State of California Guidelines for Evaluating Impacts of Generative AI on Vulnerable and Marginalized Communities

December 2024

Table of Contents

Table 1: Revision History

Revision	Date	Effective Date	Summary of Changes	Author(s)
1	December 2024	January 1, 2025	Initial publication of equity evaluation framework and deployment guidance for GenAI	Government Operations Agency, Office of Data and Innovation, California Department of Technology

1. Introduction

As the nation's largest and most diverse state, California has a long history of fighting for civil rights, freedom, and equity. Governor Newsom's [Executive Order N-12-23](#) underscores the importance of responsible adoption of Generative Artificial Intelligence (GenAI) technology across California state government. As state entities evaluate the opportunities of GenAI to support their programs and missions, it is crucial to consider the potential impacts on vulnerable and marginalized communities.

Advancing equity for all Californians means taking a systematic approach to ensuring equal opportunities and fair, just, and impartial treatment for all individuals. Equitable design creates paths to equal outcomes by recognizing that those within vulnerable and marginalized communities have unequal starting points driven by different histories, historical treatment, circumstances, strengths, and needs. Considering those unequal starting points in government program design and delivery can help illuminate opportunities for targeted actions, resulting in improved outcomes for vulnerable or marginalized communities, and reducing risks that actions or adopted tools perpetuate disparities or introduce bias into decision-making.

These guidelines encourage state department teams and leaders to consider the potential impacts a GenAI tool can have on vulnerable communities, with a particular focus on safe and equitable outcomes in the deployment and implementation of high-risk use cases. These guidelines help state entities identify communities that may be uniquely vulnerable to the impacts of GenAI and risks that adoption of GenAI tools may unintentionally introduce bias into decision-making. These recommendations will empower the state workforce with guidance, resources, and support to thoughtfully and equitably adopt GenAI technology in state programs to advance the interests of all Californians. The guidelines also include Appendices with reference materials that may assist in assessing equity considerations around potential GenAI deployment and monitoring for unanticipated impacts in implementation.

These equity evaluation guidelines will be incorporated into business readiness assessments and should be considered before the GenAI Guidelines on Procurement, Use, and Training. State entities should consult the [GenAI user journey](#) on the GenAI for California website to understand how to use these guidelines when considering the use of a GenAI tool.

This document builds upon previous reports and guidelines published by the state of California regarding GenAI such as the [State of California: Benefits and Risks of Generative Artificial Intelligence Report](#) published in November 2023 and the [State of California GenAI Guidelines for Public Sector, Procurement, Uses and Training](#) published in March 2024.

The safe and responsible adoption of GenAI tools by California state government will require thoughtful policies and continued partnership with community stakeholders. The Government Operations Agency (GovOps), Office of Data and Innovation (ODI), and California Department of Technology (CDT) will continue to take an iterative and people-centered approach to refining these guidelines and ensuring that California state government is well-positioned to equitably spread the benefits and opportunities of this technology across all communities, while safeguarding our most vulnerable communities.

2. Recommended Approach for Using the Guidelines

Understanding the potential impacts of GenAI tools on vulnerable and marginalized communities is a critical step in the responsible adoption of GenAI technology in public sector services. Below are the steps state entities can take to evaluate the equity impacts of using GenAI using these Guidelines and other resources:

1. Review the [Benefits and Risks of Generative Artificial Intelligence Report](#) so staff evaluating GenAI impacts understand the key definitions and background information related to GenAI and vulnerable communities.
2. Review the **GenAI Pre-Procurement Equity Evaluation Guidelines** for potential GenAI use cases as part of business readiness assessment.
    a. Entities should consider potential positive and negative impacts of GenAI use cases:
        i.   Consider options for mitigating potential negative impacts and consult with ODI and GovOps.
        ii. Engage with potentially impacted vulnerable communities. Gather feedback, concerns, and insights about using GenAI tools in your services.
        iii. If necessary, rescope the use of the GenAI tool or consider alternative solutions.
3. Review the **GenAI Equity Evaluation Checklist** at each iterative stage of production:
    a. Before publishing a procurement solicitation.

b. At the end of the proof of concept (POC) phase.

c. Before procuring a minimum viable product (MVP) solution.

d. Before any successive project phase.

4. Continuously assess and monitor the impact of the tool on vulnerable communities. Make adjustments when needed to ensure equitable outcomes. Create feedback loops with state employee users and communities served by the tool.

5. Comply with established laws. Your GenAI use case should comply with all relevant state or federal laws, such as but not limited to Equal Employment Opportunity policies and the California Information Practices Act. Consult your legal counsel to ensure compliance.

6. Get assistance with the GenAI Equity Guidelines. If you need help applying these guidelines, you can schedule time with ODI staff. While state entities are responsible for developing the GenAI use case, ODI staff can provide advice and clarify any questions you have about the framework. As the GenAI field evolves and matures, GovOps, ODI, and CDT will update these guidelines accordingly and develop assistive resources.

3. GenAI Pre-Procurement Equity Evaluation Guidelines

These guidelines prompt state entities to:

- Identify potential equity impacts of GenAI tools on vulnerable communities
- Assess the degree and scale of these impacts
- Select ways to manage risks if a GenAI tool is adopted

Recommended Steps for Assessing Equity Impacts

1. Identify skill sets your organization needs to assess the equity impacts of GenAI on vulnerable communities.

a. This could include program leaders, information officers, data experts, legal counsel, and equity leads.

2. Review the Equity Considerations before GenAI Procurement Recommendations (see below).

3. Consider whether mitigations such as problem statement refinement, project rescoping, or data quality improvement could resolve any identified equity impacts of the proposed GenAI tool.

a. Consult with ODI and GovOps on potential mitigations for high-risk use cases.

Equity Considerations before GenAI Procurement decision

As part of deliberations around potential adoption of GenAI technology in governmental services, state agencies should consider the following issues described below.

Communities Served

Consider the current demographic groups that your program serves and the vulnerable communities that commonly engage with your program. If state entities are not sure of which communities their programs serve, consider:

- Reviewing caseload datasets, claims databases, program utilization reports, and other data sources. These may reveal the demographics of your user base.
- Looking at existing metrics for evaluating vulnerability. Examples in California state government include:
    - [CalEnviroscreen Vulnerability Index](#)
    - [Healthy Places Index](#)
    - [Transportation Equity Index](#)
- Reaching out to stakeholders and communities. This can also help identify communities that are underrepresented in your current work streams.

Communities at disproportionate risk of GenAI impacts

Consider whether particular communities served by the program may be impacted by adoption of GenAI. A general rule is to consider whether a community will see impacts to services provided to them, potential for bias, or if the group may be excluded from the benefits of the proposed deployment of GenAI. Review GenAI risks outlined in the [State of California: Benefits and Risks of Generative Artificial Intelligence Report](#) (pages 14-26). In particular, consider:

- **Marginalized or underrepresented groups.** Some groups may be underrepresented in public or training data sets, leading to biases in GenAI. This can make some groups vulnerable to skewed representation or outcomes.
- **Groups overrepresented in public datasets.** Similarly, some groups may be overrepresented in datasets, such as social media posts, website data, and government records (like criminal justice, education, or public services data), which can lead to biases in GenAI.

Additionally, as with all technology tools, including those already deployed, it is critical to understand how reliance on GenAI tools may impact access to

services, including by **individuals with disabilities who rely on accessibility tools** and **groups that already face obstacles to digital access or technology products**, known as the "digital divide".

Degree of impact on any communities identified as being at disproportionate risk of GenAI impacts

State entities should evaluate the **positive and negative impacts** to any communities identified under the preceding considerations.

- Consider the **degree of impact:**
  - How much would this impact affect someone who experienced it?
  - Can this impact likely to be reversible through mitigation?
- Consider the **scale of impact:**
  - How many people in the community are likely to experience this impact?
  - How long will this impact last?

Using these equity considerations, state entities should evaluate whether their GenAI use case potentially falls into the high-risk category as defined by the California Department of Technology's [Generative Artificial Intelligence Risk Assessment](#). For potential high-risk use cases, state entities can seek consultation services with ODI and GovOps to discuss mitigation strategies.

4.   GenAI Equity Evaluation Checklist

This checklist prompts state entities to identify and address potential biases from human interaction with GenAI outputs to promote transparency, accountability, and fairness.

Recommended Steps for Utilizing the GenAI Equity Evaluation Checklist

1. Review the GenAI Equity Evaluation Checklist (see below).
2. Consider mitigation efforts if appropriate. ODI and GovOps can provide consultation services for potential high risk use cases as needed.
3. Regularly engage with the communities you serve. Gather feedback, concerns, and insights about the use of GenAI tools in your program or service.
4. Continuously assess and monitor the impact of GenAI tools on vulnerable communities at each iterative stage of procurement and implementation. Make adjustments as necessary to ensure equitable outcomes.

5. Review the recommendations regularly as part of the tool's re-approval process. Update materials and plans as needed.

GenAI Equity Evaluation Checklist

Government entities using GenAI should consider the following before deploying GenAI systems:

- Accountability
  - Are there clearly assigned roles and responsibilities for data governance, model deployment, monitoring, compliance, and risk management related to the proposed GenAI tool?
  - Do all project staff understand their obligations and have the necessary skills and resources to fulfill them?
- Fairness
  - Has the state entity confirmed that the proposed GenAI tool's performance metrics are equitable across communities (including protected characteristics under federal and state nondiscrimination laws, e.g., race, age, gender, disability)?
  - Is there risk of the GenAI model considering attributes strongly correlated with protected characteristics (such as using geographic zip code data that may be strongly correlated to a protected class like race)?
- Human Oversight
  - Will there be regular and ongoing human-in-the-loop oversight over the GenAI tool?
  - Will staff be trained to interpret the GenAI tool's outputs, recognize potential biases or errors, and override the system when necessary?
  - Are there clear guidelines and thresholds for when human intervention will be required, and will human operators have the authority to make final decisions overriding the GenAI tool?
- Transparency
  - Can Californians access information about how their data will be used by the GenAI tool in plain language and in-language? This can include details on what data is collected, how it is processed, and for what purposes it will be used
  - Will there be a plain language notice provided when GenAI is used in creating or recommending outputs that may impact people?

5. Community Engagement

State entities should consider engaging communities that have been identified as potentially being negatively impacted by the use of a GenAI tool before deploying a GenAI tool. State entities should consider, if appropriate:

- A plan for education, outreach, and awareness with community groups and members.
- Identifying trusted messenger networks or partners to support community communications.
- Determine how community consultation and feedback will inform decisions of whether and how the GenAI tool will be deployed.

Such outreach can improve trust in governmental services. See Appendix B for resources on community engagement strategies.

6. Appendices

Appendix A: Human-Machine Bias Reference Table

The consequences of inappropriate reliance on GenAI in government can be serious: poor decision-making, ineffective human oversight, and erosion of public trust. It is not enough for the staff SME and GenAI tool to perform well together; staff should be able to identify and mitigate problematic GenAI behaviors.

Having human reviewers provide their own assessments before seeing the GenAI tool's output can help prevent anchoring bias, when the reviewer's judgment is unduly influenced by the GenAI tool's suggestion. This approach ensures that the human's independent expertise is fully utilized and not overshadowed by the GenAI. However, in cases where the GenAI is screening large volumes of inputs, it may be more efficient for the human to review only those cases flagged by the GenAI as needing further attention.

The following table discusses five common cognitive biases that should be addressed in GenAI system design at the end of each Project Approval Workbook chapter, including any iterative project build phases.

Table 2: Review of common cognitive biases when using GenAI tools

Cognitive bias	Definition	Concern	Example
Automation bias	The tendency to favor recommendations from automated systems over information from non-automated sources, even when the automated suggestions might be wrong.	Are users over-relying on GenAI recommendations without adequately considering other relevant information?	In a GenAI system that assists with medical diagnosis, a physician may overly rely on the GenAI tool's suggested diagnosis without thoroughly examining the patient's symptoms and medical history.

Cognitive bias	Definition	Concern	Example
Confirmation bias	The tendency to search for, interpret, favor, and remember information in a way that confirms your preexisting beliefs or hypotheses.	Are users selectively focusing on GenAI recommendations that support their personal beliefs, and ignoring GenAI recommendations that don't support their beliefs?	In a GenAI system that predicts recidivism risk for parole decisions, a parole board member may give more weight to the GenAI recommendations that confirm their preconceived notions about an offender's likelihood to reoffend.
Selection bias	The bias introduced by the non-random selection of data or individuals, which can lead to a skewed representation of the population and affect the conclusions drawn from the data.	Is the data used to train the GenAI system representative of the population it will be applied to, and are there any systematic exclusions or overrepresentations?	In a GenAI system that assists with medical diagnosis, a physician may overly rely on the GenAI tool's suggested diagnosis without thoroughly examining the patient's symptoms and medical history.
Availability bias	The tendency to overestimate the likelihood of events with greater "availability" in memory, which can be influenced by how recent the memories are or how unusual or emotionally charged they may be.	Are human decision-makers overly influenced by memorable or recent examples when evaluating the GenAI tool's recommendations?	In a GenAI system that predicts recidivism risk for parole decisions, a parole board member may give more weight to the GenAI recommendations that confirm their preconceived notions about an offender's likelihood to reoffend.

Cognitive bias	Definition	Concern	Example
Anchoring bias	The tendency to rely too heavily on the first piece of information offered (the "anchor") when making decisions.	The tendency to rely too heavily on the first piece of information offered (the "anchor") when making decisions.	In a GenAI system that assists with medical diagnosis, a physician may overly rely on the GenAI tool's suggested diagnosis without thoroughly examining the patient's symptoms and medical history.

Appendix B: Recommendations on Community Consultation

Key Principles for Effective Community Engagement

- **Inclusivity and representation:** Ensure participation from diverse stakeholders. Make sure to include people from marginalized or underrepresented communities who may be disproportionately affected by AI systems.
- **Transparency:** Provide clear, accessible information about the proposed AI tool, its intended use, and potential impacts. Share details of the engagement process and how community input will be used.
- **Accessibility:** Remove barriers to participation by:
  - Offering multiple engagement formats
  - Providing necessary resources, like childcare and transportation
  - Accommodating different languages and abilities
- **Power sharing:** Move beyond token consultation to meaningful collaboration. Give community members genuine influence in decisions.
- **Continuous engagement:** Make community engagement an ongoing process instead of a one-time event. Allow for iterative feedback and long-term relationship building.
- **Capacity building:** Enable more participation through education and resources to help community members understand AI technologies and their implications.
- **Accountability:** Establish clear mechanisms for incorporating community feedback. Report back on how input influenced decisions.

Recommended structure on community consultation model

Phase 1: Pre-Engagement Planning

Strategize on who to reach out to, the timeline of the process, and key groups to partner with as trusted messengers.

Recommended Key Activities

- Reach out to local community groups, churches, schools, and neighborhood associations. Ask them to partner with you in spreading the word, organizing meetings, and sharing explainers and surveys with community members.
- Write a clear, simple explanation of the GenAI tool you are considering and why you want to use it. Train your staff on how to talk about GenAI in simple terms. Follow the [plain language equity standard provided on the California Innovation hub](#).
- Create simple handouts or videos explaining GenAI and the specific tool you're considering. Use real-life examples and provide these handouts in multiple languages when possible.

Key Questions

- Who are we missing?
- Which groups might be hard to reach?
- What do we want to learn from the community?
- What aspects of GenAI are people most confused about?
- What are people's biggest interests and concerns about GenAI?

Phase 2: Active Engagement and Deliberation

Dive deep with community members on the details and intended use of the GenAI tool.

Recommended Key Activities

- Run small group discussions (8-12 people) with people from vulnerable communities. Ask people how they think the GenAI tool might affect their daily lives. Consider the [community jury structure outlined by Microsoft](#).
- Set up an online survey to get people's thoughts and concerns on an ongoing basis. Make sure it is mobile-friendly.

Key Questions

- What specific concerns do people have about privacy, fairness, accessibility, or labor impacts?
- Are there benefits to the tool that interest community members?
- How do different groups (like elderly, youth, different ethnicities) view the GenAI tool differently?

Phase 3: Feedback Synthesis and Decision Making

Synthesize the input you gathered, come to a decision on if and how to move forward with the GenAI tool, and keep the community involved over time.

Recommended Key Activities

- Read through all the comments, survey responses, and meeting notes. Look for common themes and summarize what you heard from the community in a report.
- Make a list of recommended changes to the GenAI tool based on community input. Share a draft of your recommendations publicly.
- Announce your final decision about the GenAI tool publicly. Explain how community input shaped the decision.
- Create channels for community members to identify or appeal problems that emerge in the GenAI tool.

Key Questions

- How can we keep community members engaged over the long term?
- How can we give community members enough information on the GenAI tool's performance to be active and engaged in monitoring?

Appendix C: References

1. A Sociotechnical Approach to AI Policy
2. AI Ethics and Governance in Practice: AI Fairness in Practice
3. Algorithmic Accountability for the Public Sector - Report - AI Now Institute
4. Algorithmic Accountability: Moving Beyond Audits - AI Now Institute
5. Algorithmic Bias Explained
6. Algorithmic Discrimination Protections | OSTP | The White House
7. Algorithmic Impact Assessment: AIA template | Ada Lovelace Institute
8. Appropriate reliance on Generative AI: Research synthesis
9. Applying Sociotechnical Approaches to AI Governance in Practice
10. Assembling Accountability: Algorithmic Impact Assessment for the Public Interest
11. CIO Council: Algorithm Impact Assessment
12. City of San Francisco: Ethics & Algorithms Toolkit
13. Democratizing AI: Principles for Meaningful Public Participation
14. Datasheets for Datasets
15. Enrolling Citizens: A Primer on Archetypes of Democratic Engagement with AI
16. FACT SHEET: Vice President Harris Announces OMB Policy to Advance Governance, Innovation, and Risk Management in Federal Agencies' Use of Artificial Intelligence
17. Framework for Identifying Highly Consequential AI Use Cases | Johns Hopkins University
18. Future of Jobs Report 2023
19. Government AI Coalition | City of San José
20. Government of Canada: Algorithm Impact Assessment Tool
21. Identifying Sociotechnical Harms of Algorithmic Systems: Scoping a Taxonomy for Harm Reduction
22. Measuring Justice: Field Notes on Algorithmic Impact Assessments
23. Microsoft Responsible AI Impact Assessment Template
24. Model AI Governance Framework for Generative AI: Fostering a Trusted Ecosystem
25. NIST AI Risk Management Framework
26. NIST Artificial Intelligence Risk Management Framework: Generative Artificial Intelligence Profile
27. NTIA Artificial Intelligence Accountability Policy Report
28. Overreliance on AI: Literature Review