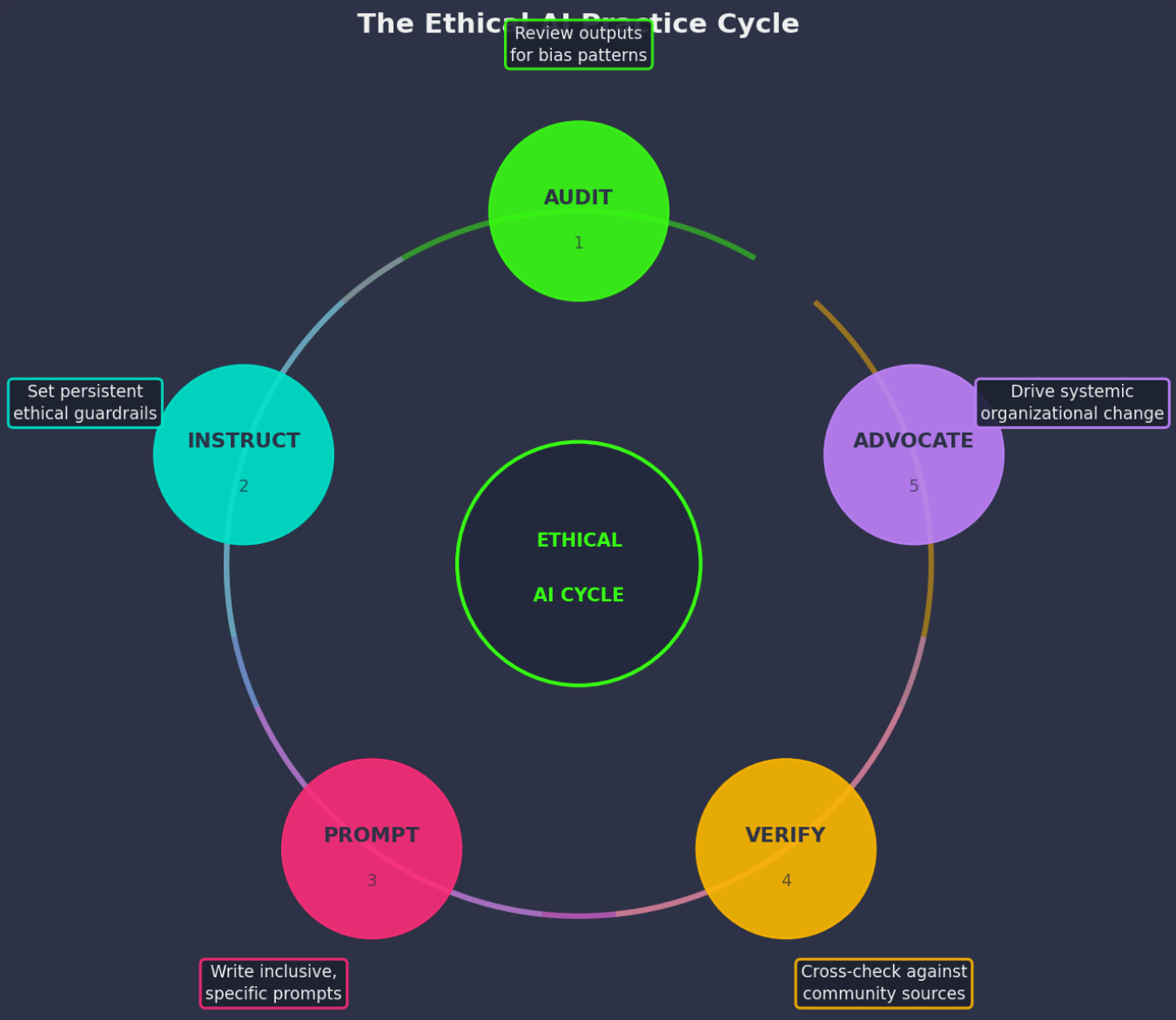

Fig. 6 — The Ethical AI Practice Cycle. This five-step loop is designed to be repeated continuously, not completed once. Each iteration deepens your practice and refines your guardrails.

1 — AUDIT

Review recent AI outputs for patterns of bias, exclusion, or stereotyping. Where are the gaps? Who is

2 — INSTRUCT

Set persistent ethical guardrails in your LLM's custom instructions or system prompt (see Part 2). Set

missing? Which roles are being assigned to which groups? This is your diagnostic step.

these instructions once and update them as your practice evolves and as new bias patterns emerge.

3 — PROMPT

Write with intentional specificity (see Part 3). Make inclusive prompting a habit, not an afterthought. Use the before/after prompt transformation framework as a mental model for every generation request.

4 — VERIFY

Cross-check all AI outputs against human-authored, community-centered sources. Do not accept the AI's version of marginalized communities' experiences. Treat every AI output as a first draft.

5 — ADVOCATE

Drive systemic change at the organizational and industry level. Individual practice is necessary but not sufficient; the structures that produce biased AI must also change. Use your voice, your purchasing decisions, and your organizational influence to demand accountability from AI vendors and policymakers.

The goal is not merely to make AI less harmful. The goal is to use AI as a tool for **actively imagining and building a more equitable world** — one prompt, one slide deck, one policy brief at a time.

GREEN LIGHTNING MARKETING

Leading the Responsible Revolution

greenlightningmarketing.com | Ethical AI Series | March 2026

REFERENCES

1. UN Office of the High Commissioner for Human Rights. "Racism and AI: 'Bias from the past leads to bias in the future'." July 30, 2024. ohchr.org
2. Stanford University HAI. "Covert Racism in AI: How Language Models Are Reinforcing Outdated Stereotypes." September 3, 2024. hai.stanford.edu
3. Articulate. "How to Create Inclusive AI Images: A Guide to Bias-Free Prompting." 2025. articulate.com
4. Microsoft Advertising. "Leading in the era of AI: Best practices for inclusive prompting." April 17, 2024. microsoft.com
5. Hofmann, V., Kalluri, P.R., Jurafsky, D., & King, S. "AI generates covertly racist decisions about people based on their dialect." *Nature*, August 28, 2024. nature.com
6. ACLU. "How Artificial Intelligence Can Deepen Racial and Economic Inequities." July 13, 2021. aclu.org
7. Cachat-Rosset, G. & Klarsfeld, A. "Diversity, Equity, and Inclusion in Artificial Intelligence." *Applied Artificial Intelligence*, 2023. tandfonline.com
8. IBM Research. "Introducing AI Fairness 360." research.ibm.com