

# **Policy Analysis: City-Level Artificial Intelligence Policies and Guidelines**

**Supplement to *Developing Artificial Intelligence Policies for Public Health Organizations: A Template and Guidance***

***Health and Human Services Region 7***

**January 2025**



## About This Initiative

*Developing Artificial Intelligence Policies for Public Health Organizations: A Template and Guidance* is a collaboration between the Kansas Health Institute (KHI), Health Resources in Action (HRiA) and the Wichita State University Community Engagement Institute (WSU CEI). The project scope included an environmental scan that informed the template's development, comprising a literature review and policy analysis at both state and city levels. The role of each organization varied by project component. This document, *Policy Analysis: City-Level Artificial Intelligence Policies and Guidelines*, was developed by the Kansas Health Institute.



## Author

- **Tatiana Y. Lin, M.A.**, Director of Business Strategy and Innovation, KHI

## Contributors

- **Emma Uridge, M.P.H.**, Analyst, KHI
- **Shelby C. Rowell, M.P.A.**, Analyst, KHI
- **Linda J. Sheppard, J.D.**, Team Leader, KHI

The author and contributors also would like to thank Health Resources in Action and the Wichita State University Community Engagement Institute for providing assistance in identifying policies to be included in the review.

## Acknowledgement

*This work is supported by funds made available from the Centers for Disease Control and Prevention (CDC) of the U.S. Department of Health and Human Services (HHS), National Center for STLT Public Health Infrastructure and Workforce, through OE22-2203: Strengthening U.S. Public Health Infrastructure, Workforce, and Data Systems grant. The contents are those of the author(s) and do not necessarily represent the official views of, nor an endorsement by, CDC/HHS or the U.S. Government.*

## Disclosure Statement

*During the policy review process, the research team utilized AI tools, specifically Petal and ChatGPT 4.0, to identify search terms, locate relevant policies and support the development of policy summaries and cross-policy analyses. These tools were used to generate initial drafts of the summaries. All content was subsequently reviewed and refined by the authors to ensure accuracy and quality. The authors take full responsibility for the final content presented in this document.*

## Table of Contents

Introduction .....	1
Structure of the Document .....	2
Methodology .....	2
Overview of Reviewed Policies and Guidelines .....	4
Scope of Use .....	5
Summary of Provisions .....	5
Acquisition of AI Tools .....	5
Bias Mitigation .....	7
Community Engagement .....	9
Copyright and Authorship .....	10
Data Privacy .....	11
Environmental Impact .....	13
Evaluation and Quality Improvement .....	13
Human Oversight .....	14
Impacted Populations .....	16
Non-Compliance and Penalties .....	18
Training and Capacity Building .....	19
Transparency .....	20
Unauthorized Use and Prohibitions .....	21
Appendix A. Methodology Protocol for Completing Steps in Using ChatGPT for Policy Analysis .....	A-1
Appendix B: Bibliography .....	B-1

*(This page intentionally left blank.)*

## Introduction

In 2024, the Kansas Health Institute (KHI), Health Resources in Action (HRiA) and Wichita State University Community Engagement Institute (WSU CEI) collaborated on a project which resulted in the document titled *Developing Artificial Intelligence Policies for Public Health Organizations: A Template and Guidance* (referred to hereafter as *AI Template and Guidance*). The document is designed to assist public health organizations, including nonprofits and government agencies at all levels, in creating policies or guidelines that facilitate ethical experimentation with artificial intelligence (AI) systems while addressing potential risks and promoting health equity and innovation.

To inform the development of the *AI Template and Guidance*, the research team conducted an environmental scan focused on considerations surrounding the use of AI, specifically identifying what should be included in the policies for public health organizations. This scan included a review of AI-related literature, state-level policies that were introduced or passed, and policies or guidelines passed by cities.

The primary goal of this document, *Policy Analysis: City-Level Artificial Intelligence Policies and Guidelines*, is to identify key themes and common provisions that are found within current local-level AI policies to inform the development of *AI Template and Guidance*. To achieve this, the analysis aimed to answer the question: *What are the common components and provisions found in AI policies across different cities, with a specific emphasis on equity, ethical considerations, transparency, human oversight and data privacy?* The analysis drew on a review of formal guidelines and policies from 14 cities regarding the responsible use of AI technologies.

The findings from this document, along with provisions included in city policies and guidelines—either verbatim or with modifications—served as the foundation for the *AI Template and Guidance*. To enhance its robustness, findings from the state policy analysis were incorporated to modify provisions. Lastly, to ensure alignment with best practices and current research, the literature review findings were used to validate, refine, remove, or propose new provisions. Additionally, the literature review findings were used to develop sections of the template that explain why specific issues are important to include in the policy and provide rationale for their inclusion.

## Structure of the Document

The document begins with an overview of the policy analysis methodology, providing readers with a clear understanding of the approach taken. Following this, the Summary of Provisions section organizes the content into 13 topics, including Acquisition of AI Tools, Bias Mitigation, Community Engagement, Copyright and Authorship, among others. For the full list of topics, please refer to the Methodology section of the report.

Within each topic, the provisions from relevant policies are summarized under subtopics, offering detailed descriptions of the types of information they encompass. The document does not include a conclusion, as its primary purpose is to present comprehensive information to support the development of *Developing Artificial Intelligence Policies for Public Health Organizations: A Template and Guidance* (referred to hereafter as *AI Template and Guidance*). The document concludes with an overview of the ChatGPT protocol used to conduct the policy analysis (*Appendix A*, page A-1), along with a bibliography of the included policies, complete with links for reference (*Appendix B*, page B-1).

## Methodology

The purpose of the analysis of local policies was to identify common provisions used in policies related to AI topical areas, such as bias mitigation, human oversight, data privacy and environmental impact, among others. See below for the full list of topics examined. The focus areas were determined based on issues typically raised by the public health community and cited in both media and research.

The findings from the review also were essential for establishing a foundational basis for the document *Developing Artificial Intelligence Policies for Public Health Organizations: A Template and Guidance*. The review focused on policies enacted at the local level within the past five years in the United States, covering a range of entities and policy categories.

### List of Examined AI Topical Areas

1. Acquisition of AI Tools
2. Bias Mitigation
3. Community Engagement
4. Copyright and Authorship
5. Data Privacy

6. Environmental Impact
7. Evaluation and Quality Improvement
8. Human Oversight
9. Impacted Populations
10. Non-Compliance and Penalties
11. Training and Capacity Building
12. Transparency
13. Unauthorized Use and Prohibitions

The documents were identified through multiple methods, including outreach via social media platforms such as LinkedIn, newsletters and emails to state and national partners. This outreach included organizations such as the National Association of City and County Health Officials, the Association of State and Territorial Health Officials and the National Network of Public Health Institutes, among others.

The team also conducted online searches using keywords such as "AI policy and city" and "AI guidelines and city." Several policies were identified by determining which local governments participated in the [MetroLab](#) GenAI for Local Government Task Force. After identifying these jurisdictions, the authors conducted further searches on their official websites to locate any MetroLab Network is a collaborative organization that partners city governments and universities to drive urban innovation through data-driven solutions, applied research and advancements in technology and public policy.

## **Analysis**

The policies were downloaded, reviewed and detailed in a table referred to as the evidence table, which captured information such as jurisdiction, the entity that enacted the policy, and detailed provisions related to the examined 13 topics, which are listed on page 2. The twenty-nine columns included in the evidence table can be found in *Appendix A*, page A-1. The evidence-based table, which is part of the policy analysis approach, provides detailed findings for each examined policy across the specified categories. The research team utilized ChatGPT 4.0 for policy review and analysis. The reasons for using ChatGPT were to evaluate the feasibility of using this tool for policy analysis and to facilitate the examination of all the policies across 29 different dimensions.

The ChatGPT 4.0 was specifically used for several purposes, such as to identify policies and guidelines based on set search parameters, to summarize policies for inclusion in the evidence table and to support the creation of final summaries across all policies/guidelines by topic.



The research team prioritized a human-in-the-loop approach in the quality assurance process, which was implemented throughout the review. Outputs were examined and validated against the policies and guidelines to ensure accuracy and reliability. For more information about the process of using ChatGPT 4.0 for policy analysis, see *Appendix A*, page A-1.

## Overview of Reviewed Policies and Guidelines

In total, 14 documents from 14 cities were identified, addressing the responsible use and regulation of AI technologies in city governance, representing a range of formal guidelines and policy documents. The documents included resolutions, executive orders, regulations, guidelines and policies, all of which are referred to throughout this document simply as “policies.”

The AI-related policies were dated from 2021 to 2024, with various authorities — including city councils, mayors and specific departments — overseeing the establishment of these policies and guidelines. *Figure 1* shows the locations of jurisdictions with policies that were included in this review.

*Figure 1. Select Jurisdictions with AI Policy or Guidelines*



*Source: Kansas Health Institute review of websites for local AI policies.*

## ***Scope of Use***

The policies listed typically apply to a range of individuals and entities connected to municipal and governmental operations, particularly in the context of AI systems and technology use. These individuals and entities include, but may not be limited to, city employees, contractors and other personnel, including department leaders, consultants and volunteers involved in AI-related projects. Policies also can apply to city departments, agencies, third-party vendors and stakeholders who engage with AI technologies in various capacities.

Some policies specifically extend to employment agencies or organizations within a particular city that use automated tools for hiring and promotion decisions. Additionally, government agencies at the municipal level, including staff and key officers responsible for technology, security, privacy and contracting, must adhere to these guidelines.

Policies also may apply to third-party providers or entities that have access to the city or district's information systems, networks and applications, specifically in cases where enterprise services are received from technology departments.

## **Summary of Provisions**

### ***Acquisition of AI Tools***

This section outlines the provisions within the reviewed policies that were related to the acquisition of AI tools and models, focusing on ethical procurement, oversight, transparency and risk management across various policies.

### **Ethical Procurement and Oversight**

The identified policies note the importance of implementing ethical standards in the procurement of AI tools and models. City departments and agencies are often required to obtain written approval from information technology (IT) departments or designated officials before acquiring AI tools. This includes ensuring that AI systems are explainable, auditable and aligned with public records and data retention requirements. Additionally, the establishment of an internal oversight team, or governance board, is recommended to evaluate responsible AI practices, oversee risk management and ensure ethical use.

## **Regular Audits and Evaluations**

Regular audits and evaluations of AI systems are a key provision across various policies. These audits are designed to ensure compliance with legal and ethical standards, including bias audits for employment decision tools. The results of these audits must be publicly available, ensuring transparency and accountability in the use of AI. Some policies also mandate ongoing evaluations through cybersecurity reviews and equity lenses to address any biases or potential risks.

## **Engagement and Public Transparency**

Provisions emphasize the importance of engaging with advisory boards, cybersecurity officers and the broader community when acquiring AI tools. Agencies are encouraged to consult internal experts, such as chief information security officers (CISOs) and privacy officers, and disclose AI usage in significant communications, particularly in public-facing applications. Transparency is further promoted by documenting how AI models are used, labeling AI-generated content and ensuring that the public is informed about the tools in use and their functionalities.

## **Error Handling and Accountability**

Policies outline processes for investigating and addressing errors made by AI systems. Provisions ensure that city employees are not held accountable for errors beyond their control when using AI tools. This helps in maintaining a clear line of accountability while also protecting employees from the unpredictable nature of AI outputs.

## **Fact-Checking and Validation**

Many provisions stress the need for fact-checking all content generated by AI, especially when it is used in public communications or decision-making processes. AI-generated content must be thoroughly reviewed for accuracy and verifiable sources should be cited for any factual data. Users also are advised to avoid sharing sensitive or private information in AI prompts, protecting against data breaches and ensuring compliance with privacy laws.

## **Compliance with Procurement and Security Policies**

The policies require strict adherence to standard procurement procedures and cybersecurity standards. Agencies must conduct due diligence when considering vendors for AI tools, ensuring compliance with citywide cybersecurity policies. This includes performing risk

assessments to evaluate potential risks related to data protection, biases and decision-making processes. Provisions also advise departments to opt out of data collection when using AI platforms and restrict AI use if non-compliance with policy or laws is detected.

### **Restrictions and Special Conditions**

Some policies impose restrictions on the types of AI tools that can be acquired, such as prohibiting the use of non-contracted generative AI systems unless approved. Non-contracted in this context means AI tools that haven't been officially approved or purchased through a formal agreement with the organization. In certain cases, agencies are required to use specific AI tools authorized by IT departments and exceptions can only be made after thorough review. This ensures that AI systems meet the required privacy, security and ethical standards before they are integrated into public services.

### **Training and Capacity Building**

To ensure safe and responsible AI usage, some policies mandate that employees receive training on the tools and models they are working with. This training helps employees understand the potential risks, ethical considerations and best practices in using AI, ensuring that AI is deployed effectively and responsibly.

### ***Bias Mitigation***

These provisions aim to reduce bias in AI systems through comprehensive evaluation, independent verification of AI outputs, maintaining transparency and promoting inclusion while safeguarding against the inappropriate use of sensitive data. Provisions also highlight the critical role of human oversight and continuous bias audits to ensure AI technology is used ethically and fairly.

### **Evaluation and Detection of Bias**

Many of the provisions note the importance of evaluating AI systems for bias before deployment, particularly through tools like bias audits. For example, before purchasing or deploying AI tools, organizations are encouraged to assess for potential discrimination, algorithmic bias or unintended harms. This proactive evaluation often includes considering training data and sampling biases, with a focus on preventing algorithmic discrimination against protected classes.

## **Content Fact-Checking and Validation**

Several policies highlight the need for fact-checking AI-generated content. Users are instructed to review outputs for inaccuracies, bias and offensive or harmful material, particularly in contexts of public communication or decision-making. This includes verifying claims, citations and links produced by AI tools, ensuring they are valid and reliable. Additionally, users are warned that AI systems may produce outputs that reflect social, cultural or economic biases, thus necessitating independent validation through trusted sources or subject matter experts.

## **Transparency and Disclosure**

Transparency in AI usage is another key theme. Policies encourage users to disclose when AI has been used in content generation, particularly in public-facing communications. This transparency extends to documenting how AI tools were used, including details such as model types and prompts, to improve understanding and safety. Some provisions also stress the importance of publicizing results of bias audits to ensure accountability, particularly in employment decision tools.

## **Privacy and Security**

To mitigate bias, provisions stress the importance of adhering to privacy and security best practices. Users are cautioned against sending sensitive, personal or confidential information to AI systems, ensuring that the information complies with privacy laws such as Maryland's Public Information Act. In the case of public records or employment screening tools, ensuring that candidates or employees are notified of the use of AI is mandated to maintain transparency.

## **Inclusive Language and Representation**

Many of the guidelines aim to promote fairness and avoid reinforcing stereotypes. For instance, job postings or AI-generated content are reviewed to eliminate language that may be associated with specific races or genders. Provisions encourage the use of gender-neutral terms and inclusive vocabulary, while avoiding stereotypical adjectives that could perpetuate biases. This also extends to ensuring that AI outputs represent diverse communities accurately.

## **Human Judgment and Empathy**

The policies recognize that AI systems lack emotional intelligence and human discretion. Users are urged to complement AI outputs with human empathy, judgment and personal contact,

particularly in decision-making processes. The reliance on AI should not replace essential human values, such as fairness and equity and any AI-generated outputs that could impact people's lives should be scrutinized with human oversight.

### **Bias Audits and Regular Monitoring**

In certain contexts, bias audits of AI tools, such as employment decision systems, are required. These audits should be conducted regularly (e.g., annually) to ensure ongoing compliance with anti-bias commitments. Moreover, continuous monitoring and evaluation of AI systems are advocated to ensure fairness across various user groups, with a particular emphasis on marginalized communities.

### **Engagement with Marginalized Communities**

Some provisions call for direct engagement with underrepresented or marginalized communities to better understand the potential impact of AI-generated content. This ensures that the voices of these communities are included in the evaluation process, helping to mitigate representational harms and promote inclusivity.

### ***Community Engagement***

Only a few policies explicitly address the topic of community engagement. These policies stress the importance of involving community members in decision-making processes, particularly when deploying AI technologies. The provisions related to community engagement are designed to foster transparency, inclusivity and ethical considerations by gathering input from affected or interested groups. Below is a breakdown of the provisions related to community engagement.

### **Involvement in Decision-Making**

Community engagement includes actively involving local residents, organizations and stakeholders in key decisions, particularly those that directly impact them. Some provisions highlight the need for consulting advisory boards and community members to ensure their perspectives are included in AI-related decisions. For instance, convening with advisory committees on multiple occasions is a necessary step to foster ongoing collaboration and feedback.

### **Engagement with Affected Groups**

Many policies and guidance suggest reaching out to communities that may be impacted by AI technologies, particularly marginalized groups. Engagement with these communities can assist with identifying potential issues in AI systems, such as bias or discriminatory content. Involving organizations that represent these groups helps ensure that the development and deployment of AI are carried out in a way that is respectful and equitable.

### **Public Disclosure and Transparency**

In some cases, policies mandate the disclosure of information to the public, particularly when AI systems, such as automated employment decision tools, are used. Employers are required to inform candidates about the use of these tools, explain how data are being collected and offer transparency about job qualifications and retention policies. These steps are intended to enhance community awareness and provide a channel for feedback from those directly affected by AI technologies.

### **Ethical Considerations and Knowledge Sharing**

The provision encourages engaging with a broader range of stakeholders, including policymakers and experts, to address ethical challenges posed by AI. This collaborative approach is designed to ensure that AI technologies benefit society as a whole and that any potential harms are identified early through community input.

### ***Copyright and Authorship***

The provisions related to copyright and authorship emphasize the need for proper attribution, thorough fact-checking and legal compliance when using AI-generated content. These policies hold humans accountable for AI-generated material and highlight the importance of respecting copyright laws and ethical standards in AI usage.

### **Attribution and Responsibility for AI-Generated Content**

Provisions emphasize the importance of providing proper attribution whenever AI-generated content is published. This includes acknowledging the use of generative AI tools for creating text, images or other forms of media, typically through references or footnotes. Staff and users are held accountable for all AI-generated content and the responsibility cannot be deferred to



the software or system, ensuring that human oversight remains central in the publication process.

### **Copyright Compliance**

Several policies highlight the importance of ensuring that AI-generated content does not infringe on existing copyright laws. Users are responsible for verifying that the content they publish is free from copyright violations. Specific attention is drawn to the U.S. Copyright Office's stance that works created solely by AI are not eligible for copyright protection unless there is significant "human authorship" involved.

### **Legal Implications and Ethical Standards**

In addition to copyright concerns, some provisions address the broader legal and ethical responsibilities of AI use. For instance, content that violates copyright, ethical standards or laws related to illegal activities is strictly prohibited. Agencies are encouraged to consult with designated officials, such as an algorithmic tools liaison, to ensure compliance with relevant laws, including local open data laws or freedom of information regulations.

### ***Data Privacy***

The provisions related to data privacy and security prioritize the protection of sensitive and personally identifiable information (PII), with city employees advised against uploading or sharing confidential information, especially when utilizing generative AI tools. Policies emphasize the careful handling of data such as names, birth dates, addresses and financial details. Additionally, there is a strong focus on ensuring that data shared publicly adheres to disclosure standards, with confidential city information managed to prevent unauthorized access or breaches. In cases where AI tools do not restrict the use of city data for staff training, they are expressly prohibited to maintain data security and privacy integrity.

### **Data Privacy and Security**

A key focus across multiple policies is the protection of sensitive and personally identifiable information (PII). City employees are instructed not to upload or share private, confidential or protected information, particularly when using generative AI tools. This includes data like names, birth dates, addresses, financial details and any information protected by state laws. There is also a strong emphasis on ensuring that any data shared aligns with public disclosure standards and that confidential city information is handled with care to prevent unauthorized



access or breaches. In some cases, AI tools that do not adequately prevent the use of city data for training purposes are explicitly prohibited.

### **Transparency and Accountability**

Employees using AI technologies are expected to take full responsibility for the content generated by AI tools. The policies repeatedly stress that AI-generated outputs must be verified for accuracy, fact-checked and validated before being used or published. Staff also must disclose when and how generative AI was used, citing or referencing AI-generated content where applicable. This transparency is seen as vital to maintaining public trust and accountability in AI's decision-making processes.

### **Usage Guidelines and Safeguards**

Many of the provisions establish strict guidelines for using AI tools, particularly generative AI systems like ChatGPT or Bard. Employees are advised not to use sensitive data in prompts, particularly on publicly accessible AI platforms and to consult IT or privacy officers when there is a need to use sensitive information in specific AI applications. Additionally, policies often require the use of anonymized or de-identified data for AI experimentation and stress the importance of opting out of data collection where possible to minimize the risks associated with data breaches.

### **Compliance with Legal and Ethical Standards**

Several provisions emphasize compliance with existing laws and regulations governing data privacy, including federal and state laws, as well as citywide policies. The need for strong safeguards, secure data storage and careful handling of PII is stressed, with a view to minimizing the risk of unauthorized access or disclosure. Furthermore, some policies highlight the importance of aligning AI use with the city's overarching privacy principles, consulting agency privacy officers before implementation and ensuring adherence to policies governing the use of identifying information.

### **Innovation and Collaboration**

While promoting responsible AI use, the provisions encourage innovation and collaboration between different city departments. They highlight the need for a balance between leveraging AI's potential for innovation and adhering to strict privacy and security guidelines.

## ***Environmental Impact***

This section outlines the provisions related to the environmental impact of AI, focusing on sustainability and responsible use across various policies. It is important to note that only a few policies included specific provisions related to the environmental impact of AI, reflecting a growing but still limited awareness of AI's ecological consequences.

### **Acknowledgment of AI's Environmental Impact**

Some provisions acknowledge that the use and experimentation with AI technologies can have environmental consequences, such as increased power consumption and greenhouse gas emissions. This recognition is important in fostering a culture of responsible experimentation, where AI usage is balanced with an understanding of its potential environmental costs.

### **Sustainability in AI Innovation**

Provisions emphasize the need for responsible exploration and evaluation of AI technologies that contribute positively to both public services and the environment. This includes a commitment to sustainability as part of the broader goal of using AI to advance beneficial outcomes for both people and the planet. Agencies are encouraged to pursue AI innovations that not only meet residents' needs but also reduce environmental impact.

### **Purposeful and Accountable AI Use**

Provisions promote a purposeful approach to AI experimentation, ensuring that the environmental impacts of AI tools are considered. This involves being mindful of how AI tools consume energy and resources and encouraging accountability in their use to mitigate negative environmental consequences.

## ***Evaluation and Quality Improvement***

This section outlines the provisions related to the evaluation and quality improvement of AI systems, focusing on transparency, continuous improvement and compliance across various policies.

### **Continuous Learning and Improvement**

Provisions emphasize the importance of continuous learning and improvement of AI systems. Regular audits, ongoing monitoring and oversight mechanisms are required to evaluate the fairness, safety and effectiveness of AI models. Cities and departments within cities are

encouraged to engage in continuous evaluation processes to improve AI systems and stay aligned with technological advancements and ethical standards.

### **Validity, Reliability and Security**

Provisions stress that AI systems must perform consistently under expected conditions. To maintain reliability, AI outputs should be regularly validated and tested for accuracy throughout the lifecycle of the system. AI models also must be governed and auditable to ensure that they operate correctly. In addition, provisions call for evaluating the resilience of AI systems, securing critical data and maintaining confidentiality to minimize security risks.

### **Public Engagement and Feedback**

Some policies encourage public engagement as part of the AI evaluation process. City employees are urged to collect feedback from residents to better understand concerns and improve AI systems used in public operations. Transparency and community engagement ensure that AI systems align with public expectations and ethical standards.

### **Documentation and Reporting**

Provisions also include requirements for documenting AI usage, including details about the version of the AI model and the prompts used. Regular reporting and audits help identify areas for improvement and ensure compliance with both internal policies and external laws.

### ***Human Oversight***

The provisions across various policies stress the critical role of human oversight in ensuring ethical AI usage. They emphasize human responsibility for AI-generated content, the necessity of reviewing and validating AI outputs, the authority to override AI decisions and the importance of governance structures to monitor AI systems. These measures ensure that AI systems are used responsibly and that human judgment remains central to decision-making processes.

### **Collaboration Between AI and Human Oversight**

Many provisions emphasize the importance of collaboration between humans and AI systems, ensuring that AI plays a supplemental role in decision-making processes. Human employees are encouraged to leverage AI's strengths, but the ultimate decision-making and control must remain with humans. This collaboration enhances the effectiveness and ethical use of AI systems while ensuring human judgment remains central.

### **Human Responsibility and Accountability**

Provisions highlight that human operators are responsible for all outcomes generated by AI systems, including generative AI tools. Users must validate AI outputs, particularly in public-facing communications or decision-making contexts. The responsibility for ensuring the accuracy of AI-generated content cannot be deferred to the software, making human oversight critical in identifying errors or AI hallucinations. Hallucination is the phenomenon in which generative AI systems produce an output that is nonsensical or untrue in response to the provided input or prompt. This phenomenon is further exacerbated when the model presents the incorrect or unverified information with high confidence and unsubstantiated authority. Additionally, staff must provide attribution when using AI-generated content, ensuring transparency and ownership of the material.

### **Review and Validation of AI-Generated Content**

Most policies require a thorough review of AI-generated content before it is published or used in official capacities. This includes fact-checking content to eliminate inaccuracies, bias or harmful material. Provisions often require personnel to verify AI outputs through multiple sources, especially when AI is used in sensitive contexts such as public communication, decision-making or determining resident services. AI outputs should be validated to ensure they meet the expected standards of accuracy and compliance with ethical guidelines.

### **Human Alternatives and Intervention**

Provisions outline that AI systems must include human intervention options, particularly when AI fails to meet expectations. City employees are expected to have the authority to override AI systems when necessary and implement corrective solutions. This ensures that human judgment can correct AI errors and provides transparency in decision-making processes. Timely human intervention is essential for addressing AI challenges and maintaining public trust.

### **Oversight and Governance Structures**

Policies call for the creation of governance boards or committees responsible for overseeing the ethical use of AI systems and managing associated risks. These oversight teams ensure that AI systems undergo regular audits and evaluations, allowing for ongoing human monitoring of AI's performance and compliance with ethical standards. Governance structures also include logging and documenting human review processes, such as Human-in-the-Loop (HITL) systems, which ensure AI outputs are consistently reviewed before being used in official capacities.

## **Prohibitions and Restrictions on Full Automation**

Provisions generally prohibit fully automating decisions, particularly in critical areas such as determining resident services or benefits. Human review is mandated to prevent AI from making autonomous decisions without oversight. Additionally, there are restrictions on the use of AI technologies that can manipulate or mimic human appearances or voices, highlighting concerns about deepfakes and ethical boundaries in AI applications. A deepfake is a specific kind of synthetic media where a person in an image or video is swapped with another person's likeness.

## **Confidentiality and Data Protection**

Some provisions emphasize the importance of human oversight in protecting sensitive data used by AI systems. Employees are tasked with ensuring that AI systems do not compromise confidential information and that data shared with AI tools is handled responsibly. This includes avoiding the sharing of sensitive or private information in AI prompts, particularly in public platforms where the data might be used to improve AI systems.

## ***Impacted Populations***

This section outlines the provisions related to specific populations and demographics, focusing on fairness, equity and the reduction of bias across various policies. These provisions aim to ensure that AI systems do not disproportionately harm or discriminate against certain groups based on characteristics such as race, gender, age or disability status.

## **Equity and Anti-Discrimination**

Several provisions emphasize the importance of evaluating AI systems through an equity lens to reduce discrimination and unintended harm. This involves assessing AI for algorithmic bias, especially before purchase and deployment, to ensure that decisions made by AI systems are fair and equitable. The goal is to align with anti-racist and anti-discriminatory commitments, ensuring that AI does not perpetuate bias against any specific group.

## **Repairing Historical Harm**

Some guidelines stress the importance of using AI to support the development of work that repairs past damage done to racial and ethnic minority groups, as well as other historically marginalized groups, such as people of all genders, sexual orientations, ages and abilities.

These provisions call for an intentional effort to avoid further harm to historically marginalized communities through the responsible use of AI.

### **Inclusion and Language**

Provisions highlight the need for inclusive and respectful language in AI outputs. AI systems are required to use language that does not exclude or harm specific demographics, ensuring that the technology supports rather than hinders the inclusion of historically marginalized populations. This includes reviewing content to ensure that language patterns are inclusive and do not reinforce stereotypes.

### **Algorithmic Discrimination**

Provisions related to algorithmic discrimination emphasize the need to avoid unjustified differential treatment of individuals based on race, gender, ethnicity, sexual orientation and other protected characteristics. AI systems should be designed and monitored to prevent algorithmic biases from impacting decisions related to housing, employment or access to services.

### **Bias Audits and Monitoring**

Some local laws require regular bias audits of automated decision tools, such as those used in employment, to assess their disparate impact on specific populations. Employers are required to disclose the use of such tools and provide transparency regarding the data collected, the criteria used and any potential impact on protected groups.

### **Racial Equity Toolkit**

Certain policies mandate the use of tools like the Racial Equity Toolkit (RET) to evaluate AI-generated content and prevent bias. This toolkit helps city departments ensure that their AI systems do not perpetuate stereotypes or use historically biased data when analyzing datasets or informing decisions. The RET process aims to document steps to avoid discrimination against protected classes.

### **Public Engagement and Fairness**

Many provisions call for public engagement and collaboration to gather diverse perspectives when developing or deploying AI technologies. This includes considering how different populations, particularly historically marginalized groups, may be impacted by AI systems.

Provisions also encourage transparency in decision-making and ensuring that the technology fosters fairness and inclusivity.

## ***Non-Compliance and Penalties***

Only a limited number of policies include explicit provisions related to penalties for non-compliance. These provisions highlight a range of potential consequences, from disciplinary actions to financial penalties and legal recourse, designed to ensure adherence to AI guidelines. The following paragraphs break down the different types of penalties outlined in the policies.

### **Disciplinary Actions and Access Restrictions**

The provision specifies that non-compliance with AI policies may result in various disciplinary actions. These include restricting access to AI tools and systems to prevent further violations. In more severe cases, such as repeated or significant breaches, the consequences could escalate to termination of employment or the ending of vendor contracts. These measures aim to ensure that individuals and organizations adhere strictly to the established guidelines.

### **Financial Penalties**

In addition to internal disciplinary actions, the policies outline financial penalties for specific violations. For instance, employers who misuse automated tools, particularly in employment decisions, may face fines. Non-compliance can result in fines ranging from \$500 for an initial violation to \$1,500 for subsequent violations. Some policies specify that each day that a violation continues is considered a separate offense, compounding the potential financial consequences.

### **Cybersecurity and Privacy Violations**

Unauthorized or unsupervised use of AI tools, particularly generative AI, may lead to data breaches or violations of citywide cybersecurity policies. Such incidents could be classified as cybersecurity threats, triggering an internal review and response. These violations, if found to contravene privacy laws or cybersecurity standards, also could be reported to the appropriate authorities, especially if they involve illegal activity.

### **Suspension of Network Access**

The policy allows for the suspension or blocking of access to government technology and network resources for individuals or entities found in violation of the policy. These actions may



be taken to safeguard the integrity, confidentiality and availability of government systems. This ensures that unauthorized access or misuse of AI tools does not compromise broader governmental operations.

### **Legal Actions and Financial Recovery**

For more severe violations that cause financial losses or other damages, the policy provides for legal recourse. The government may pursue legal action to recover any financial losses resulting from non-compliance. This highlights the government's commitment to upholding its AI policies and ensuring that violators are held accountable, particularly when their actions result in tangible harm.

### ***Training and Capacity Building***

The provisions related to training and capacity building emphasize the importance of equipping employees with the necessary skills and knowledge for effective AI/Machine Learning (ML)L governance. Agencies are tasked with implementing educational programs tailored to AI/ML initiatives, ensuring that staff involved in these projects possess the technical competencies, ethical awareness and understanding required to responsibly develop and manage AI/ML technologies.

### **Training and Capacity Building for AI/ML Initiatives**

The policies highlight the critical role of training and education in AI/ML governance. It emphasizes that agencies are responsible for providing their employees with educational programs related to AI/ML initiatives. This provision ensures that employees working with AI/ML systems are equipped with the relevant skills and understanding necessary to effectively manage, develop and implement these technologies.

### **Employee Education**

The provision calls for agencies to prioritize the development of educational programs that cater specifically to employees involved in AI/ML projects. This aims to increase their technical competencies, familiarize them with the potential risks and opportunities of AI/ML technologies and enable them to apply best practices in these areas.



## **Skill Development and Awareness**

By offering such training, agencies are expected to build the capacity of their workforce, ensuring that employees are not only proficient in AI/ML applications but also are aware of ethical considerations, potential biases and governance standards. This focus on capacity building is integral to ensuring that AI/ML tools are used responsibly and effectively within the government.

## ***Transparency***

This section outlines the provisions related to transparency in the development, use and deployment of AI systems, focusing on disclosure, compliance and accountability across various policies.

### **Transparency in AI Development and Usage**

Provisions emphasize the importance of ensuring that AI systems comply with all applicable laws and regulations prior to deployment. This includes making publicly available the documentation about the purpose, data collected, location and potential impacts of AI systems. Cities and agencies are directed to establish transparency requirements in solicitation contracts, encouraging the use of open-source code and clear details on how AI systems are trained and the data they utilize.

### **Disclosure of AI Usage**

Provisions require disclosing the use of AI, particularly in public communications or significant reports where AI-generated content is part of the final product. For instance, users must disclose when AI systems like generative AI tools (e.g., ChatGPT, Bard) are used in creating documents, reports or policy memos. Labeling AI-generated content and ensuring attribution for images, text or any other media generated by AI are critical components of maintaining public trust and transparency.

### **Public Engagement and Feedback**

Transparency provisions encourage agencies to disclose AI usage to the public, ensuring accountability. In certain contexts, city employees also are advised to engage with the community to gather feedback about the AI systems used, further promoting openness.

Moreover, transparency fosters trust and documenting how AI models and tools are used, including any prompts, helps build credibility and aids in error correction.

### **Governance and Monitoring**

Provisions require the establishment of governance boards or committees to oversee AI systems, ensuring ongoing monitoring and evaluation. Logging and auditing mechanisms should be in place to track AI use and maintain transparency across AI initiatives. These governance structures are vital for managing risks and ensuring AI compliance with ethical standards and fairness across different user groups.

### **Transparency in Employment Decisions**

Local laws mandate transparency in the use of automated tools for employment decisions. Employers must notify candidates or employees when such tools are used, disclose the types of data collected and provide information on data retention policies. This ensures that individuals are fully informed about how AI systems are affecting their employment opportunities.

### **Transparency in AI Content Creation**

Policies recommend disclosing the version and type of AI model used when generating content, including citations and verifiable sources for facts and figures. This helps ensure the credibility of AI-generated content and aligns with transparency requirements by clearly marking AI involvement in the production process. Watermarking capabilities and labeling AI-generated content, even when edited by humans, also are encouraged to maintain a clear distinction between human and machine-generated outputs.

### ***Unauthorized Use and Prohibitions***

This section outlines the provisions related to unauthorized use and prohibitions in the use of AI technologies, focusing on safeguarding sensitive information, ensuring compliance and restricting certain uses across various policies.

### **Prohibition of Sharing Sensitive Information**

A key provision across many policies is the strict prohibition against entering sensitive or confidential information into AI systems, especially public generative AI platforms such as ChatGPT. This includes personally identifiable information (PII) and other private data. Users

are repeatedly cautioned that such information may be accessible to the companies running these AI systems or even the public.

### **Restrictions on AI Capabilities**

Several policies outline prohibitions on using AI tools to make fully automated decisions without human review, particularly when these decisions impact residents, such as determining eligibility for services or benefits. Other restrictions include banning the use of AI technologies that manipulate or fake someone's appearance or voice without consent.

### **IT Department and Authorization Requirements**

In some cases, city employees are required to obtain approval from IT departments before accessing or using AI products. Unauthorized AI tools, or those not reviewed and approved by IT, are prohibited from use in city operations. Additionally, personal AI accounts are banned for work purposes and city employees must ensure AI tools comply with public records retention policies.

### **Prohibitions on Unauthorized AI Content and Data Use**

Several policies strictly prohibit the use of non-enterprise AI and ML platforms due to data security and privacy risks. *Non-enterprise* refers to AI and ML platforms or tools that are not officially approved, managed or supported by an organization's IT department or infrastructure. These platforms are typically consumer-grade, publicly available or personal-use tools that lack the enterprise-level security, compliance and integration features required for organizational use.

Employees are barred from sharing, uploading or transmitting government data to unauthorized third-party AI platforms. Furthermore, using AI tools to bypass security controls or engage in illegal activities, such as creating content related to weapons or terrorism, is explicitly forbidden.

### **Revocation of AI Capabilities**

Some policies grant IT departments the authority to revoke or restrict the use of AI capabilities if they pose risks that cannot be mitigated. This includes cases where AI technologies cannot prevent city data from being used to train external AI models.

## Appendix A. Methodology Protocol for Completing Steps in Using ChatGPT for Policy Analysis

This protocol was used to analyze AI policies and guidelines at the local level. The goal was to identify common components and provisions, with an emphasis on equity, ethical considerations, transparency, human oversight and data privacy.

The analysis aimed to answer the question: *What are the common components and provisions found in AI policies across different jurisdictions, with a specific emphasis on equity, ethical considerations, transparency, human oversight and data privacy?*

### Step-by-Step Approach

#### Step 1: Upload and Prompt

After downloading the policies, researchers attached one policy file at a time using the attachment icon in ChatGPT. They then added the following prompt, *“Populate the table listed below with a detailed summary in each column, to address the research question: “What are the key components of AI policy and how does it address equity/ethics, bias, human oversight and data privacy? List section and subsection numbers from the document at the end of each finding.”* The columns in the evidence table included:

1. Number
2. Jurisdiction
3. Title
4. Signed
5. Effective Date
6. Document Type
7. Passed By
8. Reviewed By
9. Approved By
10. Enacted By
11. Applies To
12. Collaborators
13. Relevant Policies
14. Policy Summary
15. Bias Mitigation
16. Data Privacy
17. Equity & Ethics
18. Penalties

19. Community Engagement
20. Training/Capacity Building
21. AI Acquisition
22. Evaluation/Quality
23. Transparency
24. Human Oversight
25. Environmental Impact
26. Specific Populations
27. Section Outline
28. Copyright/Ownership
29. Unauthorized Use

### **Step 2: Review Results**

Researchers read the results included in the table and compared them to the policy by reading the provision.

1. If any information was incorrect, a correction was made and it was noted that AI had produced an error.
2. If any information was unclear, a follow-up prompt was used to improve clarity.

### **Step 3: Complete Evidence Tables**

Researchers completed the evidence tables.

- Researchers copied and pasted information from ChatGPT into Excel.
- Researchers attached the next policy and repeated the process.
- By using section and subsection numbers referenced by ChatGPT in the summaries, researchers checked the documents to ensure that the information was accurate.

### **Step 4: Summarize Information Using ChatGPT**

Researchers summarized the information using ChatGPT.

- Researchers copied the information from each relevant column, one item at a time.
- Researchers pasted all the copied information into ChatGPT.
- Researchers used the sample prompt found in *Figure A-1* to request a summary based on the provided table information.

**Figure A-1. Sample Prompts for Summary Creation and Quality Assurance Strategies**

Category	ChatGPT Prompt	Quality Assurance
A-1. Jurisdiction Level	"Summarize the jurisdiction levels across all policies."	Ensure the summary covers all relevant jurisdiction levels.
B-1. Policy Title	"Summarize the titles of policies in this category and list key themes."	Verify that all policy titles are listed and accurately summarized.
C-1. Responsible Agency	"Summarize the responsible agencies across all policies."	Confirm that all responsible agencies are correctly identified.
D-1. Scope of Policy	"Summarize the provisions across all polices and list the results by theme."	Check that the scope of each policy is thoroughly summarized.
E-1. Equity Provisions	"Summarize equity provisions across all policies and list the results by theme."	Ensure all equity provisions are included and summarized accurately.
F-1. Bias Mitigation	"Summarize the bias mitigation strategies across all policies and list them by theme."	Verify the completeness and accuracy of the bias mitigation summary.
G-1. Human Oversight	"Summarize the human oversight mechanisms across all policies."	Confirm that all human oversight mechanisms are included and summarized.

## **Lessons Learned from Using AI Tools**

The information provided below is not intended to endorse any specific AI tool but is primarily meant to provide a high-level description of the lessons learned related to their usage in the conducted literature review.

The team used two tools for this policy review. In general, both tools — ChatGPT and Petal — demonstrated their utility in the policy review process. The team utilized both free and paid subscription versions of ChatGPT. Advantages included the identification of articles and the development of summaries organized by themes based on prompts with reasonable accuracy. However, in some cases, the summaries were too general and required additional editing through further prompting. An additional issue occasionally observed was that some summaries included information not sourced from the policy provided. To mitigate this, the research team implemented measures, including disabling ChatGPT's memory and internet browsing capabilities in the settings, to ensure outputs were based on the policy content.

The second AI tool used was Petal. Given that the functionality of the Petal tool differs from ChatGPT — it does not browse the web and works solely with the documents provided — the research team did not need to implement any extra measures beyond standard review and quality assurance to ensure that the summaries accurately reflected the content of the provided policies. However, some limitations noted included challenges in recognizing specialized terminology and phrases used in the policies which sometimes resulted in outputs that did not fully capture the intended meaning. The team addressed this issue by conducting manual reviews to correct AI-generated outputs, ensuring they aligned with the intended meaning and adding additional prompts to guide the tool toward producing more accurate and relevant results.

Further testing of these and other AI tools is essential to identify the most effective ways to leverage their capabilities for policy reviews in the future, while also addressing potential ethical considerations.

## Appendix B: Bibliography

1. Austin, Texas. City Council. (2024, February 15). Resolution on AI. Retrieved from <https://services.austintexas.gov/edims/document.cfm?id=423417>
2. Baltimore, Maryland. Mayor of Baltimore. (2024, March 20). Mayoral Executive Order Establishing Principles and Policy Governing Use of Generative Artificial Intelligence. Retrieved from <https://mayor.baltimorecity.gov/news/press-releases/2024-03-20-mayor-scott-issues-executive-order-use-generative-artificial>
3. Boise, Idaho. City of Boise. (2023, December 1). City Use of Artificial Intelligence (AI) Regulation. Retrieved from <https://www.cityofboise.org/departments/human-resources/employee-policy-handbook/section-400-general-provisions/430q-city-use-of-artificial-intelligence-ai-regulation/>
4. Boston, Massachusetts. Department of Innovation, City of Boston. (2023, May 18). Interim Guidelines for Using Generative AI. Retrieved from <https://www.boston.gov/sites/default/files/file/2023/05/Guidelines-for-Using-Generative-AI-2023.pdf>
5. Clearlake, California. City Management. (n.d.). Generative AI Usage Policy. Retrieved from <https://mccmeetingspublic.blob.core.usgovcloudapi.net/clearlakca-meet-2525f632061c4c79bcfdeede22a76bf/ITEM-Attachment-001-36be984a1a6f451a88b17ff59bbdae5c.pdf>
6. Lebanon, New Hampshire. City of Lebanon. (2023, December 19). Use of Artificial Intelligence. Administrative Policies & Procedures. Retrieved from <https://digitalgovernmenthub.org/examples/city-of-lebanon-use-of-artificial-intelligence-policy/#:~:text=On%20December%2019%2C%202023%2C%20the,%2C%20fair%2C%20and%20secure%20application>
7. Long Beach, California. City of Long Beach. (2024, August 13). City of Long Beach Generative AI Guidance v1.1. Retrieved from <https://longbeach.gov/globalassets/smart-city/media-library/documents/generative-ai-guidance-v1-1>



8. New York City, New York. City Council. (2021, November 1). Automated Employment Decision Tools. Retrieved from <https://legistar.council.nyc.gov/LegislationDetail.aspx?ID=4344524&GUID=B051915D-A9AC-451E-81F8-6596032FA3F9&Options=ID%7CText%7C&Search=>
9. New York City, New York. Office of Technology and Innovation (OTI). (2024, March 4). Preliminary Use Guidance: Generative Artificial Intelligence. Retrieved from <https://www.nyc.gov/assets/oti/downloads/pdf/about/preliminary-use-guidance-general-artificial-intelligence.pdf>
10. San Francisco, California. City Administrator's Office. (2023, November 9). San Francisco Generative AI Guidelines. Retrieved from <https://www.sf.gov/sites/default/files/2023-11/Generative%20AI%20Guidelines%20-%20CCSF.pdf>
11. San Jose, California. City of San Jose. (2023, September 23). Generative AI Guidelines. Retrieved from <https://www.sanjoseca.gov/home/showpublisheddocument/100095/638314083307070000>
12. Seattle, Washington. City of Seattle. (2023, October 23). Generative Artificial Intelligence Policy (POL-209). Retrieved from <https://mrsc.org/getmedia/771150fb-9ec5-47d2-aa63-596318da1890/s42aipol.pdf>
13. Tempe, Arizona. City of Tempe. (2023, May 31). Ethical Artificial Intelligence (AI) Policy. Retrieved from <https://tempe.hylandcloud.com/AgendaOnline/Documents/ViewDocument/ETHICAL%20ARTIFICIAL%20INTELLIGENCE%20POLICY.DOCX.pdf?meetingId=1451&documentType=Agenda&itemId=5692&publishId=9354&isSection=false>
14. Washington, D.C. Chief Technology Officer (CTO). (2024). AI/ML Governance Policy. Retrieved from <https://octo.dc.gov/page/aiml-governance-policy>