State of California
Department of Technology
Office of Information Security

Generative Artificial Intelligence Risk Assessment

SIMM 5305-F

August 2025

REVISION HISTORY

Revision	Date of Release	Owner	Summary of Changes
Initial Release	January 2024	California Office of Information Security - Payam Hojjat	Initial Release of SIMM 5305-F Generative Artificial Intelligence Risk Assessment
Minor Update	July 2024	California Office of Information Security - Payam Hojjat	Added clarifying questions for GenAI regarding: • Review and consultation. • State entity policies. • User training.
Major Update	February 2025	California Office of Information Security - Kory Fesliyan	• Updated Format • Restructured and revised questions • Added a new section to document safeguards • Added a new section for data types • Updated risk assessment table
Minor Update	August 2025	California Office of Information Security - Kory Fesliyan	• Added Labor Relations Engagement

Table of Contents

I. Introduction

Generative Artificial Intelligence (GenAI) has the potential to improve the delivery of government services and operations. GenAI enables enhancements to the development, adoption, and implementation of new technologies to streamline and optimize business operations and state services to Californians. With that, it is critical for entities to be cognizant to ensure that GenAI does not lead to a state in which human life, health, property, or the environment is endangered, nor have public services be solely contingent upon these systems. GenAI systems are to be used only to augment and improve workflows, not to replace or impair the services received by the public.

As described in the State of California Report: Benefits and Risks of GenAI report, GenAI offers a wide variety of potential applications with varying impacts. Any application of GenAI tools within the California state government will adhere to appropriate protocols and testing procedures. To proactively address potential threats to state-owned information assets, privacy, and the welfare of California's citizens, the Statewide Information Management Manual (SIMM) 5305-F, GenAI Risk Assessment introduces a risk assessment methodology that will aid state entities in evaluating the risks associated with GenAI systems.

This SIMM is to ensure alignment with:

- Executive Order (EO) N-12-23
- NIST Artificial Intelligence Risk Management Framework

Please note that a completed SIMM 5310-C Privacy Threshold Assessment and Privacy Impact Assessment must be accessible upon request.

II. Risk Assessment - Part 1

Generative Artificial Intelligence Risk Assessment Part 1:

This section completed by the Chief Information Officer (CIO)

Instructions:

- This risk assessment is required for ***all*** GenAI procurements, acquisitions, renewals and internally developed systems.
- This risk assessment applies to **free** GenAI products and services, defined as any free software or service that interacts with state data. Examples include users entering state data into conversational GenAI platforms or installing GenAI plugins and extensions. However, this assessment does not apply in cases where state data is not being used, such as browsing the web with search engine GenAI results or using conversational GenAI systems where no state data is entered.
- This risk assessment is required for ***existing*** GenAI tools that were acquired prior to the release of this document. These assessments must be retained and accessible by the entity as it may be requested during an information security program audit or assessment.
- State entities complete SIMM 5305-F, Risk Assessment Part 1 to determine the level of risk associated with a GenAI system.
- After completion, submit a Case via the New Technology Consultation and Assessment request, in the CDT IT Service Portal for all risk levels. When the request has been processed, a CDT Customer Engagement Services (CES) representative will reach out to provide a secure location to upload the required documents.
- *CDT reserves the right to audit and consult on "**Low**" GenAI Risk Levels with potential higher risk concerns.*
- Only complete SIMM 5305-F, Risk Assessment Part 2 if the GenAI system risk level is rated **Moderate** or **High**.
- **Important**: Once completed, this form is confidential and may be exempt from disclosure pursuant to Government Code sections 7929.210 and 8592.45.

Data Types

Select all that apply to identify the data types at risk. Take into account the potential impact to data types in this risk assessment questionnaire (**note**: data types may overlap across categories).

☐ **Personal Information (Privacy) or PII** - Information that can be used to distinguish or trace an individual's identity, including those in accordance with Civil Code 1798.3 & 1798.29.

1. ☐ Individual's first name or initial and last name
2. ☐ Date of Birth
3. ☐ Social Security Number
4. ☐ Physical description
5. ☐ Home address
6. ☐ Telephone number
7. ☐ Education
8. ☐ Financial matters, including financial/bank account number, credit or debit card number, or any code that would permit access to an individual's financial account
9. ☐ Medical history and medical information
10. ☐ Health data – other identifiable medical data not covered by Health Insurance Portability and Accountability Act (HIPAA). (e.g., occupational health, services qualification, personal health record)
11. ☐ Employment history
12. ☐ Statements made by & attributed to an individual
13. ☐ Driver's license number
14. ☐ California identification card number
15. ☐ Tax identification number
16. ☐ Passport number
17. ☐ Military identification number
18. ☐ Other unique identification number issued on a government document commonly used to verify identity of a specific individual
19. ☐ Security code, access code, password, and username/email address
20. ☐ Health insurance information
21. ☐ Unique biometric data, such as fingerprints, retina, or iris scans, used for authentication
22. ☐ Physical or digital photograph, if used or stored for facial recognition purposes
23. ☐ Information collected through automated license plate recognition systems (GC Section 1798.90.5).
24. ☐ Genetic data
25. ☐ Personal data as defined by European Privacy Law (EEA and UK GDPR)
26. ☐ Special data as defined by EEA and UK GDPR
27. ☐ Other:

☐ **Confidential** - Information maintained by state agencies that is exempt from disclosure under the provisions of the California Public Records Act (Government Code Sections 7923.75-7923.55, 7929.210, 7930.000-7930.005, 7930.100-7930.215) or has restrictions on disclosure in accordance with other applicable state or federal laws.

1. ☐ Intelligence information in support of Homeland Security and National Defense
2. ☐ Attorney/Client Privileged and/or Work Product
3. ☐ Controlled Technical Information (CTI) (e.g., Network security info, diagrams, security & transaction logs, encryption keys)
4. ☐ Intellectual property, such as patents, copyright, trade secrets
5. ☐ Proprietary
6. ☐ Electronic Health Record (EHR) information subject to Health Insurance Portability and Accountability Act (HIPAA) Privacy and Security Rules
7. ☐ Protected Health Information (PHI) subject to HIPAA Privacy or Security Rules
8. ☐ Medical information as defined by the California Confidentiality of Medical Information Act (CMIA)
9. ☐ Family Educational Rights and Privacy Act (FERPA) data
10. ☐ Federal Tax Information (FTI) subject to Internal Revenue Service (IRS) Publication 1075 requirements
11. ☐ Controlled Unclassified Information (CUI)
12. ☐ Criminal Justice Information Services (CJIS) information
13. ☐ Department of Defense Covered Defense Information (CDI)
14. ☐ Export Controlled Research information subject to International Traffic in Arms Regulations (ITAR) and Export Administration Regulations (EAR)
15. ☐ Sensitive data that requires a higher-than-normal assurance of accuracy and completeness
16. ☐ Other:

☐ **Public Information -** Information maintained by state agencies that is not exempt from disclosure under the provisions of the California Public Records Act, Government Code Sections 7929.210, or other applicable state or federal laws.

1. ☐ Non-Human Research Data (Animals, etc.)
2. ☐ Other Program-Related Research Data
3. ☐ Aggregated Data (Census, traffic, crime, etc.)
4. ☐ Federal Acquisition Regulations information other than CUI
5. ☐ Sensitive data that requires a higher-than-normal assurance of accuracy and completeness
6. ☐ Other:

III. GenAI Risk Table Assessment Scale

This table categorizes a system's risk level by evaluating the GenAI system's nature (use case), data type, and potential impact (FIPS 199) of a security breach, with the highest applicable category determining the overall risk level. For example, FIPS 199 Low + Mission Related (Moderate) + Processes confidential data (High) would result in a risk level of high based on the use case examples.

FIPS 199 Impact	Data Type	Use Case Examples
HIGH (Red) Depending on the nature & sensitivity of the affected systems or information, losses related to confidentiality, integrity, & availability may still result in ***severe or catastrophic*** adverse impacts	**Confidential/Personally Identifiable Information** • The data inputs, activities, and outputs involve processing Personally Identifiable Information (PII) or confidential data. **Safety Related** • The data inputs, activities, and outputs support decision-making processes that impact public safety.	• Inputs and processing of confidential data, including data that identifies or describes an individual, authentication, biometric identification, financial, medical, procurement, investigative, national security, network architecture schematics, ports, protocols, and others. • Inputs and processing of data that inform safety decisions, such as water treatment ratios, load-bearing specifications for bridges, and similar critical information.
MODERATE (Yellow) Depending on the nature & sensitivity of the affected systems or information, losses related to confidentiality, integrity, & availability may still result in ***serious*** adverse impacts	**Decision Related, Non-Confidential/Non-PII Related, Resident (Public) Related** • The data inputs, activities, and outputs are non-PII and non-confidential but lead to decision making in public (resident) programs and services. **Decision Related, Non-Confidential/Non-PII Related, Not Validated** • The data inputs, activities, and outputs are non-PII, non-confidential, and support decision-making, but the GenAI model/output lacks verification by a qualified subject matter expert against the original data source used by the GenAI. **Mission Related, Resident (Public) Facing Web Apps** • The data inputs, activities, and outputs involve mission-critical, state-critical, critical infrastructure, or public-facing systems.	• Code Analysis & Development Tools (e.g., static and dynamic code analysis, code generation tools, code assistant tools, platforms used to create GenAI solutions). • Chatbots (e.g., internal chatbots for resource finding and process advice, resident-facing chatbots for public interaction and information retrieval). • Public & Direct Contact Services (e.g., customer service, public relations, jurisprudence, output providing recommendations such as legal, tax, regulatory information). • Critical Infrastructure & Safety (e.g., handling mission-critical apps/systems, service eligibility assessments for housing or income assistance, drafting organizational documents, generating data for public use like soil composition or seismic considerations).
LOW (Green) Depending on the nature & sensitivity of the affected systems or information, losses related to confidentiality, integrity, & availability may still result in ***limited*** adverse impacts	**Non-Decision Related, Non-Confidential/Non-PII Related** • The data inputs, activities, & outputs are non-PII, non-confidential, & neither involve nor lead to decision-making throughout the end-to-end business process **Decision Related, Non-Confidential/Non-PII Related, Validated** • The data inputs, activities, and outputs are non-PII, non-confidential, and support decision-making, with the original data source of the GenAI output verified by a qualified subject matter expert.	• Network & Security Tools (e.g., packet inspection, system monitoring, intrusion prevention/detection, spam filtering tools, malware detection tools like anti-virus, anti-malware, & endpoint detection & response). • Content & Communication Tools (e.g., grammar correction tools, summarization tools, reference generating tools like laws, case law, policies, translation tools, text to speech, & content creation tools for image, video, ads, and marketing).

Risk Assessment Questionnaire

GenAI Description and Use Case:
(a) What is the vendor's name that offers the GenAI?
(b) What is the application and/or product name?
(c) What is the model and version of the product?
(d) What is the license tier of the GenAI product, if applicable (free, enterprise, platinum, etc.)?
(e) How is the GenAI solution delivered: IaaS, PaaS, SaaS, or will it be deployed on-premises? *(Indicate if this is a thin client, thick client, web extension, plugin, etc.)*

(f) What is the current end-to-end business process? *(Provide a brief overview of the workflow that is currently being performed).*
(g) Which aspects of the end-to-end business process will the GenAI support AND how will you use GenAI? *(Provide a brief overview of what the GenAI solution will replace, enhance, or introduce as a new functionality and/or services. If only certain features will be enabled, explain their purpose and use in detail.)*

(h) What safeguards will be deployed with this GenAI product?

Work with your security teams to identify safeguards that apply. Review section V - GenAI Sample Safeguards to help complete this section if applicable. Please note that CDT Safeguards are inherited when CDT managed services are used.

- ***For example:*** *Safeguard: Will ensure all GenAI input and output data is validated by an adequately qualified subject matter expert.*
- *State agencies/entities must consider their own tailored mitigation processes and procedures specific to their GenAI product.*

(i) What GenAI features of this product will be disabled, if any (e.g., incidental GenAI features that come with renewals but will be disabled)?	(j) Will there be tailored training and specific rules of engagement for stakeholders to ensure proper usage of the system, with adherence required for each use case?
	☐ Yes ☐ No

FIPS 199 Categorization Level: to be completed by ***ISO*** or equivalent *All GenAI data must be taken into context, and an associated categorization must be given based on the highest severity. This includes prompt data, output data, data source, training data, etc.*	
(a) Is the GenAI tool being used for mission-critical applications or processes? Briefly explain.	
(b) **Confidentiality**: What is the potential impact to the organization if this system results in unauthorized disclosure of information.	
Level (choose only one): ☐ High ☐ Moderate ☐ Low	Impact Description:
(c) **Integrity**: What is the potential impact of unauthorized modification or destruction of information.	
Level (choose only one): ☐ High ☐ Moderate ☐ Low	Impact Description:
(d) **Availability**: What is the potential impact of disruption to the availability of the system or its services.	
Level (choose only one): ☐ High ☐ Moderate ☐ Low	Impact Description:
(e) Based on the responses to the FIPS 199 Categorizations, what is the overall level of the protection for this system/application/service?	
FIPS 199 - Protection Level Needed (choose only one): ☐ High ☐ Moderate ☐ Low	Comments or additional notes (if any):

GenAI Risk Level: According to section III. GenAI Risk Table Assessment Scale, what is the highest risk level based on the FIPS 199 System Categorization, Data Type, and Use Case? *(CDT reserves the right to audit and consult on "Low" GenAI Risk Levels with potential higher risk concerns)*

☐ Low ☐ Moderate ☐ High

Safeguard Level: Based on the safeguards in (h), what level of safeguards have been identified for implementation for this GenAI system.

☐ **Not Identified** - Safeguards have not yet been reviewed or considered for this system or use case
☐ **Partially Identified** - Some safeguards have been recognized, but further evaluation is needed
☐ **Mostly Identified** - Most safeguards have been identified, with minor gaps remaining
☐ **Fully Identified** - All safeguards have been identified and documented for implementation
☐ **Not Applicable** - Safeguards are deemed unnecessary or irrelevant for this system or use case

Signatures and Acknowledgement (Required Completion)

Required Signatures for Risk Assessment Part 1

You are required to ensure that the Acceptable Use Policy is updated to address GenAI. A completed SIMM 5310-C Privacy Threshold Assessment and Privacy Impact Assessment must be accessible upon request. If additional GenAI features are enabled in the future beyond those outlined in this document, a new SIMM 5305-F must be submitted for review.

By signing this document, the signatory is confirming that the state entity certifies the intended GenAI use case, its risk level, and understands that all procurements are mandated to comply to all CDT-published security and privacy policies (SAM Sections 5100 and 5300 - 5399).

Labor Relations Engagement

California state hiring authorities are responsible for ensuring that the implementation of changes affecting state employees' terms and conditions of employment complies with relevant policies and codes. For specific guidance on applicable labor relations policies and procedures, please coordinate with your labor relations office or CalHR Labor Relations.

☐ *As the CIO (or equivalent), I acknowledge that our Department's Labor Relations officer (or equivalent) has received and reviewed a copy of this 5305-F Risk Assessment form.*

_____ ISO Name (print)	_____ CIO Name (print)
_____ _____ ISO Signature (or equivalent) Date	_____ _____ CIO Signature (or equivalent) Date

*End of Risk Assessment Part 1 – Only continue if the GenAI Risk Assessment is **Moderate** or **High**.*

IV. Risk Assessment - Part 2

Generative Artificial Intelligence Risk Assessment Part 2:

This section completed by the Chief Information Officer (CIO)

Instructions:

- State entities complete this form when required.
- Complete SIMM 5305-F, Part 2 if the **GenAI system risk level is rated Moderate or High.**
- After completion, submit a Case via the New Technology Consultation and Assessment request, in the CDT IT Service Portal for all risk levels. When the request has been processed, a CDT Customer Engagement Services (CES) representative will reach out to provide a secure location to upload required documents as part of the GenAI consultation process.
- **Important:** Once completed, this form is confidential and may be exempt from disclosure pursuant to Government Code sections 7929.210 and 8592.45.

Mandatory Minimum Safeguards:

Instructions:

- Checkmark all the safeguards your system is in compliance with. All safeguards must be met and will be discussed with CDT during consultation for compliance.

Yes	No	N/A	
☐	☐	☐	The GenAI system workflow includes human verification to ensure accuracy and factuality of the output.
☐	☐	☐	The GenAI system will not have the potential to degrade public services.
☐	☐	☐	The GenAI system will not adversely impact the availability of resources and services provided by the State of California.
☐	☐	☐	If the GenAI system is a shared system, is there an existing data-sharing agreement between parties including roles & responsibilities for data owner, custodian, user, etc.?
☐	☐	☐	User accounts for the GenAI tool is managed by a state-owned identity access and management tool (e.g. Active Directory).
☐	☐	☐	Business services are not contingent on the system's use. In the event of system failure or inaccurate results, the State of California can continue to provide the same level of services without disruption.
☐	☐	☐	The state entity has safeguards in place to protect data used by the GenAI tool from being exposed to the internet.
☐	☐	☐	The state entity uses safeguards that comply with the state-defined security parameters for NIST SP 800-53, SIMM 5300-A, and SAM Section 5300.5.
☐	☐	☐	Cloud-based GenAI systems comply with Cloud Computing Policy SAM 4983.1 and Cloud Security Guide SIMM 140, which states that all data will remain in the United States and that no remote access will be allowed outside of the United States.
☐	☐	☐	All remote access uses Multi-Factor Authentication (MFA) and complies with the Telework and Remote Access Security Standard (SIMM 5360-A).
☐	☐	☐	All confidential, sensitive, or personal information is encrypted in accordance with SAM 5350.1 (Encryption) and SIMM 5305-A (Information Security Program Management Standard) and at the necessary level of encryption for the data classification pursuant to SAM 5305.5 (Information Asset Management).
☐	☐	☐	All data, hardware, software, internal systems, and essential third-party software, including for on-premises, cloud, and hybrid environments, are aligned with a zero-trust architecture model in accordance with NIST 800-27.

☐	☐	☐	All data is subject to Civil Code 1798.99.80 – 1798.99.89 and will not be sold or advertised to data brokers.
☐	☐	☐	Unless specified in the contract, prompts or Generated Data resulting from such Prompts constitute a Work Product. Contractors may not use, copy, modify, distribute, or disclose any such Prompts or Generated Data for any purpose other than performing their obligations under the Contract unless expressly authorized by the State in writing.
☐	☐	☐	To the extent any Prompts or Generated Data constitute Work Product, the State will retain Government Purpose Rights.
☐	☐	☐	The GenAI system will opt out of any data collection and model training features that may be used to train commercial instances of GenAI systems.
☐	☐	☐	GenAI output will not infringe on copyright or intellectual property laws and is compliant with open-source licenses, if applicable. GenAI output will be cited (from credible sources) if any statements used as facts are generated and published for consumer use. All generated images and videos will cite any GenAI used in their creation, even if the images are substantially edited afterward.
☐	☐	☐	The GenAI system will not spoof or engage in fraud, including deepfake creation, impersonation, phishing, other social engineering, or manipulation of other GenAI systems.
☐	☐	☐	The GenAI system is designed to avoid generating or creating illicit content that may be controversial, subjective, or potentially not widely accepted by the public.
☐	☐	☐	The GenAI system will not improperly systematically, indiscriminately, large-scale monitor, surveil, or track individuals.

Details of Transparency:	
(a) What mechanism will the GenAI system use to notify a user that they are interacting with a GenAI system rather than a human?	**(b)** What mechanisms can be used to audit the system and its data?
(c) How will the system disclose to the customer that the data generated is by GenAI? (e.g. watermark, banner)	**(d)** How will customers receive an output, and what is the mechanism to correct or appeal an error?

Human Oversight and Monitoring:	
(a) How will system owners identify and mitigate hallucinations and that data outputs are accurate and factual? What ability will system owners have to accept, reject, and correct data?	**(b)** Will the system be publicly accessible or only within a state-managed environment? Who is the intended audience, and will it impact a specific group of individuals or communities? Briefly explain.
(c) How will system owners test, evaluate, and verify that the GenAI system's original designated GenAI Risk Level will or has not changed? (e.g., changes to use, data, privacy, cost)	**(d)** Will logs be available in a non-proprietary format, that can be ingested into a Security Information and Event Management (SIEM) tool? Briefly explain.

Ensuring Equity:	
(a) Does the output of the system make decisions that impact access to, or approval for, housing or accommodations, education, employment, credit, health care, or criminal justice? If so, please describe.	**(b)** Will the output of the system make decisions that factor in the Diversity, Equity, Inclusion, and Accessibility (DEIA) of individuals? Briefly explain.
(c) Will the system impact a minor under the age of 18? Briefly explain.	**(d)** Will system decisions impact public safety? (e.g., water pollution metrics)? Briefly explain.

Signatures and Acknowledgement (Required Completion)

Required Signatures for Risk Assessment Part 2
You are required to ensure that the Acceptable Use Policy is updated to address GenAI. A completed SIMM 5310-C Privacy Threshold Assessment and Privacy Impact Assessment must be accessible upon request. If additional GenAI features are enabled in the future beyond those outlined in this document, a new SIMM 5305-F must be submitted for review. By signing this document, the signatory is confirming that the state entity certifies the intended GenAI use case, its risk level, and understands that all procurements are mandated to comply to all CDT-published security and privacy policies (SAM Sections 5100 and 5300 - 5399).
Labor Relations Engagement
California state hiring authorities are responsible for ensuring that the implementation of changes affecting state employees' terms and conditions of employment complies with relevant policies and codes. For specific guidance on applicable labor relations policies and procedures, please coordinate with your labor relations office or CalHR Labor Relations. ☐ *As the AIO (or equivalent), I acknowledge that our Department's Labor Relations officer (or equivalent) has received and reviewed a copy of this 5305-F Risk Assessment form.*

_____ AISO Name (print)	_____ AIO Name (print)
_____ _____ AISO Signature (or equivalent) Date	_____ _____ AIO Signature (or equivalent) Date

End of Risk Assessment part 2

V. GenAI Use Cases & Safeguard Samples

The following section lists high-level categories for the wide variety of functionality for GenAI with sampled public sector use cases. The example use cases are only intended to help illustrate the potential uses for state government adoption of GenAI tools.

While these use cases and safeguards offer a solid foundation for mitigating risks associated with GenAI tools, state entities are **required** to tailor their approach to the specific use cases and products that they employ. Given the diverse range of GenAI applications and their associated risks, it is crucial to conduct a thorough risk assessment for each specific use case. By analyzing factors such as system categorization, data type and business process, entities can optimize their security posture and minimize potential risks.

Safeguards (Common)

This section outlines a list of general risks along with their corresponding safeguards. These risks are not tied to any specific use case. State agencies/entities should review this section in conjunction with the section relevant to their specific use case to gain a better understanding of potential risks and their mitigation strategies.

Common Safeguards	
Risk: Inadvertent exposure of sensitive or confidential information, either through data input, processing, storage, or generated outputs.	**Mitigation**: Anonymize or obfuscate sensitive information before inputting it into GenAI systems. **Mitigation**: Implement strict role-based access to GenAI platforms and their outputs, limiting access to GenAI tools and data to authorized personnel, enforcing the principle of least privilege. **Mitigation**: Use end-to-end encryption for all data interactions with GenAI tools. **Mitigation:** Apply data masking or redaction to sensitive data before processing it through GenAI tools. **Mitigation**: Avoid using third-party or public GenAI solutions for sensitive use cases. **Mitigation**: Ensure acceptable use policies include a data handling component that addresses the use of sensitive and confidential information with publicly available GenAI tools; including the use of internally hosted generative AI tools that are on commercial clouds that could still expose that data to the internet. **Mitigation**: Review vendor privacy statements to identify what type of information is being shared with affiliates and that privacy and confidential data does not get shared. **Mitigation:** Use vendors with clear data protection and non-retention policies. **Mitigation**: Obtain copies of vendor certificates to ensure their stated security compliance matches their security certificate (e.g., vendor claims SOC Type2 compliance but uses a certificate from their hosting provider and not their own). **Mitigation**: Review data sharing statements to ensure data transmitted does not leave the continental United States. **Mitigation:** Review vendor privacy statements to ensure that the data is encrypted in transmit and at rest. **Mitigation:** Opt for in-house or private cloud-based GenAI solutions when handling sensitive data. **Mitigation:** Require review of outputs by network security professionals to ensure that no sensitive details are unintentionally revealed or misrepresented. **Mitigation:** Deploy GenAI tools in isolated, secure environments without external internet connectivity to prevent data leaks. **Mitigation:** Train employees about appropriate usage and discourage inputting personal or sensitive data.

Risk: GenAI may produce errors, inconsistencies, or outputs that do not align with the business context, leading to flawed decisions.	**Mitigation:** Require subject matter expert (SME) review and approval of all outputs before implementation. **Mitigation:** Use well-curated, domain-specific datasets to improve the model's understanding and accuracy. **Mitigation:** Integrate tools to validate outputs for correctness and alignment with predefined rules. **Mitigation:** Perform regular audits of GenAI-generated data to ensure continued accuracy and to detect any emerging errors or discrepancies. **Mitigation:** Include metadata or disclaimers with summaries indicating potential limitations or areas requiring manual review. **Mitigation:** Review laws and regulations to ensure that GenAI tools are following disclosures and disclaimers from regulatory entities (government codes, assembly bills, election laws, etc.) **Mitigation:** Allow users to flag inaccuracies and provide feedback for system improvement. **Mitigation**: Configure the tools to provide links back to the source material. **Mitigation**: Train staff on how to verify outputs for accuracy and find source of truth. **Mitigation:** Use a curated set of validated responses for frequently asked questions.
Risk: Outputs may reflect biases present in the training data, leading to unfair or discriminatory results.	**Mitigation:** Regularly audit GenAI systems for bias and apply techniques to reduce bias in outputs. **Mitigation:** Ensure datasets are inclusive, representative, and free of systemic bias. **Mitigation:** Use explainable AI (XAI) techniques to assess and justify outputs. **Mitigation:** Provide clear and transparent mechanisms for individuals to appeal decisions made using GenAI tools. **Mitigation:** Involve DEIA professionals in the review and validation of outputs used for decision-making. **Mitigation:** Use representative test cases to evaluate the model's behavior across different demographics or situations.
Risk: Excessive dependence on GenAI could reduce human oversight and critical thinking, increasing susceptibility to undetected errors.	**Mitigation:** Combine human expertise with GenAI insights rather than fully automating decisions. **Mitigation:** Educate users on the limitations of GenAI and the importance of maintaining human judgment. **Mitigation:** Establish manual processes to take over in case GenAI fails or outputs unreliable results. **Mitigation:** Use models that provide clear explanations for their decisions to enable better oversight. **Mitigation:** Implement new GenAI-driven decision-making systems in controlled environments before full deployment to evaluate reliability and fairness.
Risk: GenAI systems and their outputs may be targeted by attackers to steal or manipulate sensitive information.	**Mitigation:** Host GenAI systems in secure environments with firewalls, intrusion detection, and vulnerability scanning. **Mitigation:** Conduct periodic reviews to identify and address vulnerabilities in GenAI systems. **Mitigation:** Limit the flow of sensitive data to only the required systems and processes.
Risk: GenAI outputs may unintentionally breach compliance with laws and regulations, e.g., data privacy regulations, procurement standards.	**Mitigation:** Use automated tools to validate GenAI outputs against relevant laws and standards. **Mitigation:** Have legal teams review processes and outputs that may affect compliance. **Mitigation:** Maintain detailed documentation of how data is processed by GenAI to ensure auditability.

Risk: GenAI-generated content may unintentionally replicate copyrighted material or proprietary information from training data.	**Mitigation:** Use plagiarism detection tools to ensure outputs are unique. **Mitigation:** Train models only on licensed, public domain, or organizationally owned datasets. **Mitigation:** Include IP experts to review outputs for potential infringement risks.
Risk: Outputs may include insensitive, offensive, or harmful content, leading to ethical concerns or reputational harm.	**Mitigation:** Implement safeguards to identify and block offensive or inappropriate outputs. **Mitigation:** Transparently disclose the use of GenAI and its role in decision-making. **Mitigation:** Use diverse and representative datasets to train GenAI to minimize the risk of generating harmful stereotypes. **Mitigation:** Have culturally aware reviewers assess content targeted at specific demographics.
Risk: It may be unclear who is accountable for errors or harm caused by GenAI-generated outputs.	**Mitigation:** Establish clear policies assigning responsibility for validating and approving GenAI outputs. **Mitigation:** Maintain logs that track who used the GenAI system, what inputs were provided, and how outputs were utilized. **Mitigation:** Periodically assess the system's performance and outputs to identify areas for improvement.
Risk: The computational demands of GenAI-based solutions (e.g., processing large datasets or analyzing complex code samples) might strain system resources, leading to reduced system performance or slower response times.	**Mitigation:** Use lightweight GenAI models or distributed processing to reduce system impact. **Mitigation:** Implement tools to monitor and manage resource utilization in real time. **Mitigation:** Ensure backup systems are available in case of resource exhaustion during high-demand periods.
Risk: The public (residents) may struggle to understand the reasoning behind GenAI-generated eligibility determinations, leading to confusion or mistrust.	**Mitigation:** Integrate tools that provide clear explanations of how eligibility decisions were made. **Mitigation:** Maintain detailed logs of input data, the model's decision process, and outcomes for auditability and accountability. **Mitigation:** Ensure eligibility decisions are conveyed in a clear and accessible manner to applicants. **Mitigation:** Ensure all GenAI input and output data is validated by an adequately qualified subject matter expert.
Risk: The GenAI may not be fully accessible to individuals with disabilities or may fail to support diverse user groups effectively.	**Mitigation:** Ensure the GenAI tool complies with accessibility standards, such as WCAG, to accommodate users with disabilities. **Mitigation:** Provide support for multiple languages or dialects, depending on the demographic it serves. **Mitigation:** Conduct user testing with diverse groups to optimize ease of use and accessibility.
Risk: GenAI might fail to identify when an issue or interaction requires escalation to a human representative, leading to unresolved or improperly handled situations.	**Mitigation**: Implement clear rules for escalating complex or sensitive interactions to human agents. **Mitigation:** Use intent recognition models to identify and flag interactions that may require escalation. **Mitigation:** Collect user feedback to identify gaps in the escalation process and improve the system.

Safeguards (Use-Case Specific)

Grammar correcting tools

e.g., grammar assistant tools, rewording and revising text, providing alternative text

Risk: Inadvertently exposing data to the internet due to working with sensitive and confidential information throughout the course of work.	**Mitigation**: Host the tools internally on the state network. **Mitigation**: Leverage API instead of working directly with the vendor's website. **Mitigation**: Update acceptable use policy to address data handling of sensitive and confidential information with grammar correcting tools. **Mitigation**: Disable the tool by default and allow user to enable as needed. **Mitigation**: Read privacy statements to identify what type of information gets processed by the vendors cloud, how long it is stored, and how it gets deleted. **Mitigation**: Purchase the right license level that gives the state more configuration options over the tools. **Mitigation**: Negotiate and redline contracts on data collection and data sharing with vendors to minimize the data exposure.

Mission, State, or Critical Infrastructure

e.g., handling systems classified as mission, state, or critical infrastructure	
Risk: Tools may be hosted in an environment with insufficient safeguard that expose mission, state, or critical infrastructure.	**Mitigation**: Ensure that the safeguards implemented match the level of the criticality of the mission, state, or critical infrastructure (e.g., mission critical web apps at moderate safeguards replicate to a moderate development environment). **Mitigation:** Segment the network so that systems categorized as low, moderate, and high have security zones that match low, moderate, and high safeguards. **Mitigation:** Adopt a Zero Trust security model, where every interaction with critical infrastructure—whether initiated by GenAI or a human operator—requires verification before granting access. **Mitigation:** Implement comprehensive logging and monitoring of GenAI activities within mission-critical apps/systems, including auditing interactions that involve sensitive configuration or system design tasks.

Network Analysis Tools

e.g., packet inspection, system monitoring, intrusion prevention/detection	
Risk: GenAI models could generate false positives (incorrectly flagging benign activities as threats) or false negatives (failing to detect actual threats), potentially undermining the reliability of intrusion detection and prevention systems.	In addition to the safeguards given for inaccuracies in GenAI performance in Common Risks/Safeguards section, consider the following: **Mitigation**: Regularly retrain GenAI models using updated threat intelligence to reduce false positives and false negatives. **Mitigation:** Implement a human review process for flagged threats to validate or dismiss potential issues. **Mitigation:** Fine-tune detection thresholds based on feedback and observed network behavior to reduce incorrect alerts.

Spam and Malware Detection

e.g., web gateway filtering, email gateway filtering e.g., anti-virus, anti-malware, endpoint detection & response	
Risk: GenAI models may incorrectly identify legitimate emails, web content, and files as spam or malware, causing disruptions to business communications and workflows.	In addition to the safeguards given for the risk of inaccuracies in GenAI performance in Common Risks/Safeguards section, consider the following: **Mitigation**: Implement mechanisms for users to report false positives and retrain the model based on this feedback. **Mitigation:** Allow for exceptions through whitelisting of critical email addresses, domains, web content, or files.
Risk: The model may fail to detect sophisticated spam, phishing attempts, malicious content, or malware potentially leading to security breaches.	In addition to the safeguards given for the risk of inaccuracies in GenAI performance in Common Risks/Safeguards section, consider the following: **Mitigation**: Combine traditional rule-based filtering with GenAI models to improve detection accuracy.

	Mitigation: Continuously retrain GenAI with up-to-date spam, phishing patterns, and malware samples to identify evolving threats. **Mitigation:** Augment spam filtering and malware detection solutions with tools that monitor for malicious activity post-delivery (e.g., link-click analysis, sandbox execution).
Risk: Attackers may deliberately craft spam or malicious content to bypass GenAI filters and poison the model's learning process, reducing its effectiveness.	**Mitigation:** Regularly test the spam filter and malware detectors against adversarial examples to ensure robustness. **Mitigation:** Vet incoming training data to avoid incorporating malicious or misleading inputs. **Mitigation:** Use isolated environments to test model updates before deploying them into production.

Reference Generating Tools

e.g., laws, case law, judgements, policies, procedures	
Risk: GenAI tools may generate references based on outdated laws, policies, or judgments, or omit critical updates, leading to incorrect or incomplete advice.	**Mitigation**: Continuously update the training data with the latest legal, policy, and procedural information. **Mitigation:** Validate generated references against official or authoritative sources before use. **Mitigation:** Include metadata indicating when the reference was last verified to highlight potential currency issues.
Risk: GenAI may generate references that are applicable to a different jurisdiction or context, leading to misapplication of laws or policies.	**Mitigation:** Train GenAI models with data specific to the relevant contexts. **Mitigation:** Require users to provide jurisdictional or contextual information to guide GenAI outputs. **Mitigation:** Require legal experts to confirm applicability and accuracy of references in the given context.

Content Creation Tools

e.g. image creators, video creators, ads, marketing	
Risk: GenAI tools may collect and process personal data, which could be vulnerable to breaches.	**Mitigation**: Redline contracts to minimize data collection. **Mitigation**: Review license levels to identify if there are product versions that enable administrative configuration by state staff to restrict data collection. **Mitigation**: Use template contracts that do not have personal, proprietary or organizational information.
Risk: GenAI tools may produce content that is misleading, manipulative, or violates ethical marketing practices, potentially harming trust or misleading consumers.	**Mitigation:** Establish clear organizational policies for acceptable content creation, ensuring compliance with advertising standards and consumer protection laws. **Mitigation:** Require a review process to validate that content aligns with ethical and legal standards before publication. **Mitigation:** Disclose when content has been AI-generated to maintain trust. **Mitigation:** Prepare plans to mitigate fallout from potential reputational harm caused by AI content.

Translation Tools

e, g., meeting summarization, audio/video to text	
Risk: AI tools may collect and process sensitive information, such as personal conversations and proprietary business data.	In addition to the safeguards given for the risk implying inadvertent exposure of sensitive information given in Common Risks/Safeguards section, consider the following: **Mitigation**: Obtain consent before recording meetings. **Mitigation:** Limit how long audio, video, and transcription data are stored and ensure proper deletion protocols.

Generative AI Platforms & Code Analysis

e.g., static and dynamic code analysis, code generation tools, code assistant tools e.g., developer tools	
Risk: AI models can introduce bugs and backdoors over time, or do not adhere to best practices of security (e.g., lack of input validation, hardcoding credentials)	**Mitigation**: Generate small batches of code at a time and have staff review it to ensure that they understand everything that is being generated. **Mitigation**: Avoid generating thousands of lines of code that can't be thoroughly inspected. **Mitigation**: Add comments to each line of code that is AI generated that explains what it is doing to confirm staff understand the code that was generated.
Risk: mismatched safeguard levels with supported computing resources (e.g., development server has Low safeguards, but the supported systems are classified as moderate, thereby exposing moderate development (e.g., code, data records) in the development environment with Low safeguards.	**Mitigation**: Deploy the same level of safeguards as the systems that are being supported. **Mitigation:** Sanitize data being used in development environments. **Mitigation**: Implement strong access controls and strong authentication mechanisms to limit unauthorized access to LLM model repositories and training environments. **Mitigation**: Restrict the LLMs access to network resources, internal services and API's. **Mitigation**: Regularly monitor and audit access logs and activities related to LLM model repositories to detect and respond to any suspicious activities.
Risk: Developers may over-rely on AI-generated code, leading to reduced understanding of underlying logic and potential propagation of errors.	**Mitigation:** Require thorough documentation of all AI-generated code to promote understanding and accountability. **Mitigation**: Encourage collaborative review processes where one developer writes while another reviews the code (pair programming).
Risk: AI-generated code may suggest or introduce third-party libraries or dependencies that are outdated, insecure, or poorly maintained.	**Mitigation:** Use tools to identify and validate the security and maintenance status of suggested dependencies. **Mitigation:** Restrict AI tools to a pre-approved list of secure and well-maintained libraries. **Mitigation:** Regularly review third-party dependencies in generated code for vulnerabilities.
Risk: AI-generated code might inadvertently replicate copyrighted material or violate licensing terms, leading to intellectual property disputes.	**Mitigation:** Use automated tools to check generated code for licensing conflicts. **Mitigation:** Ensure AI models are trained on datasets with clear, appropriate licensing. **Mitigation:** Involve legal teams to review and approve policies governing the use of AI-generated code.
Risk: AI-generated code may be suboptimal, leading to performance issues such as increased memory usage or slower execution times.	**Mitigation:** Apply performance profiling tools to assess and optimize generated code. **Mitigation:** Compare AI-generated outputs against best-practice implementations to ensure efficiency (benchmarking). **Mitigation:** Allow developers to iteratively refine AI outputs for better performance.
Risk: Generated code might not comply with industry regulations, security standards, or organizational policies (e.g., SIMM 5300 series).	**Mitigation:** Incorporate compliance checks as part of the code review process. **Mitigation:** Train developers to understand applicable compliance requirements for their industry. **Mitigation:** Use AI tools that are specifically designed to comply with relevant regulations.
Risk: When errors arise in AI-generated code, it may be difficult to assign accountability or trace the origin of the problem.	**Mitigation:** Require AI-generated code to be committed to version control with clear labeling. **Mitigation:** Enable logging of AI tool activities to maintain traceability of code generation.

	Mitigation: Ensure developers take ownership of AI-generated code by requiring signoffs.
Risk: False positives (granting access to unauthorized users) or false negatives (denying access to legitimate users) due to errors in biometric recognition models.	**Mitigation:** Use biometrics in combination with other authentication factors (e.g., PIN, token). **Mitigation:** Regularly validate the accuracy of biometric algorithms with diverse datasets. **Mitigation:** Implement secondary authentication methods for denied users to verify their identity, aka fallback mechanisms.
Risk: Sensitive biometric data (e.g., fingerprints, facial data) could be exposed, stolen, or misused, leading to identity theft or privacy breaches.	**Mitigation:** Encrypt biometric data at rest and in transit using strong cryptographic methods. **Mitigation:** Store only necessary biometric templates instead of raw biometric data. **Mitigation:** Process data locally rather than transmitting it to external servers, where feasible. **Mitigation:** Implement strict access controls to ensure only authorized personnel can access biometric data.
Risk: LLM documents become stale resulting in inaccurate information being provided to staff or the public.	**Mitigation**: Create an LLM update and release strategy to minimize the gap between changes to documents and the re-training of the LLM model. **Mitigation**: Communication to staff when policies change prior to updates to the LLM.
Risk: Reliance on open LLMs data sources can weaken model accuracy due to malicious actors manipulating the training data.	**Mitigation**: Continuously benchmark the model's performance against established benchmarks to identify unexpected drops in accuracy or changes in behavior. **Mitigation**: Enforce privilege control on LLM access to backend systems. **Mitigation**: Segregate external content from user prompts and limit the influence when untrusted content is used. **Mitigation**: Maintain fine user control on decision making capabilities by LLM. **Mitigation**: Use content safety filters for prompt inputs and its responses. **Mitigation**: Use TLS to encrypt all HTTP-based network traffic. Use other mechanisms, such as IPSec, to encrypt non-HTTP network traffic that contains customer or confidential data.
Risk: Insufficient scrutiny of LLM output, unfiltered acceptance of the LLM output could lead to unintended code execution.	**Mitigation**: Treat the model as any other user. Adopt a zero-trust approach. Apply proper input validation on responses coming from the model to backend functions. **Mitigation**: Encode model output back to users to mitigate undesired code execution by JavaScript or Markdown.
Risk: Extensions, plugins, API's that are out of date or unknown can expose the development environment and reach mission, state, or critical infrastructure.	**Mitigation**: Ensure there is a vulnerability management policy in place. **Mitigation**: Ensure there is a vulnerability patch and update procedures in place. **Mitigation**: Carefully vet data sources and suppliers, including T&Cs and their privacy policies, only using trusted suppliers. **Mitigation**: Only use reputable plug-ins and ensure they have been tested for your application requirements. **Mitigation**: Implement sufficient monitoring to cover component and environment vulnerabilities scanning, use of unauthorized plugins, and out-of-date components, including the model and its artifacts.

Chatbots

e.g. used internally by entity for finding resources, getting advice on processes and procedures, locating documents.	
Risk: Employees may be unable to access critical internal resources if the chatbot is	**Mitigation:** Maintain alternative methods for accessing internal resources. **Mitigation:** Use load balancing to handle high usage volumes and ensure consistent availability.

unavailable due to outages or high demand.	**Mitigation:** Set up alerts for system downtime and have a defined escalation process to resolve issues quickly.
Risk: Staff input personal information into the chatbot to try and get tailored information.	**Mitigation**: Create or update the acceptable use policy and outline how the tool should be used. **Mitigation**: Add a one-time popup disclaimer that users acknowledge that the tool should be used for work related activities only.
Risk: Staff input personal and confidential information into prompts that get logged.	**Mitigation**: Update acceptable use to mitigate for misuse. **Mitigation**: Communicate to staff that inputs get logged and safeguard their own information by not entering personal information into GenAI tools.
Risk: Staff input questions related to bias or DEIA	**Mitigation**: provide default prompt responses letting staff know that the information is not available and to only use the tool as outlined in the acceptable use policy.
Risk: Staff use GenAI outputs to make decisions.	**Mitigation**: Add disclaimers, banners, and contact reminders that AI tools can make mistakes and to verify and validate the source of the outputs. **Mitigation**: Verify and validate the outputs by an adequately qualified subject matter expert.
Risk: Administrators see input/output logs that contain confidential information.	**Mitigation**: Ensure administrators who have access to the input/output logs have the right background clearance and security training to be able to view that data. **Mitigation**: Configure data anonymization.
Risk: Crafty inputs can train the back end LLM.	**Mitigations**: Create a test plan and include various crafty prompts that try to get the LLM to respond in an unintended manner.

Input Processing

e.g., inputting and processing of hiring information e.g., inputting and processing of personal or sensitive data in an open environment e.g., processing sensitive data for national security or intelligence purposes	
Risk: GenAI models may unintentionally introduce or perpetuate bias in hiring processes, leading to discriminatory practices (e.g., favoring or disfavoring candidates based on gender, ethnicity, or other protected characteristics).	In addition to the safeguards in Common Risks/Safeguards regarding bias: **Mitigation:** Clearly document and explain how hiring decisions are made to ensure fairness and compliance with anti-discrimination laws.
Risk: GenAI may inadvertently violate laws regarding equal opportunity employment by relying on non-compliant criteria for candidate assessment.	**Mitigation**: Incorporate legal reviews into the hiring workflows to verify that outputs comply with relevant employment regulations. **Mitigation**: Implement constraints within GenAI systems to disallow processing or consideration of protected attributes.
Risk: Sensitive data might be inadvertently exposed to unauthorized parties in an open environment (e.g., public Wi-Fi, shared workspaces, or non-secured devices).	In addition to the safeguards in Common Risks/Safeguards regarding exposure of sensitive information: **Mitigation:** Require secure connections (e.g., VPNs, HTTPS) when handling sensitive data in open environments. **Mitigation:** Ensure devices used in open environments are equipped with firewalls, antivirus software, and updated security patches. **Mitigation:** Implement Data Loss Prevention solutions to monitor and prevent sensitive data from being transmitted outside authorized channels. **Mitigation:** Train personnel to identify and mitigate risks of working with sensitive data in open environments (e.g., avoiding public Wi-Fi, locking screens when leaving devices unattended).
Risk: The compromise of GenAI systems or outputs could expose	In addition to the safeguards in Common Risks/Safeguards regarding exposure of sensitive information:

highly sensitive national security or intelligence data, potentially endangering public safety or operational integrity.	**Mitigation:** Use physically isolated environments to process national security data, ensuring GenAI tools are disconnected from external networks. **Mitigation:** Implement stringent clearance procedures for individuals accessing GenAI systems used for national security. **Mitigation:** Use data compartmentalization to ensure that only specific, authorized segments of data are accessible to any given process or user.
Risk: Adversaries or insiders might intentionally manipulate GenAI outputs to mislead or disrupt decision-making processes related to national security.	**Mitigation**: Require multiple layers of review and cross-validation by authorized intelligence personnel. **Mitigation**: Maintain immutable logs of GenAI inputs, processing, and outputs for forensic auditing. **Mitigation:** Frequently audit and test models to ensure they are not compromised or manipulated.
Risk: GenAI systems may become specific targets of nation-state adversaries or advanced threat actors aiming to disrupt or infiltrate intelligence workflows.	**Mitigation:** Deploy advanced cybersecurity measures such as intrusion prevention systems, endpoint detection and response (EDR), and zero-trust architecture (ZTA). **Mitigation:** Use updated threat intelligence feeds to anticipate and mitigate evolving APT tactics. **Mitigation:** Establish robust recovery mechanisms to ensure continuity of operations in the event of a successful attack.

Confidential Data Handling

e.g., processing or analyzing medical records or health data	
Risk: GenAI systems may inadvertently process or generate outputs that violate health-specific legal or regulatory requirements for data protection.	In addition to the safeguards in Common Risks/Safeguards regarding lack of compliance with laws and regulations: **Mitigation:** Ensure that GenAI systems are explicitly designed to comply with health data regulations like HIPAA. **Mitigation:** Create and enforce tailored policies for the use of GenAI in healthcare contexts.
Risk: GenAI may produce outputs that misinterpret medical data (e.g., symptoms, diagnoses, treatment plans), leading to incorrect or unsafe conclusions.	In addition to the safeguards in Common Risks/Safeguards regarding inaccuracies: **Mitigation**: Limit GenAI usage to non-critical medical contexts unless extensively validated.
Risk: GenAI's use of health data could lead to unintended ethical dilemmas, such as perpetuating biases in diagnoses or treatments.	In addition to the safeguards in Common Risks/Safeguards regarding biases: **Mitigation:** Obtain explicit patient consent before using their health data in GenAI systems. **Mitigation:** Involve ethics boards to oversee and evaluate the application of GenAI in healthcare.

Resident (Public) Facing

e.g., chatbots that residents (public) interact with to find web resources and information e.g., direct contact with the public, customer service, public relations, jurisprudence e.g., output provides recommendations, legal, tax, regulatory compliance advice, benefit qualifications	
Risk: Crafty inputs can train the back end LLM.	**Mitigations**: Create a test plan and include various crafty prompts that try to get the LLM to respond in an unintended manner.
Risk: liability for the actions or statements of the GenAI tools used by residents.	**Mitigation**: Include legal in the review process. **Mitigation**: Provide a banner somewhere on the tool that indicates something along the lines of “GenAI can make mistakes, double-check important information.”
Risk: liability for non-compliance with laws and regulations.	**Mitigation**: Review all relevant Generative AI laws and regulations and incorporate requirements into the build of the tools (e.g., statements, disclaimers, banners, that are in Gov Codes).

Mobile Device

e.g., mobile GenAI solutions and GenAI app Downloads	
Risk: GenAI apps may be downloaded and misused by staff.	**Mitigations**: Establish an acceptable use policy outlining permitted and prohibited downloads on company devices. **Mitigations** Configure devices to disable app store downloads prior to issuing phones. **Mitigations**: Restrict all app downloads to the company portal. **Mitigations:** Enforce the use of company certificates as a prerequisite for downloading apps on devices.
Risk: GenAI may be built into mobile devices resulting in inadvertent disclosures of sensitive & confidential info	**Mitigations:** Disable GenAI mobile capabilities, including GenAI intelligence features.
Risk: Work and Personal data may be mixed causing data ownership issues.	**Mitigations:** Configure devices to prevent logging out of company accounts, ensuring mobile devices remain subject to mobile device management (MDM) configurations.

VI. Definitions

Relevant definitions for this guidance are available in SAM 4819.2 and 5300.4

Generative Artificial Intelligence means an artificial intelligence system that can generate derived synthetic content, including text, images, video, and audio that emulates the structure and characteristics of the system's training data.

Human Verification is the process in which one or more people review and validate the output generated by an automated system, such as a GenAI tool, to ensure its accuracy, correctness, and alignment with factual information before it is used or implemented. This step acts as a safeguard against errors or misleading content produced by the automated system.

Personal Identifiable Information is information that can be used to distinguish or trace an individual's identity, either alone or when combined with other information that is linked or linkable to a specific individual.

Proprietary Information is material and information relating to or associated with a company's products, business, or activities, including but not limited to financial information; data or statements; trade secrets; product research and development; existing and future product designs and performance specifications; marketing plans or techniques; schematics; client lists; computer programs; processes; and know-how that has been clearly identified and properly marked by the company as proprietary information, trade secrets, or company confidential information. The information must have been developed by the company and not be available to the Government or to the public without restriction from another source.

Redlining is the process of marking up a contract with suggested changes and revisions. It's usually done during contract negotiations.

Resident Facing Service refers to any service or application that interacts directly with residents or the public. This typically involves systems or tools that provide information, make decisions, or deliver services to residents.

VII. References

Please refer to the latest version of the following resources when implementing this standard:

1. Generative Artificial Intelligence
   https://www.genai.ca.gov/

2. Algorithm Risk Management
   Ethics & Algorithms Toolkit (beta) (ethicstoolkit.ai)

3. Artificial Intelligence Risk Management Framework (AIRMF1.0)
   https://nvlpubs.nist.gov/nistpubs/ai/nist.ai.100-1.pdf

4. AI Risk Management Framework:
   https://www.nist.gov/itl/ai-risk-management-framework/

5. Definition of High-Risk Automated Decision System:
   https://legiscan.com/CA/text/AB302/id/2814759

6. Executive Department State of California Executive Order N-12-23
   https://www.gov.ca.gov/wp-content/uploads/2023/09/AI-EO-No.12-_-GGN-Signed.pdf

7. Federal Information Processing Standards, Standards for Security Categorization of Federal Information and Information Systems (FIPS 199)
   https://nvlpubs.nist.gov/nistpubs/fips/nist.fips.199.pdf

8. California Government Operations Agency website
   GovOps | Government Operations (ca.gov)

9. NIST Special Publication 800-53 Security & Privacy Controls for Information Systems & Organizations
   https://csrc.nist.gov/pubs/sp/800/53/r5/upd1/final

10. Statewide Administrative Manual (SAM) policies:
    https://www.dgs.ca.gov/en/Resources/SAM/TOC/

11. Statewide Information Management Manual (SIMM) policies:
    https://cdt.ca.gov/policy/simm/#SIMM

12. San José Generative AI Guidelines
    https://www.sanjoseca.gov/home/showpublisheddocument/100095/638255600904300000/

13. San José Digital Privacy and GenAI Manual
    https://www.sanjoseca.gov/home/showpublisheddocument/82093/637889898788170000/

14. State of California Benefits and Risks of Generative Artificial Intelligence Report
    State of California Benefits and Risks of Generative Artificial Intelligence Report

VIII. Questions

Please reference SIMM 71B for questions regarding procurement. For all other inquiries and implementation of this standard, please contact the California Department of Technology, Office of Information Security at Security@state.ca.gov.