*Number:* 03.020.00
*Last Review Date:* 07/2024

# City of Riverside
# **Administrative Manual**

| | |
|---|---|
| *Effective Date:* | 07/2024 |
| *Latest Revision Date:* | 07/2024 |
| *Next Review Date:* | 07/2025 |
| *Policy Owner(s):* | Innovation & Technology Department |

*Approved:*

George Khalil (Jul 2, 2024 13:12 PDT)

Department

Mike Futrell

Mike Futrell (Jul 5, 2024 10:59 PDT)

City Manager

## SUBJECT:

## Artificial Intelligence (AI) Policy

## PURPOSE:

The City of Riverside is committed to harnessing the power of Artificial Intelligence (AI) responsibly, ethically, transparently, and securely. This policy outlines the requirements for City departments to leverage AI effectively for improved efficiency, better decision-making, and enhanced service delivery for residents. The City of Riverside focuses on developing and using AI in a human-centered, ethical, responsible, accurate, inclusive, and transparent way across all departments. This approach will accelerate service delivery, improve city services for residents, and empower our employees.

## AI Background and Definitions:

### 1. Artificial Intelligence (AI):

For the purposes of this policy, Artificial Intelligence (AI), also known as machine intelligence, is the simulation of human intelligence processes, such as problem-solving by machines. AI tools are computer programs capable of many activities, including but not limited to writing assistance, content generation, data analysis, pattern detection, predictive analytics, translation, cybersecurity, programming code generation, image, audio, or video creation.

### 2. Artificial Narrow Intelligence (ANI):

- This is the most common type of AI, focused on performing specific tasks in well-defined domains. Examples include chess engines, language translation, cybersecurity, and image recognition software.
- ANIs excel in tasks requiring high-volume pattern recognition and data analysis but lack general human intelligence, context awareness, and adaptability.

### 3. Generative AI:

Generative AI can be considered a specialized form of ANI, focusing on creating original content rather than simply completing a task. It is often provided as a web-based or integrated software solution that can generate entirely new content. This content creation goes beyond simple manipulation or modification of content.

Generative AI leverages complex algorithms and machine learning models to understand existing data's underlying patterns and structures. It then utilizes this understanding to produce original content in a wide range of formats like:

- **Text**: From creative writing and poetry to code generation and report writing, Generative AI can create various forms of text content with remarkable fluency and coherence.
- **Images and Video**: Generative AI can generate realistic and visually stunning images, animations, and videos, often indistinguishable from human-created work. This allows for applications in design, entertainment, and virtual reality.
- **Audio and Music**: Generative AI can compose original music, generate sound effects, and even mimic human voices with uncanny accuracy, opening doors for new musical and audio-based experiences.
- **Other Forms of Data**: Generative AI can even translate input data into entirely new forms, such as converting **written** instructions into computer code. This opens possibilities for automating tasks and streamlining workflows.

### 4. Generative Adversarial Networks (GANs):

GANs are a specific type of Generative AI that trains two neural networks (or machine learning processes) to compete against each other to learn and generate new content. These neural networks include a generator and a discriminator:

- Generator: This network takes noise or random input and attempts to transform it into realistic and convincing outputs such as images, video, text, or music.
- Discriminator: This network acts as a critic, trying to distinguish between known content and the generator's creations. It provides feedback to the generator, helping it improve its output over time.

This adversarial process creates a feedback loop where the generator constantly strives to "fool" the discriminator while the discriminator continuously refines its ability to detect fakes. This dynamic training process allows both networks to become highly skilled. For example:

- Generator: It learns to generate increasingly realistic and diverse outputs, capturing the essential characteristics of the input data.
- Discriminator: Develops a deep understanding of accurate data, enabling it to identify and reject forgeries accurately.

However, GANs also share some challenges with other Generative AI models, such as potential biases in their training data that require intentional and careful human oversight to avoid unintended consequences like generating harmful or offensive content.

### 5. AI Training Data:

AI training data is the fuel that powers machine-learning models. It's the collection of information, examples, and patterns that an AI model learns from to perform its tasks. Depending on the model's purpose, this data can be anything from text, images, and videos to audio and sensor readings. The diversity and accuracy of training data are foundational for an ethical, inclusive, and transparent AI implementation.

### 6. Bias risk:

Bias risk refers to the potential for AI models to learn and perpetuate unfair, discriminatory, or biased patterns or missing perspectives or minority groups in their training data. This can lead to inaccurate or unfair outcomes when the AI is deployed in real-world applications.

**7. Prompts:**

Any form of text, question, information, or coding that communicates to AI what response you're looking for.

**8. AI Training Data:**

Training data could come from images, audio, video or text in the form of spreadsheets, PDFs or HTML code – to name just a few. AI uses the data to improve its recognition of patterns, to make predictions and to improve its learning capabilities.

## SCOPE:

This policy applies to the selection, design, development, and deployment of AI for:

- All departments, employees, contractors, and stakeholders involved in developing, deploying, and utilizing AI within or on behalf of the City of Riverside.
- All cases where AI functionality is known to be included, such as enhancements for existing products, new products being considered for use, or AI technology developed by City of Riverside employees, contractors, partner agencies, or other stakeholders for City use.

## POLICY:

The following policy outlines the guiding principles and details the processes necessary for the successful use of AI within the City of Riverside. Thoughtful implementation and adherence to these approved processes will allow the successful automation of routine tasks where AI will streamline service delivery**,** reduce wait times, and reduce cost and response delays. This will allow our public servants to dedicate their time and skills to solving complex problems and providing world-class and personalized assistance to City residents, businesses, and visitors.

**AI Adoption Guiding Principles:**

- **Prioritizing People:** Human well-being, safety, equity, and dignity are at the heart of our AI strategy**.** We design technologies that amplify capabilities, embrace inclusion, and contribute to a better future for everyone.
- **Human & AI Amplifying Each Other:** We harness the synergy of humans and AI to make better decisions while **maintaining** ultimate control in human hands.
- **Building Fair AI:** We actively eliminate bias and build inclusive, ethical AI technology. Every step, from design to deployment, ensures fairness, transparency, and equity, protects privacy, and mitigates harm**.** We continuously learn and improve to deliver just outcomes powered by AI.
- **Transparently Powered AI:** We build, procure, or leverage transparent and explainable AI**.** Users will know when they interact with it and why it makes decisions. Clear documentation and accessible information keep everyone informed and empowered.
- **AI Accountability:** We take ownership of the AI technology we use - establishing rigorous standards for its selection, design, development, and deployment. We proactively monitor and

evaluate its performance, implementing safeguards and adjusting where needed to address unintended consequences. By upholding ethical standards, we strive for responsible AI that benefits everyone.

- **Privacy and Security:** We establish rigorous standards for development, deployment, and oversight, ensuring accountability at every stage. We handle personal data with utmost security and privacy, adhering to strict minimization and purpose-limitation principles, while robust safeguards protect information from unauthorized access, misuse, or breaches. Continuous monitoring, evaluation, and mitigation strategies address potential unintended harms of AI while proactively engaging with stakeholders and transparency in algorithms and data usage.
- **Embracing a continual journey towards ethical AI:** We actively learn from experience, research trends, and refine best practices. By staying attuned to evolving ethical perspectives, we refine our AI technology, ensuring its positive impact on society.
- **Equipping our workforce:** We offer dedicated training programs fostering AI literacy, ethical awareness, privacy protection, and responsible practices. This investment empowers employees to shape AI for good.
- **Legal and regulatory obligations:** Navigating the evolving legal landscape of AI, we adhere to all applicable laws and regulations while actively contributing to developing ethical frameworks that promote responsible AI for a brighter future.

**AI Governance:**

- **City of Riverside administrative policies**
    - **Technology Selection and Acquisition Policy 03.017.00**
        - Consistent with the City's standards for acquiring technology, City employees and departments may be authorized to use pre-approved AI software tools listed at the following URL.
            - https://riversideca.sharepoint.com/sites/thehive-it-ops/SitePages/Pre%20Approved%20Software.aspx
        - City employees and departments seeking non-standard AI tools not included on the pre-approved list must submit exception requests to the Department of Innovation and Technology (IT) for consideration through the Needs Assessment form on TechHub.
        - The policy applies to free, on-premises, cloud-hosted, or Software-as-a-Service (SaaS) software.
        - The City of Riverside reserves the right to revoke or restrict the use of any technology, including added AI capabilities if such capabilities pose unacceptable risks in the City's judgment.
    - **Technology Use and Security Policy (TUSP) 03.002.00**
        - Employees of the City of Riverside are prohibited from submitting classified, regulated, personal, or Confidential data to AI systems unless IT has implemented appropriate controls and data protection measures. This also applies to any data deemed unacceptable for public disclosure.
        - The City of Riverside prohibits using City data or records, including inputs or prompts, for training or optimization of Generative AI models that operate outside the City's direct control. As such, City employees are prohibited from using AI technologies that lack mechanisms to prevent City data or records from contributing to their language models.
        - To ensure proper record retention and data security, City employees using

Generative AI must register their usage through a dedicated City-issued email address assigned by IT who then integrates the usage into a centralized system for record retention and public record request laws.

- City data security and legal obligations apply to all AI tools to safeguard privacy and control; the City's internal or confidential records and prompts are strictly for internal Generative AI training and tuning. AI technology lacking robust safeguards against City data leaks is strictly prohibited. AI tools shall not prevent, hinder or circumvent the City's legal obligations under the California Public Records Retention Act and e-discovery obligations.
- City-approved AI solutions must offer seamless retrieval and export of prompts and outputs. Corresponding records of inputs, prompts, and outputs are subject to public records requests and e-discovery in accordance with established public disclosure policies and practices.

- **Safeguarding Against Misinformation: Policy on Generative AI Outputs**

  - To ensure responsible deployment of Generative AI, all City-generated outputs must undergo documented human quality assurance review, adhering to this document's guiding principles. Departments are required to establish and document these reviews which must demonstrate information accuracy, transparency, and compliance for all AI-generated items.

    AI-generated content by City staff falls under California's public records law and is subject to retention schedules, balancing accountability, legal obligations, public trust, and fostering Innovation.

  - All products created by both humans and AI must give credit to the specific AI system and its human partner to further enhance public trust and transparency.

    - Sample watermark: ***"Insights in this report, graphics, video, audio, software development etc. were crafted by Microsoft Co-Pilot, Google Gemini, Adobe Firefly, OpenAi ChatGPT 4, Anthropic Claude 3, etc., and meticulously fact-checked through human review by the City Planning team, ensuring accuracy and alignment with City goals."*** By naming the AI and its human counterpart, we transparently celebrate the synergy that powers responsible innovation. Clarity strengthens public trust by ensuring the AI fingerprints and human handprints that guide them are always visible.
    - Fact-based statements should be attributed to a credible source such as official City or government documents, published and peer-reviewed research, press releases, etc., not simply accepted as truth from the AI or citing non-subject matter expert opinions.
    - AI-generated video shall embed citations for its source into each frame of every image and video.

  - To guard against bias and discrimination, City employees purchasing, developing, or training a new Generative AI system shall prioritize the diversity of training data to ensure accurate, equitable and representative AI outcomes before purchasing, developing or training AI applications.
  - AI implementations and the associated training data shall be reviewed and approved by a Citywide AI steering committee led by the IT, the Human Resources Department, and subject matter experts from all City departments. The committee will conduct an analysis

of all AI training data and meticulously document steps taken to evaluate AI-generated content for accuracy and freedom from potential bias that may be underrepresented or missing altogether from AI data. Responsible AI starts with proactive bias-busting during AI training and implementation. The City of Riverside is committed to a fair and equitable future.

- The below chart provides guidance on when AI usage can be used without citation, with citation, or when AI use is not acceptable to guide responsible and transparent AI usage and adoption:

| Breadth of Distribution | Proofreading, Grammar | Brainstorming | First Draft | Collaborative Writing with human fact checks | Human Edited without fact checks | Copy-Paste Generated Text |
|---|---|---|---|---|---|---|
| Press release, prepared remarks | Use | Cite Use | Cite Use | Not Acceptable | Not Acceptable | Not Acceptable |
| Replies to Public Inquiry | Use | Cite Use | Cite Use | Not Acceptable | Not Acceptable | Not Acceptable |
| Public-facing content | Use | Use | Cite Use | Cite Use | Not Acceptable | Not Acceptable |
| Memos, broad internal communication | Use | Use | Cite Use | Cite Use | Not Acceptable | Not Acceptable |
| Internal process documentation | Use | Use | Use | Cite Use | Not Acceptable | Not Acceptable |
| Emails | Use | Use | Use | Cite Use | Cite Use | Not Acceptable |
| Chat | Use | Use | Use | Cite Use | Cite Use | Not Acceptable |

Credit: Microsoft Corporation 2023 AI maturity guidelines

### Prohibited Uses

Generative AI can create realistic fake content, such as fake news or deepfake videos, images, or voices, which can be used to spread misinformation and/or manipulate public opinion. AI tools shall not be used for illegal, harmful, or malicious activities. This includes activities perpetuating unlawful bias, automating unlawful discrimination, and producing other harmful outcomes.

### Responsible AI Usage

Employees are not currently required to use AI systems and tools. Members of the public should be informed when interacting with an AI tool and be provided with an available alternative to using AI tools. Employees are responsible for ensuring that the AI-generated content aligns with the City's values, ethics, and quality standards. Generated content shall not be used if it is misleading, harmful, offensive, or discriminatory.

### Access and Security

<u>Authorized Access</u>

Access to AI tools, platforms, or related systems should be restricted to authorized personnel. Users may request access through the City's helpdesk ticketing system MAC form and shall not share their access credentials or allow unauthorized individuals to use AI tools on their behalf.

Secure Configuration

AI tools and platforms must be configured securely, following industry best practices and vendor recommendations. This includes ensuring the latest updates, patches, and security fixes are applied on time and subject to security, technology, and bias audits. IT is responsible for implementing, maintaining, and securing the City's technology infrastructure, including AI systems.

User Authentication

Strong authentication mechanisms, such as SAML multi-factor authentication (MFA), shall be implemented to access generative AI tools and platforms. Passwords used for access should be unique and complex and changed regularly in adherence to the City's Technology Use and Security Policy (TUSP).

Data Protection

Users shall handle any personal, sensitive, or confidential data generated or used by AI tools in accordance with the City's data protection policies and applicable laws. Encryption and secure transmission should be employed whenever necessary. Inputting sensitive or confidential organization data into a public online AI prompt exposing City data to third parties is prohibited.

**Monitoring and Incident Response**

Logging and Auditing

Appropriate logging and auditing mechanisms shall be implemented to capture activities related to AI usage, including but not limited to prompts, output, training data used, and human quality assurance actions. These logs shall be regularly reviewed to detect and respond to any suspicious or unauthorized activities.

Incident Reporting

Any suspected or confirmed security incidents related to AI usage shall be reported promptly to IT.

Vulnerability Management

IT shall conduct regular vulnerability assessments and security testing on AI tools and platforms to identify and address security weaknesses.

**Copyright and Intellectual Property:**

Users are responsible for ensuring any content they create using Generative AI systems does not infringe copyright. Tools like plagiarism checkers and reverse image searches can aid in verifying text and image ownership; they are not foolproof. When in doubt, edit the content to be original or avoid using it altogether. Generally, the City owns the input and output from these services, but many free or individually licensed consumer-focused AI companies retain usage rights for both. City data entered into a consumer-focused AI tool could be used by the company to train their models or even in their marketing activities, only input information you are authorized to make public in consumer-focused AI or use city-provided and authorized enterprise AI through IT when available. Copyright and ownership considerations demand carefulness when using Generative AI.

**Training and Awareness**

When City-provided AI training becomes available, employees shall participate to ensure appropriate and secure use of AI, data handling, and ongoing adherence to City policies. This training should cover ethical considerations, potential risks, security best practices, and compliance requirements.

Employees using AI tools are encouraged to educate themselves on effective and appropriate AI usage. Regular awareness campaigns and communications should reinforce the importance of cybersecurity, responsible AI usage, and adherence to this policy.

**Policy Review**

This policy will be reviewed periodically and updated to address emerging risks, technological advancements, and regulatory changes.

**Sources and References**

Some insights in this policy were crafted by Google Gemini and meticulously fact-checked through human review by IT, ensuring accuracy and alignment with City goals.

https://seattle.gov/documents/Departments/SeattleIT/City-of-Seattle-Generative-Artificial-Intelligence-Policy.pdf

https://www.sanjoseca.gov/home/showpublisheddocument/100095/638314083307070000

https://tempe.hylandcloud.com/AgendaOnline/Documents/ViewDocument/ETHICAL%20ARTIFICAL%20INTELLIGENCE%20POLICY.DOCX.pdf?meetingId=1451&documentType=Agenda&itemId=5692&publishId=9354&isSection=false