AI Governance Framework for Nonprofits

AI POLICY TEMPLATE

SEPTEMBER 2024

ANB Advisory | Afua Bruce & Rose Afriyie

How to Use This Template

This template provides the foundation for an AI policy for your organization. It is written with the understanding that teams or individuals at your nonprofit may already be experimenting with AI tools. We encourage approaching AI adoption with curiosity and experimentation while balancing this with policies that enable leaders to manage risk, protect liability, and achieve compliance in pursuit of your organization's mission. If your organization has existing data and/or privacy policies, you may wish to copy the relevant sections from this template into your existing policies to those documents. This template is intended to be a starting point – keep the sections you need, modify sections as needed, and delete sections that are not relevant to your organization. For ease of use, please refer to the color-coded icons in the left margin at the start of each paragraph. They indicate which nonprofits may find the paragraph most relevant.

Although you can simply have one individual tailor this template for your organization, it is suggested that a cross-functional group at your organization discuss the template before finalizing your AI policy.

Note: Resources may not be available to implement every aspect of the AI Policy fully. Remember that some progress is better than no progress. It is recommended organizations begin with overarching principles and general AI guidelines.

KEY

Nonprofits not building technology
This category is for nonprofits that are using or are considering using commercial off-the-shelf AI tools. Nonprofits in this category may customize tools using non-coding methods.

Nonprofits building technology
This category is for nonprofits that are writing code, developing new tools, and utilizing APIs to create new, custom tools and functionality for their organization.

Overarching AI Principles

Generative AI offers a chance to help nonprofit organizations like [Organization Name] in new ways. Individual tools, as well as AI systems, can affect how organizations meet their community's needs. In a world with limited resources, nonprofit organizations need to explore all available tools to fulfill their mission to serve. However, valid concerns about AI's potential for bias and unethical use exist. To balance these concerns and support our community service efforts, when [Organization Name] uses AI, we will do so responsibly by applying the following principles:

- **Accountability and responsibility:** We will hold the individuals using AI tools accountable for their decision-making based on the tools. AI tools and systems support the work our staff remains committed to.
- **Equity and access:** We will use AI in ways that do not create new inequities or barriers to accessing vital services. We will maintain safeguards that promote fair access to our services.[1]
- **Fairness and non-discrimination:** We will aim to use AI tools in ways that do not discriminate against the communities we serve. We will seek to use AI tools that minimize bias and ensure fair outcomes for everyone, regardless of race, gender, ethnicity, or other factors.[2]
- **Reliability and accuracy:** We will use AI tools that perform as intended. We will select AI tools that consistently produce accurate outputs.
- **Transparency:** We will explain when and how we use AI products when asked by our stakeholder communities.
- **Trust:** We will use AI in ways that allow us to maintain trust with the community we support.
- **In service of mission:** We will use AI in support of human decision-making, expertise, and creativity, and not in place of human expertise. AI tools will be selected because there is a way for them to support our mission, not just because it is a new technology.

These Overarching AI Principles support and supplement our organization's stated mission, vision, and values. We recognize that both our organization and AI technology will evolve; therefore, we will review and update this policy [frequency].

[1] The USDA has developed a framework for use of AI in public benefit administration. Here they mention key principles that can be found in section 1.4 with equity and access being one of the core principles lifted up for this example: https://www.fns.usda.gov/framework-artificial-intelligence-public-benefit

[2] This principle is adapted from the Microsoft AI principles.

General AI Guidelines

Specific AI tools and functions change regularly. Individual staff may want to use AI tools to support their work; others may want to experiment with developing their own tools. The following AI guidelines govern how [Organization Name] can use AI, rather than constantly deciding on a case-by-case basis. Given some of the risks of AI, these guidelines also serve as a risk management tool for the organization.

- Does staff need approval to use or develop AI tools? [Yes/No]
  - If yes, what level (eg: who) must approve?
  - If yes, does this apply for all tools?
- Can you use AI to generate or modify external-facing written content? [Yes/No]
  - If yes, do you need to disclose? In what circumstances?
- Can you use AI to generate or modify internal-facing written content? [Yes/No]
  - If yes, do you need to disclose? In what circumstances?
- Can you use AI to generate or modify external-facing visual content? [Yes/No]
  - If yes, do you need to disclose? In what circumstances?
- Can you use AI to generate or modify internal-facing visual content? [Yes/No]
  - If yes, do you need to disclose? In what circumstances?
- Can you use AI to capture and summarize meeting notes with internal or external parties? [Yes/No]
- Can you use AI for translation? [Yes/No]
- Can you use AI for research purposes? (eg: individuals, other organizations, market trends, etc.) [Yes/No]
- Can you use AI for brainstorming and editing purposes? (eg: revising human-generated drafts) [Yes/No]
- Can you use AI for analysis and prediction purposes? [Yes/No]
- Can you use generative AI tools wholesale, or must the output be human-modified and/or verified in some way? [Yes/No]
  - If yes, where should documentation of the review be kept?
- Must all contractors adhere to the guidelines in this policy document? [Yes/No]
  - If no, what are the exceptions?
- Must all staff take data literacy and AI training? [Yes/No]
  - If no, which staff must take the training?
  - How often should training be taken?
  - Link to the training here: [link]

Data/IT Governance & Privacy

[Organization Name] will follow all relevant industry guidelines and regulations (eg: HIPAA standards[3]), as well as all applicable data laws and policies (eg: GDPR, EU AI Act).

AI tools use a lot of data. The tools themselves are trained from different data sources, and to work inside [Organization Name], the AI tools will use [Organization Name]'s data as well. It is important to know what data [Organization Name] is sharing with these tools. The next section documents what information can be shared with less concern, and what information should be discussed before deciding whether or not to share with an AI product. Tier 1 data is considered acceptable for AI tools without approval. Before Tier 2 data is entered into any AI tool, additional approval from [position title/internal committee] must be obtained.

- **Tier 1:** Less sensitive. This includes publicly available program names and descriptions, high-level budget information, etc., as well as internal draft documents (such as grant application materials).

  ◦

  ◦

  ◦

- **Tier 2:** More sensitive. This includes personally identifiable data[4] including: individual client or staff names, social security numbers, financial information tied to specific individuals, phone numbers, email addresses; information covered under non-disclosure agreements or other data-sharing agreements.

  ◦

  ◦

  ◦

If sensitive (Tier 2) data is put into generative AI tools, staff should inform their supervisor and either employ masking to prevent disclosure or other methods to limit unintended disclosure.

[3] The HIPPA Journal states that "however, organizations that are required to comply with the HIPPA are not permitted to use [generative AI] tools in connection with any ePHI unless the tools have undergone a security review and there is a signed. ChatGPT (not API) is not HIPAA compliant as per OpenAI web site."

[4] The National Institute of Standards and Technology defines personally identifiable information (PII) as "Any representation of information that permits the identity of an individual to whom the information applies to be reasonably inferred by either direct or indirect means."

Strong **data management practices** should be followed. Where possible, we will employ the following practices concerning user data[5]:

- **Regular reviews:** We will conduct routine reviews to ensure that permissions to our organization's data used in AI tools remain appropriate as our organization scales or roles change to prevent privilege creep, privileges over time that are not necessary for duties.

- **Division of duties:** We will distribute duties so that one person does not have excessive control over AI tools and the data used by them for [Organization Name]. In practice, this means that programs and operations teams, in addition to the tech team, should have a say in how data and AI tools are used in [Organization Name].

- **Principle of least privilege:** We will grant users the minimum permissions necessary to perform tasks to reduce risk and prevent accidental error,[6] and ensure access changes as employment status or position changes.

- **Principle of minimization:** We will collect and store the minimum amount of information needed to execute our organization's mission. We will not collect and store information just because we can ask for it.

- **Multi-factor authentication (MFA):** Where possible, we will use tools that give us the option of MFA to reduce the risks of unauthorized users receiving access to data.

Regular **reviews** of data and AI tools will be conducted by a committee consisting of the following roles, including technical and non-technical perspectives:

-
-
-

This committee will meet [frequency] and will document any decisions made in this meeting in [location (eg: Slack channel, shared document, folder, etc.)].

When selecting new **vendors**, we will ask vendors to adhere to the data governance practices outlined in this document. We will ask them to share, in writing, their data management guidelines for their products.

[5] FrontEgg focuses on securing end-users access experience through granular, multi-tenant authentication. They have written about MFA, Principles of Least Privilege and Regular Reviews here and we have modified their language for this practice.

[6] Cloud Security Alliance (CSA) is a not-for-profit organization with the mission to "promote the use of best practices for providing security assurance within cloud computing, and to provide education on the uses of cloud computing to help secure all other forms of computing. They have written at length about Just In Time Privilege and it's relationship with the principle of least privilege

Key security practices inspired by the NIST Framework[7] that [Organization Name] will adopt in the use of AI in our organization include but are not limited to:

- **Anonymity:** We will limit the use of sharing personally identifiable information in AI technologies.
- **Confidentiality:** We will abide by key confidentiality policies by not inputting confidential information into AI tools. Where it is unclear, we will seek consent from those who own the data.
- **Intentional data disclosure:** When we use large language models, we will make a decision about whether to opt in or out of sharing nonprofit data with the tool being adopted for the purposes of training their model.
  - For example, it may be appropriate to help train models on marketing information or information we want other nonprofit organizations to benefit from. However, it may not make sense to train models on fundraising proposals that we leverage large language models to proofread.

To increase **transparency and accountability** of AI tools used within the organization, the following information should be documented for each new AI tool developed, and this is particularly important for tools built on predictive AI:

- Origin of data
- Associated models and metadata
- Overall data pipelines for audit.
- Testing metrics

This documentation should be captured in the [Organization Name] AI Technical Resource, located [link to internal document]. The [Role] holds approval responsibility for the document.

Understanding that data can introduce bias into AI tools, [Organization Name] will work to **minimize bias** in its own data. We will do this by:

- Validating the source and accuracy of data
- Ensuring informed consent for data collection

[7] NIST has long been a leader in the privacy domain. Their values are excerpted and adapted for the nonprofit use case from pg. 17 of their Artificial Intelligence Risk Management Framework (AI RMF 1.0)

- Considering how individual data sources could be combined to reveal personally identifiable information
- Ensuring completeness of datasets
- Developing mitigation strategies for incomplete and/or inaccurate datasets
- Securing and protecting data

Recognizing that the security of our data sources matters, we will also engage in **best practices to secure our data and systems**. These practices include:

- Setting up strong data encryption protocols
- Establishing user authentication systems
- Conducting regular security audits

If AI tools begin to hallucinate with clients, or if it is discovered that an inaccurate output from an AI tool led to bad decisions about operations or in materials provided to stakeholders, the internal cross-functional AI committee should be notified within [timeframe]. If the error is deemed to be a reputational and/or operational risk, the committee will notify the board within [timeframe].

Inspired by the NIST Framework[8], we will support the development of privacy-enhancing technology. Practices we will adopt in the use of AI in our organization include:

- **De-identification**
    - Data collection practices will require the minimum amount of PII needed to achieve our mission and satisfy our agreements.
    - We will store data in our database in a way that minimizes harm in the event of a breach.
    - We will anonymize or de-identify any health data that we may handle in the course of our work, stripping away identifiable information and ensuring data cannot be traced back to specific individuals.
- **Aggregation**
    - We will present data from our database in aggregate when possible.
    - We will ensure we have the proper consent from users (constituents/consumers/anyone providing data) in place if deviations from an aggregate summary standard are needed.

[8] NIST has long been a leader in the privacy domain. Their section on PET considerations are excerpted from pg. 17 and adapted for the nonprofit use case from their Artificial Intelligence Risk Management Framework (AI RMF 1.0)

AI Tool Analysis and Development

We recognize that selecting or developing AI tools requires us to consider how the tool can positively impact our organization, the level of effort required to implement the product, and how the product will fit into our current workflows.

Key preferences

- In selecting AI products to use internally, we will prefer tools that allow us to opt out of having our data used in the product's training data.
- We will also prefer tools that have worked with non-profits or other community members and constituents.

Licenses and financial costs for tools

- We will provide licenses for our teams when it will increase privacy, accuracy, and data controls, contingent upon available financial resources.
- Financial approval for licenses should be approved by [Role].
- Personal email addresses [are/are not] not permissible for AI tools.

For **decisions about AI adoption**, we will observe the following protocols:

- **Ensure adequate review and testing.** Before adopting consumer-facing tools, we will set a deadline for when we will decide on adoption and rollout to our constituents. We will involve community members, as appropriate, and current staff who previously engaged in the task we are leveraging AI for, to ensure functionality and reliability.
- **Set adoption goals or at least one key performance indicator (KPIs) for AI adoption.** KPIs are measurable criteria that are set for a specific objective that we can point to so that will help us assess how our use of AI can be evaluated for effectiveness.
- **Monitor errors and wins to adjust our behavior.** We will monitor how often the AI tools produce errors and depending on frequency and severity, will decide to provide additional training, change how we use the tool, or discontinue use of the tool. We will share wins with the broader team to support transferrable uses across the organization.

For **decisions about AI adoption**, we will observe the following protocols:

- **Ensure real-time review.** Before launching consumer-facing tools, we will pilot them with experts to ensure functionality and reliability. For example, if we use AI to automate responses to frequently asked questions on our site, we will pilot with customer support to review answer choices for a period of time to monitor error, hallucination, or bias before minimizing human involvement in fielding questions.
- **Make it easy to transition to humans.** All tools that serve a customer support function will be designed to "fail safely," seamlessly transitioning to human intervention, or a queue for human follow-up, when necessary.
- **Set adoption goals or key performance indicators (KPIs) for AI adoption.** KPIs are measurable criteria that are set for a specific objective that we can point to so that we can assess how our use of AI can be evaluated for effectiveness.
- **Monitor errors and correct them.** We will develop the infrastructure to monitor error rates associated with our use of AI and recovery plans for issues such as hallucination, which is crucial for maintaining trust with our stakeholders and constituents.
- **Establish data quality validation.** In instances when we use predictive AI we will validate data quality to ensure that internal and external consumers understand the accuracy of predictions.

We will strategically leverage both open-source and commercial products to maximize cost-effectiveness, impact, and to prevent duplication of efforts.

For **AI-related decisions for our Cloud Infrastructure and Development, Security, and Operations**, we will observe the following practices:

- We will **utilize cloud platforms** when we have limited in-house computing resources or require intense computing capabilities. Leveraging cloud-based AI solutions will allow us to significantly enhance efficiency by handling tasks that free up staff time, allowing us to focus on more mission-aligned activities. This approach not only optimizes resource use but also ensures scalability and flexibility in managing AI workloads.
- For programs using sensitive data, we will **consider self-hosted Large Language Models** (LLMs), and when possible, we will consider 1) keeping GPT models on secure, in-house servers or 2) maintaining any extremely sensitive information within our controlled environment.
- We will engage in the following as **overall practices to support responsible AI** to support the responsible deployment of AI within our technical stack:
  - Focusing on data hygiene through regular accuracy checks

    - Curating and strategically labeling data directly related to AI training goals
    - Allowing time for testing and fine-tuning of LLMs.

- We will assemble and maintain a **skilled team to support AI maintenance** with security in mind. We will integrate a diversity of roles internally, and, when needed, supplement them with external contractors or vendors, to ensure seamless AI delivery. Prioritizing security from the start allows us to have a collaborative environment, embedding security throughout the development lifecycle rather than treating it as an afterthought. By automating security testing with cloud-based tools, we can ensure continuous monitoring by identifying and addressing vulnerabilities early while maintaining a more secure codebase.

About the authors

ANB Advisory Group

Afua Bruce and Rose Afriyie

This resource was authored by Afua Bruce and Rose Afriyie for the ANB Advisory Group. ANB Advisory Group is a consulting firm that supports a responsible tech ecosystem. We work with clients that fund, develop, or implement responsible tech and data strategies and programs, and we bring a strong equity frame to our work across all sectors. Learn more at anbadvisory.com.

nten

This resource was supported by NTEN. We are creating a world where missions and movements are successful through the skillful and equitable use of technology. We build transformative power by connecting people who are putting technology to work for social change. We strengthen their individual and collective capacity for doing good by offering expert trainings, researching effective approaches, and providing places where relationships can flourish. We relentlessly advocate for the redesign of the systems and structures that maintain inequity. Learn more at nten.org.