State of California

Department of Technology

Office of Information Security

Generative Artificial Intelligence Risk Assessment

SIMM 5305-F

March 2024

REVISION HISTORY

Revision	Date of Release	Owner	Summary of Changes
Initial Release	March 2024	California Office of Information Security – Payam Hojjat	Initial Release of SIMM 5305-F Generative Artificial Intelligence Risk Assessment

TABLE OF CONTENTS

I. INTRODUCTION

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, Generative Artificial Intelligence (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, The White House Blueprint for an GenAI Bill of Rights, The White House EO on Safe, Secure, and Trustworthy Artificial Intelligence, and the National Institute of Standards and Technology (NIST) Artificial Intelligence Risk Management Framework.

GenAI Use Cases:

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 that California provides to its citizens. 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 to only 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. They will incorporate feedback from impacted stakeholders, serving as guidance to better streamline the use of this technology. Looking ahead, we will evaluate additional potential use cases to maximize the benefits of GenAI and continually improve service to the public.

The following table 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:

Table 1: GenAI Use Cases

Operational Need	GenAI Outcome	Common GenAI Use Cases
Content generation (text, image, video)	Generates completely novel content, instead of remixing and modifying existing content.	• Generate public awareness campaign materials like flyers, website content, posters, and videos. • Generate visualizations of data.

Chatbots	Leverages conversational models trained on massive dialogue datasets. Can have coherent discussions and execute tasks via conversation naturally.	● Build a virtual assistant for common constituent questions. ● Voice enabled digital assistance. ● Create chatbot to guide users. through services in their preferred language. ● Increase first-call resolution for state service centers. ● Reduce call wait and handle time at state customer service centers. ● Create greater language access equity for program beneficiaries.
Data analysis	Finds insights and relationships in data through learned knowledge about the world, without hand-coded rules or labeled training data.	● Analyze healthcare claims or tax filing data to detect fraud. ● Analyze network activity logs, identify cybersecurity anomalies and threats, and propose remediation actions. ● Ticket triaging. ● Root cause identification. ● Resolution recommendation from historical tickets.
Explanations and Tutoring	Generates natural language explanations and tutoring through dialogue without human-authored content.	● Explain program eligibility to potential enrollees. ● Provide interactive tax assistance.
Personalized Content	Leverages user data, information and/or models to adaptively generate personalized content without explicit rules or large amounts of user data.	● Auto-populate tax information and filing instructions based on a person's needs. ● Help auto-populate public program applications based on a person’s situation and household composition.
Search and Recommendation	Uses contextual cues to improve search relevance and provide useful recommendations.	● Searching or matching state code regulations concerning specific topics. ● Recommend government services based on eligibility. ● Search regulations nationwide.
Software code generation	Generates code by learning underlying structure and patterns of code, without the need for human written examples. Can expand short descriptions into full programs.	● Translate policy specifications, such as Web Content Accessibility Guidelines (WCAG) and Americans with Disability Act (ADA) requirements, into software code. ● Generate data transformation scripts from instructions.

		● Accelerate adoption of human-centered design in state web-based forms and pages. ● Reduce administrative costs and burden to developing and maintaining best-in-class state government websites.
Summarization	Does not require human-written summaries as training data. Can learn underlying patterns of language to generate summaries.	● Summarize public comments to identify key themes. ● Summarize public research to inform policymakers. ● Summarize statutory or administrative codes.
Synthetic data generation	Allows generation of new diverse, anonymized data from existing datasets for analysis and experimentation.	● Generate synthetic patient data for training healthcare AI. ● Generate simulated tax records for training tax auditing AI.

Source: State of California Benefits and Risks of Generative Artificial Intelligence Report, November 2023.

Risk Assessment Framework:

The following references the NIST Risk Assessment Framework and reflects California's GenAI Risk Management Principles as outlined in the State of California Generative AI Toolkit for Procurement, Use, and Training.

The following risk assessment enables state entities to address the Quality, Safety, and Security Controls by assessing specific risk factors around GenAI systems. However, it is up to the entity to identify their risk tolerance and apply risk mitigation strategies that align with their organizational acceptable risk standards.

Potential risks associated with GenAI differ upon each instance and use case. This GenAI risk assessment is determined by two factors including data type and expected use of the data. Cross examining these factors will determine the level of GenAI risk, what security controls need to be implemented, and how the system will be categorized - Low, Moderate, or High. The factors are:

1. Information Type: The risk of data breach that can occur if the information becomes compromised and accessible to an unauthorized party based on its data classification. GenAI data also includes inputs to systems that are used to train or tune a large language model (LLM) as well as prompts submitted to an LLM through an application interface.

2. Information Expected Use: There is a risk when utilizing GenAI output for decisions, tasks, or services. The ramifications of risk vary depending on its use and can generate biases, misinformation, and inaccurate results. These inaccuracies may adversely impact diversity, equity, inclusion, accessibility (DEIA) decisions throughout departments. This may put individuals at a disadvantage if factors like race, age, gender identity, or disability come into play for a decision.

State entities must assign a level of risk for each GenAI use case or system. The assigned level will help an entity understand the permissibility and risk associated with the GenAI use case. Assignable risk levels include High (Red), Moderate (Yellow), and Low (Green). Table 2 below provides the criteria used to assign levels and examples for each:

Table 2: GenAI Risk Level

Risk Level	Criteria Used to Assign Level	Examples
High (Red) Data loss could pose a severe or catastrophic adverse effect.	Input and processing of confidential information. Resident-facing service decision making using public data is also considered a high-risk level.	• Decision making that impact financial or benefit claims • Evaluating a proposal • Inputting and processing of personal or sensitive data in a state managed environment • Inputting and processing of personal or sensitive data in an open environment • Decisions that could affect public safety • Output data that impacts the decision making for privacy or DEIA related services or processes • Authentication and biometric identification
Moderate (Yellow) Data loss could still pose serious adverse effects.	Non-identifying and non-confidential information, that is potentially human related, used to make resident-facing service decisions.	• Drafting organizational documents with public information (for state use) • Human interaction with a chatbot • Static code analysis • Output generated for public use
Low (Green) Data loss could pose limited adverse effects.	Non-human related data that is used for system and technology processing optimization.	• Spam filtering • Malware detection • Network Analysis • Grammar correcting tools

II. RISK ASSESSMENT PART 1

Generative Artificial Intelligence Risk Assessment Part 1: *This section to be completed by the Chief Information Officer (CIO)*	
Instructions: • This risk assessment is required for all new GenAI procurements and acquisitions. • State entities complete SIMM 5305-F, Part 1 to determine the level of risk associated with a GenAI system. • Only complete SIMM 5305-F, Part 2 if the GenAI system risk level is rated Moderate or High. • **Important:** Once completed, this form is confidential and exempt from disclosure pursuant to Government Code sections 7929.210 and 8592.45.	
Section 1 - GenAI Description and Use Case:	
(a) Describe the project use case and problem, current process, and the impact of the desired outcome with the implementation of the GenAI system.	**(b)** Were there other options presented to solve the use case and/or problem? What other approaches to solving this problem were considered (if any) and what led to the decision to use GenAI to solve this problem?
(c) Will the GenAI system be shared or procured with any other state entity or third-party organization? If so, please specify whom and the use case of the project. (e.g. data custodian, data user, etc.)	**(d)** If the GenAI system is a shared system, is there an existing data sharing agreement?
(e) Has a Privacy Threshold Assessment (PTA) and Privacy Impact Assessments (PIA) (SIMM 5310 – C) been completed for the GenAI system?	

☐ Yes ☐ No
(f) If No was selected for Risk Assessment Part 1, Section 1 (e), please provide justification on why a PTA/PIA was not completed. *****Please note that if selected No, the GenAI Risk Level will immediately be marked as High and will require a CDT consultation.***

(g) Are funds appropriately allocated for this procurement, and are all related costs including development, integration, operation, maintenance, and potential scaling of the GenAI system identified and accounted for?	**(h)** Are ongoing costs contingent on a future Budget Change Proposal (BCP) or reallocation of resources?
☐ Yes ☐ No	☐ Yes ☐ No

GenAI Risk Level:

In addition to the GenAI Toolkit, use the previously answered questions to help determine the system's level of risk.

☐ Low ☐ Moderate ☐ High

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

III. RISK ASSESSMENT PART 2

Generative Artificial Intelligence Risk Assessment Part 2:

This section to be completed by the Chief Information Officer (CIO)

Instructions:

- State entities complete this form when required.
- Complete SIMM 5305F, Part 2 if the **GenAI system risk level is rated moderate or high.**
- I**mportant:** Once completed, this form is confidential and exempt from disclosure pursuant to Government Code sections 7929.210 and 8592.45.

Section 1 – Mandatory Quality and Security and Safety Controls:

Section 1 Instructions:

- Check mark all the controls that your system is in compliance with. All controls must be met and will be discussed with CDT during consultation for compliance.

☐ The GenAI system will have human verification to ensure accuracy and factuality of the output.

☐ The GenAI system will not impact physical equipment that may pose a risk to public health and safety.

☐ The GenAI system will not adversely impact the availability of resources and services provided by the State of California.

☐ If the GenAI system poses a risk to California security, national economic security, or national public health and safety, it has been reported to the federal government during the training of the model.

☐ State-owned user accounts are used for general use of the system to ensure segregation between public and personal records for future audit use and logging.

☐ Business services are not contingent on the use of the system. In the event of system failure or inaccurate results, the State of California can continue to provide the same level services without disruption.

☐ State entity has a data loss prevention system, and it will be analyzing input and training data for the GenAI system.

☐ State entity is using Federal Information Processing Standards (FIPS) and NIST Special Publication (SP) 800-53 in the planning, development, implementation, and maintenance of their information security programs. (California has adopted the NIST SP 800-53 as minimum information security control requirements to support implementation and compliance with the FIPS. Adoption of these standards will facilitate a more consistent, comparable, and repeatable approach for securing state assets; and create a foundation from which standardized assessment methods and procedures may be used to measure security program effectiveness.)

☐ State entity is using security controls complying with the State-defined Security Parameters for NIST SP 800-53, SIMM 5300-A and SAM Section 5300.5. Unless otherwise specifically noted the current final publications revisions (Rev.'s) are required. (California has adopted additional standards and procedures to address

more specific requirements or needs unique to California. These standards are referenced in the applicable policy section and maintained in the SIMM.)

- ☐ Cloud based GenAI systems shall comply to Cloud Computing Policy SAM 4893.1, which includes all data will remain in the United States and 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 compliant to a zero-trust architecture model in accordance with Government Code 11549.45.
- ☐ All data is subject to Civil Code 1798.99.80 – 1798.99.89 and will not be sold or advertised to data brokers.
- ☐ Input and prompt data will not be stored by the vendor for future prompt engineering.
- ☐ All generated output will be owned by the State of California.
- ☐ 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 any copyright or intellectual property laws and are compliant with open-source licenses as 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 must cite any GenAI used in their creation, even if the images are substantially edited afterwards.
- ☐ The GenAI system will not spoof or conduct acts of fraud including deepfake creation, impersonation, phishing and 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.
- ☐ State entity will provide sufficient training to stakeholders and customers for the use of the GenAI system.

Section 2 – Vendor Details:	
(a) Will the GenAI system be designed, developed, deployed, or maintained by a vendor or third party?	**(b)** How will the GenAI solution, residing on state infrastructure, be tested (including all systems interacting with AI)?
(c) What kind of access will the vendor provide to the system owner, if any?	**(d)** What type of model(s) and/or network(s) will be used in the GenAI system? Please reference all and explain their specific applicational use and purpose.
Section 3 – 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?	**(d)** How will customers receive an output, and what is the mechanism to correct or appeal an error?
(e) Data Output Standards:	**(f)** Level of Autonomy:
1. ☐ Data output from GenAI systems are analyzed and fact checked via a human reviewer before it is used for services. 2. ☐ The State of California will own all rights and intellectual property of data output and consultants are to release all ownership of data that was generated.	1. ☐ System operates automatically with no human intervention. 2. ☐ System operates automatically with occasional retrospective reviews by humans. 3. ☐ System produces recommendations but cannot act without human intervention.
Section 4 – 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?
(c) How will system owners test, evaluate, and	**(d)** Will logs be available in a non-proprietary

verify that the GenAI system's original designated GenAI Risk Level will or has not changed? (i.e. changes to use, data, privacy, cost, etc.)	format, that can be ingested into a Security Information and Event Management (SIEM) tool?
Section 5 – 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 factors in the Diversity, Equity, Inclusion, and Accessibility of individuals?
(c) Will the system impact minors under the age of 18?	**(d)** Will system decisions impact the environment? (e.g., water pollution metrics)?

Section 6 – FIPS 199 Categorization Level:

All GenAI data must be taken into context, and an associated categorization must be given based on the highest watermark. This includes prompt data, output data, data source, training data, etc.

(a) How critical is the confidentiality of the data in this system/application/service to achieving the business/mission objective?

Level (choose only one):	Criticality / Impact Descriptor:
☐ High ☐ Moderate ☐ Low	

(b) How critical is the integrity of the data in this system/application/service to achieving the business/mission objective?

Level (choose only one):	Criticality / Impact Descriptor:
☐ High ☐ Moderate ☐ Low	

(c) How critical is the availability of the data in this system/application/service to achieving the business/mission objective?

Level (choose only one):	Criticality / Impact Descriptor:
☐ High ☐ Moderate ☐ Low	

(d) Based on the responses to key questions, the overall protection level categorization for this system/application/service is?

FIPS 199 - Protection Level Needed (choose only one):	Comments or additional notes (if any):
☐ High ☐ Moderate ☐ Low	

Section 7 – Required Signatures for Risk Assessment Part 2:

By signing this document, the signatory is confirming that the state entity certifies the intended GenAI use case and its risk level and understands that any procurements are mandated to comply to all CDT-published security and privacy policies (SAM Sections 5100 and 5300 through 5399).	Agency Information Officer (AIO) Signature _____ Date Agency Information Security Officer (AISO) Signature _____ Date

IV. DEFINITIONS

Relevant definitions for this guidance are available in SAM 4819.2.

V. REFERENCES

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

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

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

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

4. Blueprint for an AI Bill of Rights
   https://www.whitehouse.gov/ostp/ai-bill-of-rights/

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. Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence
   https://www.whitehouse.gov/briefing-room/presidential-actions/2023/10/30/executive-order-on-the-safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence/

8. FACT SHEET: President Biden Issues Executive Order on Safe, Secure, and Trustworthy Artificial Intelligence
   https://www.whitehouse.gov/briefing-room/statements-releases/2023/10/30/fact-sheet-president-biden-issues-executive-order-on-safe-secure-and-trustworthy-artificial-intelligence//

9. 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

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

11. NIST Special Publication 800-53 Security and Privacy Controls for Information Systems and Organizations: https://csrc.nist.gov/Projects/risk-management/

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

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

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

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

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

VI. 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