PUERTO RICO INNOVATION & TECHNLOGY SERVICE

PRITS

GOBIERNO DE PUERTO RICO

PRITS-POL-0012

Política para el uso de inteligencia artificial en el Gobierno de Puerto Rico

Policy for the Use of Artificial Intelligence in the Government of Puerto Rico

- Aplicación: Todas
  Application: All

- Sistema: Todos
  System: All

Tabla de Contenido / *Table of Contents*

1. Descripción

Esta Política establece un marco estratégico y operativo para la adopción, implementación y supervisión responsable de las tecnologías de Inteligencia Artificial (IA) dentro del Gobierno de Puerto Rico. Esta proporciona los principios para asegurar que la integración de los sistemas de IA en las entidades gubernamentales mantenga los más altos estándares de transparencia, rendición de cuentas, ética y confianza pública.

En consonancia con las mejores prácticas reconocidas a nivel internacional, entre ellas el Marco de Gestión de Riesgos de Inteligencia Artificial (AI RMF 1.0) y el Marco de Ciberseguridad (CSF 2.0) del Instituto Nacional de Estándares y Tecnología (NIST), esta Política promueve el desarrollo y la utilización de sistemas de IA seguros, justos y confiables. También fomenta la innovación en cumplimiento con las leyes y regulaciones aplicables, considerando los avances tecnológicos y las implicaciones éticas emergentes. A través de este marco, el Gobierno de Puerto Rico busca fomentar una cultura de uso responsable de la IA que mejore la calidad, eficiencia y equidad de los servicios públicos.

2. Base legal

Esta Política se emite conforme a la Ley Núm. 75-2019, “Ley del Puerto Rico Innovation and Technology Service” (PRITS), la Ley Núm. 40-2024, “Ley de Ciberseguridad del Gobierno de Puerto Rico”, la Ley Núm. 151-2004, según enmendada, “Ley de Gobierno Electrónico”, y la Ley Núm. 122-2019, “Ley de Datos Abiertos del Gobierno de Puerto Rico”.

3. Propósito

Mediante la integración de principios de gestión de riesgos, protección de la privacidad y uso ético, esta Política tiene como objetivo establecer un marco de gobernanza unificado y aplicable para el uso responsable, seguro y eficaz de la inteligencia artificial en las entidades gubernamentales de Puerto Rico. Su propósito es salvaguardar los derechos individuales y la integridad de los datos públicos, al tiempo que promueve el desarrollo y

la implementación de sistemas de IA que sean transparentes, rindan cuentas y estén alineados con las mejores prácticas reconocidas internacionalmente.

4. Alcance

Las disposiciones de esta Política se aplican a la Rama Ejecutiva del Gobierno de Puerto Rico, incluyendo cualquier departamento, junta, dependencia, comisión, negociado, oficina, agencia, administración u organismo, subdivisión política, corporaciones públicas y municipios[1]. También aplica a cualquier persona natural o jurídica que haga negocios o tenga contratos con el Gobierno de Puerto Rico, incluyendo, pero sin limitarse a, personas naturales o jurídicas que realicen funciones y servicios públicos, pero solo con respecto a las funciones y servicios públicos realizados; a cualquier ejercicio de la administración pública o privada en el que se hayan dedicado o invertido fondos o recursos públicos, ya sea directa o indirectamente, o sobre el que se haya ejercido la autoridad de cualquier servidor público, respecto a los datos generados como resultado de tales actividades.

Específicamente, esta Política rige:

- Todos los sistemas y servicios de IA utilizados, desarrollados, adquiridos o mantenidos por o en nombre de organismos gubernamentales. Esto incluye, pero no se limita a, modelos de aprendizaje automático, algoritmos, automatización robótica de procesos con componentes de IA, servicios de procesamiento de lenguaje natural (PLN), herramientas de IA generativa, sistemas de apoyo a la toma de decisiones que utilizan IA y cualquier otro sistema que incorpore funcionalidades de inteligencia artificial o aprendizaje automático.

- Todas las etapas del uso de la IA, desde la planificación inicial, la propuesta, el desarrollo o la adquisición, hasta las pruebas, la implementación, el mantenimiento y la eventual desactivación. La Política también se aplica a modificaciones significativas, mejoras o reutilización de sistemas de IA que ya están en uso.

[1] Véase, artículo 2 de la Ley 40-2024.

- Todo el personal y los contratistas de la Agencia involucrados en el diseño, implementación, gestión, supervisión o uso de sistemas de IA dentro de las operaciones gubernamentales. Esto incluye empleados gubernamentales, contratistas o consultores externos y proveedores de terceros o de servicios. Todas estas personas y entidades están obligadas a cumplir con las disposiciones de esta Política. Además, cualquier acuerdo contractual que involucre servicios de IA debe incluir términos que reflejen y hagan cumplir los requisitos de esta Política.

5. Abreviaciones, acrónimos, definiciones y significados

Abreviación / Acrónimo	**Significado**
AI RMF	Marco de Gestión de Riesgo de Inteligencia Artificial *(Artificial Intelligence Risk Management Framework)*
API	Interfaz de programación de aplicaciones (*Application Programming Interface*)
BC	Continuidad del negocio (*Business Continuity*)
CISO	Principal Oficial de Seguridad Cibernética *(Chief Information Security Officer)*
CSF	Marco de Ciberseguridad (*Cybersecurity Framework*)
DEIA	Diversidad, equidad, inclusión y accesibilidad
DR	Recuperación ante desastres (*Disaster Recovery*)
EIA	Evaluación de impacto algorítmico
EIP	Evaluación de impacto a la privacidad
FedRamp	Programa Federal de Gestión de Riesgos y Autorizaciones (*Federal Risk and Authorization Management Program*)
GRE	Gestión de riesgos empresariales
HIPAA	Ley de Portabilidad y Responsabilidad del Seguro Médico (*Health Insurance Portability and Accountability Act*)
IA	Inteligencia artificial
IEC	Comisión Electrotécnica Internacional (*International Electrotechnical Commission*)
IIP	Información de identificación personal
IPS	Información protegida de salud

Abreviación / Acrónimo	**Significado**
ISO	Organización Internacional de Normalización (*International Organization for Standardization*)
MFA	Autenticación multifactor (*Multi-factor authenthication*)
MOU	Memorandos de entendimiento (*Memorandum of Understanding*)
NIST	Instituto Nacional de Estándares y Tecnología *(National Institute of Standards and Technology)*
OPI	Oficial Principal de Informática
PEII	Principal Ejecutivo de Innovación e Información
PI	Propiedad intelectual
PLN	Procesamiento de lenguaje natural
PoLP	Principio de Privilegios Mínimos (*Principle of Least Privilege*)
PRITS	Puerto Rico Innovation and Technology Service
RBAC	Control de acceso basado en roles (*Role-Based Access Control*)
RCA	Análisis de la causa raíz (*root cause analysis*)
RFP	Solicitud de propuestas (*Request for proposals*)
SDLC	Ciclo de vida de desarrollo de *software* (*software development life cycle*)
SIEM	Gestión de información y eventos de seguridad (*Security Incident and Event Management*)
SLA	Acuerdo de nivel de servicio (*Service Level Agreement*)
TI	Tecnología de la información

Término	**Definición**
Acceso basado en roles	Conjunto de autorizaciones concedidas a un usuario al asumir un rol, ya sea de forma explícita o implícita. Los permisos de un rol pueden heredarse a través de una jerarquía de roles y, por lo general, reflejan los permisos necesarios para realizar funciones específicas dentro de una organización. El mismo rol puede asignarse a una o más personas.
Agencia u organismo gubernamental	Incluye todas las entidades y organismos que componen la Rama Ejecutiva del Gobierno de Puerto Rico y sus municipios. Esto incluye, pero no se limita a cualquier departamento, junta, dependencia, comisión, negociado, oficina, agencia, administración, organismo, subdivisión política y corporaciones públicas. También comprende el conjunto de funciones, cargos y

Término	Definición
	puestos que constituyen toda la jurisdicción de una autoridad nominadora de estas agencias, independientemente de cómo se denomine.
Apetito de riesgo	Nivel general y tipos de riesgo que una organización está dispuesta a aceptar para cumplir su misión, visión o creación de valor.
Arquitectura de confianza cero	Un modelo de seguridad, un conjunto de principios de diseño de sistemas y una estrategia coordinada de gestión de ciberseguridad y sistemas basados en el reconocimiento de la existencia de amenazas tanto dentro como fuera de la red. Este modelo elimina la confianza implícita en cualquier elemento, componente, nodo o servicio, y en su lugar requiere una verificación continua y en tiempo real del entorno operativo para determinar el acceso y otras respuestas del sistema.
Ataque adversario	Actividad maliciosa en la que un atacante intenta corromper o manipular un sistema de inteligencia artificial o aprendizaje automático. Esto se logra al introducir datos engañosos o maliciosos durante las fases de entrenamiento o de operación, con el fin de forzar al sistema a tomar decisiones incorrectas o a comportarse de manera no deseada.
Inventario de IA	Una lista actualizada de todos los sistemas de IA en uso o en desarrollo dentro de una Agencia, incluyendo detalles clave como su propósito, los datos de entrada, el proveedor o desarrollador, el estado de implementación y la clasificación de riesgo designada.
Junta / Comité de Gobernanza de la IA	Un cuerpo establecido dentro de una Agencia (o a nivel de Gobierno) para supervisar la estrategia, la gestión de riesgos y el cumplimiento de la IA. Puede estar compuesto por los líderes de tecnología de la Agencia (por ejemplo, el OPI), los oficiales de seguridad, los oficiales de datos, los asesores legales y las partes interesadas del programa.
Caso de uso de la IA	Una aplicación o función específica para la cual se emplea un sistema de IA (por ejemplo: un *chatbot* para servicios al ciudadano, asistencia de IA para la toma de decisiones en la determinación de beneficios, análisis predictivo para el mantenimiento, etc.).
Sistema de Inteligencia Artificial	Cualquier *software*, plataforma, algoritmo o sistema que realiza tareas que generalmente requieren inteligencia humana, procesando datos para tomar o informar decisiones, recomendaciones o predicciones. Esto incluye modelos de aprendizaje automático, sistemas expertos, redes neuronales, sistemas de generación y comprensión de lenguaje natural,

Término	Definición
	sistemas de visión por computadora y otras tecnologías de IA emergentes.
Ciclo de vida (de la IA)	Todas las fases en la existencia de un sistema de IA, incluyendo el diseño, la recopilación de datos, el desarrollo, las pruebas, la implementación, la operación, el monitoreo, el mantenimiento y la desactivación.
Clasificación de datos	La categorización de los datos según su sensibilidad y criticidad, tal como se define en la política PRITS-POL-0001 (Política de Clasificación de Datos del Gobierno de Puerto Rico). Por ejemplo, los datos pueden clasificarse como públicos, internos, confidenciales o restringidos (o niveles similares definidos por PRITS), con sus correspondientes requisitos de manejo.
Clasificación de Riesgo (de Sistemas de IA)[2]	Designación del nivel de riesgo general de un sistema de IA (bajo, moderado o alto), según el impacto potencial que su uso podría tener en individuos, comunidades u operaciones gubernamentales. En general, un sistema de IA de alto riesgo es aquel que influye en decisiones con efectos significativos en los derechos civiles, la seguridad, el acceso a servicios o recursos críticos, o el bienestar de individuos o comunidades. El bajo riesgo se refiere a las aplicaciones con un impacto o sensibilidad mínimos, mientras que el riesgo moderado se encuentra entre ambos.
Confidencialidad	Preservar las restricciones de acceso y divulgación, incluyendo los medios para proteger la privacidad e información confidencial.
Datos	Cualquier secuencia de uno o más símbolos a los que se les da significado mediante actos específicos de interpretación.
Datos personales	Cualquier información relacionada con un individuo identificado o identificable (persona natural). Esto incluye datos que pueden identificar directa o indirectamente a una persona, de acuerdo con las definiciones de las leyes y regulaciones de privacidad.
Datos sensibles	Cualquier información que esté clasificada como confidencial, restringida o de cualquier otro tipo de sensibilidad según la Política de Clasificación de Datos del Gobierno de Puerto Rico (PRITS-POL-0001), por ejemplo, información de identificación personal (PII), información protegida de salud (PHI), registros financieros, detalles de infraestructura crítica o cualquier dato que, si se accede

[2] Los criterios para cada nivel se proporcionan en el Apéndice I: Matriz de Clasificación de Riesgos de la IA.

Término	Definición
	o se utiliza de forma incorrecta, podría dañar a las personas o al interés público.
Envenenamiento de datos	Un ataque de envenenamiento en el que un adversario controla parte de los datos de entrenamiento del sistema de IA.
Evaluación de impacto a la privacidad (EIP)	Análisis integral y una documentación formal sobre cómo se recopila, utiliza, procesa, almacena, mantiene, divulga, comparte y elimina información personal identificable dentro de un sistema de información. El propósito de una EIP es: (i) asegurar que el manejo de dicha información cumpla con los requisitos legales, reglamentarios y de política aplicables en materia de privacidad; (ii) determinar los riesgos y los posibles efectos asociados a dichas actividades; y (iii) examinar y evaluar las protecciones, salvaguardas y procesos alternativos que permitan mitigar los riesgos a la privacidad. Las EIP se realizan a lo largo del ciclo de vida de desarrollo de un sistema o programa, con el fin de identificar, evaluar y reducir preocupaciones potenciales de privacidad.
Evaluación de impacto algorítmico	Proceso sistemático para identificar, analizar y mitigar los riesgos y los impactos potenciales que un sistema de inteligencia artificial puede tener en los individuos, la sociedad y los derechos fundamentales. Este análisis examina el diseño, desarrollo y despliegue del algoritmo para determinar su nivel de riesgo y las salvaguardas necesarias. Su objetivo es garantizar que el sistema de IA sea transparente, equitativo y responsable, especialmente cuando se utiliza en decisiones críticas con consecuencias legales, financieras o civiles.
IA Confiable	Sistemas de IA que son demostrablemente válidos y fiables, seguros y resilientes, responsables y transparentes, explicables e interpretables, con privacidad mejorada y justos (con sesgos dañinos gestionados). Estas características se derivan del Marco de Gestión de Riesgos de la IA de NIST y representan las cualidades que los sistemas de IA deben tener para mantener la confianza del público.
Impacto	Magnitud del daño que se puede esperar como resultado de las consecuencias de la divulgación, modificación o destrucción no autorizadas de la información, pérdida de esta, o por la indisponibilidad del sistema de información.
Información de identificación personal (IIP)	Es cualquier representación de información que es legible sin la necesidad de una clave criptográfica especial para acceder a ella o facilita el rastreo de la identidad de un individuo, incluyendo el

Término	Definición
	nombre o la primera inicial y apellido paterno de un individuo combinado con otra información que está vinculada o que se debe vincular a un individuo específico, como: - Número de Seguro Social - Número de licencia de conducir, tarjeta electoral u otra identificación oficial - Números de cuentas bancarias o financieras de cualquier tipo, con o sin claves de acceso que puedan habérsele asignado - Nombres de usuario y claves de acceso a sistemas informáticos públicos o privados - Información médica protegida por la Ley HIPAA - Información contributiva - Evaluaciones laborales
Infraestructura crítica	Servicios, sistemas, recursos y activos esenciales, ya sean físicos o virtuales, cuya incapacidad o destrucción tendría repercusiones perjudiciales en la seguridad cibernética, la salud, la economía, la seguridad de Puerto Rico o cualquier combinación de esos asuntos.
Marco de Ciberseguridad de NIST (CSF) 2.0	Marco de NIST actualizado en 2024 para la gestión de riesgos de ciberseguridad, organizado en seis funciones principales: gobernar, identificar, proteger, detectar, responder, recuperar. El CSF 2.0 proporciona resultados y controles para proteger los sistemas (incluidos los sistemas de IA) y los datos contra las amenazas cibernéticas, e incluye un nuevo énfasis en la gobernanza.
Marco de Gestión de Riesgos de la IA de NIST (AI RMF)	Un marco publicado por NIST (Versión 1.0, 2023) que proporciona un enfoque estructurado (con funciones de gobernar, mapear, medir y gestionar) para identificar, evaluar y mitigar los riesgos de los sistemas de IA a lo largo de sus ciclos de vida. El AI RMF enfatiza la integración de características de fiabilidad en el diseño y la implementación de la IA.
Municipio(s)	Cualquiera de los 78 municipios de Puerto Rico. Para propósitos de esta Política se utilizarán los términos municipio(s) y Agencia(s) indistintamente.
Oficial Principal de Informática	Individuo responsable de la gestión y supervisión de la infraestructura y operaciones tecnológicas de una Agencia gubernamental.
Principio de Privilegios Mínimos	Cada módulo (proceso, usuario, o programa, dependiendo del tema) solo puede acceder a la información y recursos necesarios para su propósito legítimo.

Término	**Definición**
Proveedor, proveedor externo, proveedor de servicios	Entidad, ya sea persona natural o jurídica, pública o privada que provee servicios como redes, aplicaciones, programas, infraestructura o medios de seguridad mediante el soporte continuo y habitual, así como servicios de administración activa ya sea en las instalaciones de una Agencia, en el centro de procesamiento de datos de la Agencia (*hosting*), o en el centro de procesamiento de datos de un tercero.
Riesgo	Se refiere a toda circunstancia o hecho razonablemente identificable que tenga un posible efecto adverso en la seguridad de las redes y recursos de información. Generalmente se determina en función de: (i) los impactos adversos que surgirían si se produce la circunstancia o evento; y (ii) la probabilidad de ocurrencia. Los riesgos de seguridad relacionados con el sistema de información son aquellos que surgen de la pérdida de confidencialidad, integridad o disponibilidad de información o sistemas de información y reflejan los potenciales impactos adversos a las operaciones de la organización (incluida la misión, las funciones, la imagen o la reputación), los activos de los organismos gubernamentales, los individuos, otras organizaciones y el Gobierno.
Suma de verificación (*checksum*)	Valor calculado a partir del contenido de un objeto de datos mediante una función específica, el cual se almacena o transmite junto con dicho objeto. Su propósito principal es detectar errores, alteraciones no autorizadas o manipulaciones de los datos, verificando si el *checksum* calculado coincide con el valor original.
Tecnología de la información (TI)	Para una Agencia, significa cualquier sistema o recurso interconectado o subsistema de equipo utilizado en la adquisición, almacenamiento, análisis, evaluación manipulación, manejo, movimiento, control, visualización, conmutación, intercambio, destrucción, transmisión o recepción automática de datos o información, si el equipo es utilizado por la Agencia directamente o por un tercero bajo un contrato con la Agencia que requiere el uso (i) de ese equipo; o (ii) de ese equipo en una medida significativa para la prestación de un servicio o el suministro de un producto. Incluye computadoras, equipos auxiliares (incluidos periféricos de imágenes, dispositivos de entrada, salida y almacenamiento necesarios para la seguridad y vigilancia), equipos periféricos diseñados para ser controlados por la unidad central de procesamiento de una computadora, *software*, *firmware* y procedimientos y servicios similares (incluyendo servicios de apoyo) y recursos relacionados.

6. Política

6.1 Gobernanza y rendición de cuentas de la IA

6.1.1 Estructura de Gobernanza

Cada Agencia deberá establecer un mecanismo de gobernanza interna para supervisar el uso responsable de la IA. El Jefe de la Agencia (por ejemplo, Secretario, Director o Alcalde) deberá asignar una responsabilidad clara en la supervisión de la IA. Esta responsabilidad puede ser delegada al Oficial Principal de Información (OPI), a otro alto funcionario designado, o a un Comité de Gobernanza de IA interno.

Si se establece un Comité de Gobernanza de IA, este deberá incluir representantes de la alta dirección, tecnología de la información (TI), ciberseguridad, asuntos legales y las partes programáticas relevantes. Este organismo, junto con el Jefe de la Agencia y el Oficial Principal de Información (OPI), será responsable de hacer cumplir esta Política, de asegurar que se realicen las evaluaciones de riesgos y de garantizar que se obtengan todas las aprobaciones necesarias antes de usar sistemas de IA.

6.1.2 Políticas y procedimientos

Las Agencias deberán desarrollar o actualizar sus procedimientos internos para asegurar la implementación de esta Política. Estos procedimientos deberán abordar la propuesta de nuevos casos de uso de IA, la realización de evaluaciones de riesgo e impacto, la coordinación con PRITS para su revisión y aprobación, el monitoreo continuo de los sistemas de IA y los mecanismos para reportar el cumplimiento. Todos los procedimientos deberán estar alineados con esta Política y con cualquier directriz técnica adicional emitida por PRITS. Las Agencias deben asegurarse de que los procesos para identificar, evaluar y gestionar los riesgos relacionados con la IA estén claramente definidos, se apliquen de manera consistente y sean totalmente transparentes.

6.1.3 Rendición de cuentas

La máxima responsabilidad por el uso apropiado de la IA en cada Agencia reside en el Jefe de la Agencia. La persona designada como líder de IA, ya sea el OPI o el presidente del organismo de gobernanza de IA de la Agencia, es responsable del cumplimiento y la gestión de riesgos en el día a día. Las Agencias deben mantener una documentación exhaustiva de las decisiones relacionadas con el diseño, la adquisición, la implementación, el monitoreo y el desmantelamiento de los sistemas de IA, incluyendo la identidad de la autoridad que aprueba. Las Agencias también deben asegurar que cualquier sistema de IA que influya en los servicios públicos o en los derechos individuales incorpore mecanismos para la supervisión, revisión y rendición de cuentas por parte de un ser humano.

6.1.4 Supervisión de PRITS

Las Agencias deben documentar el propósito, el uso previsto y la justificación para implementar cualquier sistema de IA. Esta documentación debe ponerse a disposición de PRITS y de otros organismos de supervisión pertinentes cuando se solicite. PRITS, actuando a través de la Oficina del Oficial Principal de Seguridad de la Información (CISO) y bajo el liderazgo de su Director Ejecutivo, supervisará el uso de la IA en todo el Gobierno. PRITS es responsable de mantener esta Política, ofrecer orientación y revisar las solicitudes de las Agencias relacionadas con la IA. PRITS también puede emitir estándares o avisos suplementarios según sea necesario para abordar los riesgos o tecnologías emergentes relacionados con la IA. Las Agencias no deben implementar sistemas de IA ni realizar modificaciones significativas en ellos sin la aprobación previa por escrito de PRITS. Para esta Política, las "modificaciones significativas" se refieren a cambios sustantivos que podrían alterar la funcionalidad, el perfil de riesgo o las obligaciones de cumplimiento del sistema. Los ejemplos incluyen la introducción de nuevos conjuntos de datos, el reentrenamiento de modelos con datos diferentes, el cambio del alcance o propósito de la toma de decisiones del sistema, la integración con nuevos sistemas externos o la aplicación de actualizaciones importantes de *software* o algoritmos. El mantenimiento de rutina, los parches de *software* menores o los ajustes de configuración que no

afecten el riesgo o la funcionalidad principal no se consideran modificaciones significativas.

PRITS puede imponer condiciones, ordenar modificaciones o denegar el uso de sistemas de IA cuando sea necesario para mitigar el riesgo o asegurar la alineación con los objetivos de política pública.

6.1.5 Cumplimiento con los principios de gobernanza de NIST

La gobernanza de la IA, tanto a nivel de Agencia como a nivel de Gobierno, deberá alinearse con la función de *gobernar* de los marcos de NIST. Esto incluye establecer roles y responsabilidades claras, políticas y estrategias de gestión de riesgos; incorporar principios relacionados con la diversidad, la equidad, la inclusión y la accesibilidad (DEIA); e involucrar a las partes interesadas de manera significativa.

Los mecanismos de gobernanza deben asegurar que las iniciativas de IA apoyen el interés público y aborden de manera proactiva, en lugar de reaccionar ante, los posibles daños y las consecuencias no deseadas.

6.2 Gestión y evaluación de riesgos de IA

6.2.1 Marco de gestión de riesgos de IA

6.2.1.1 Las Agencias deberán aplicar el Marco de Gestión de Riesgos de IA (AI RMF 1.0) de NIST para guiar el proceso de gestión de riesgos en todas las etapas de la adquisición, desarrollo, implementación y mantenimiento de los sistemas de IA. Al seguir los principios del AI RMF, las Agencias se aseguran de que las características de confiabilidad (como la validez, seguridad, etc.) se consideren en cada etapa. El AI RMF se estructura en torno a cuatro funciones principales que las Agencias deben implementar:

6.2.1.1.1 *Gobernar*

Según se describe en la Sección 6.1, las Agencias deben establecer un mecanismo de supervisión que

asegure una revisión continua y la rendición de cuentas en la gestión de riesgos de IA.

6.2.1.1.2 *Mapear*

Identificar y contextualizar el propósito previsto, las partes interesadas y los posibles impactos del sistema de IA. Las Agencias deben definir claramente el contexto, el alcance y los objetivos de cada sistema de IA, incluyendo los requisitos legales y expectativas.

6.2.1.1.3 *Medir*

Evaluar y analizar los riesgos y el rendimiento del sistema de IA en comparación con las métricas y los puntos de referencia definidos. Las Agencias deberán evaluar los sistemas de IA en cuanto a su precisión, confiabilidad, robustez, imparcialidad, privacidad, seguridad y alineación con los resultados esperados.

6.2.1.1.4 *Gestionar*

Implementar estrategias de mitigación de riesgos y tomar decisiones informadas con respecto al uso, restricción o modificación de los sistemas de IA, considerando los riesgos asociados. Las Agencias deben priorizar los escenarios de alto riesgo y establecer un monitoreo y controles continuos.

6.2.1.2 La gestión de riesgos de la IA no debe estar aislada. Las Agencias deben integrarla en sus programas generales de gestión de riesgos empresariales (GRE) y de gestión de riesgos de ciberseguridad. Las decisiones relativas a los riesgos relacionados con la IA deben alinearse con el apetito de riesgo de la Agencia y con su postura general de ciberseguridad.

6.2.2 Evaluación de riesgos empresariales

6.2.2.1 Antes de implementar cualquier sistema de IA nuevo o de realizarle modificaciones significativas a uno existente, las Agencias deben llevar a cabo una evaluación exhaustiva de riesgos e impacto de la IA. Esta evaluación debe realizarse antes de la implementación en producción y debe ser enviada a PRITS para su revisión y aprobación, según la Sección 6.1.4.

6.2.2.2 La evaluación de riesgos e impacto incluirá, como mínimo, una evaluación de las siguientes áreas:

6.2.2.2.1 *Impactos potenciales*

Las Agencias deben analizar cómo el sistema de IA puede afectar a las personas, las comunidades y las operaciones gubernamentales. Esto incluye el potencial de daños no intencionados, el uso indebido del sistema de IA o la interrupción de servicios esenciales. El uso indebido puede consistir en utilizar el sistema de IA fuera de su alcance previsto, manipularlo con fines no autorizados o implementarlo sin las salvaguardas adecuadas. Se deberá prestar especial atención al riesgo de resultados discriminatorios, la violación de los derechos civiles, la erosión de la confianza pública y los impactos en la privacidad, la salud pública y la seguridad.

6.2.2.2.2 *Sensibilidad de los datos*

Las Agencias deben revisar los tipos de datos que el sistema de IA utilizará, incluyendo los datos de entrada, entrenamiento y salida. Todos los datos deben ser clasificados de acuerdo con la Política de Clasificación de Datos del Gobierno de Puerto Rico (PRITS-POL-0001) o cualquier otra regulación emitida por PRITS, y ser gestionados según su nivel de sensibilidad, utilizando salvaguardas adecuadas

como cifrado, controles de acceso y registro de auditoría.

6.2.2.2.3 *Sesgo e imparcialidad*

Las Agencias deben evaluar si el sistema de IA podría desenlazar en resultados discriminatorios o sesgados. Cuando los sistemas de IA se utilicen para la toma de decisiones sobre individuos (por ejemplo, elegibilidad para un servicio), las Agencias deben realizar una Evaluación de Impacto Algorítmico (EIA) e implementar medidas correctivas para abordar las disparidades identificadas, asegurando la imparcialidad y la equidad.

6.2.2.2.4 *Transparencia y capacidad de explicación*

Las Agencias deben evaluar en qué medida la lógica y los resultados del sistema de IA pueden explicarse a los usuarios finales y a las personas afectadas. Para cualquier sistema de IA de alto riesgo, en particular aquellos que afecten a individuos, las Agencias deben asegurarse de que las decisiones puedan ser explicadas a solicitud.

6.2.2.2.5 *Riesgos de seguridad*

Las Agencias deben evaluar los riesgos de ciberseguridad asociados con el sistema de IA a lo largo de su ciclo de vida. Esto incluye evaluar la vulnerabilidad del sistema a ataques adversarios (como el envenenamiento de datos o la manipulación del modelo), así como el riesgo de acceso no autorizado al propio sistema de IA o una violación de los datos sensibles que procesa.

Las Agencias deben asegurar la alineación con las funciones del NIST CSF 2.0 (Identificar, Proteger, Detectar, Responder y Recuperar) e implementar

salvaguardas para proteger tanto el sistema como sus datos contra amenazas internas y externas.

6.2.2.2.6 *Confiabilidad y seguridad*
Las Agencias deben evaluar si el sistema de IA funciona según lo previsto en condiciones normales y de estrés. Se deben documentar los modos de fallo y establecer procedimientos de respaldo. Ningún sistema de IA será implementado si su fallo pudiera poner en peligro la vida, la seguridad o la propiedad (especialmente en el caso de sistemas de IA que tengan efectos en el mundo físico, como en la gestión de tráfico, la atención médica, etc.).

6.2.2.2.7 *Cumplimiento*
Las Agencias deben verificar que el sistema de IA cumple con todas las leyes y regulaciones aplicables. Esto incluye, pero no se limita a, las leyes locales y federales, así como las disposiciones aplicables en materia de privacidad, derechos civiles, propiedad intelectual y otras áreas relevantes.

6.2.2.3 Todos los hallazgos deben documentarse en un Informe Formal de Evaluación de Riesgos de IA, que servirá como registro oficial de la debida diligencia de la Agencia. El informe debe incluir:
- 6.2.2.3.1 Un resumen del análisis de riesgos en los siete dominios enumerados anteriormente.
- 6.2.2.3.2 Los riesgos identificados y las estrategias de mitigación recomendadas.
- 6.2.2.3.3 Una justificación para proceder, modificar o detener la implementación del sistema de IA.

6.2.2.4 Este informe debe ser enviado a PRITS como parte de la solicitud de aprobación antes de la implementación en producción (según la sección 6.1.4). PRITS puede emitir plantillas o listas de verificación estandarizadas para dichas

evaluaciones. Las Agencias deben consultar cualquier guía suplementaria de este tipo.

6.2.3 Clasificación de riesgos

6.2.3.1 Como parte de la evaluación de riesgos, las Agencias deberán clasificar cada sistema de IA como de riesgo bajo, moderado o alto, basándose en la naturaleza y gravedad de los impactos potenciales (consulte el Apéndice I para ver los criterios de clasificación). Esta clasificación determina el rigor de los controles y la supervisión requeridos:

6.2.3.1.1 Riesgo bajo

Sistemas de IA que presentan un riesgo mínimo para las personas, los servicios o las operaciones. Estos sistemas utilizan únicamente datos no sensibles y no tienen un impacto significativo en las decisiones o servicios (por ejemplo, una herramienta de IA utilizada para resumir u organizar información disponible públicamente). La documentación, los controles y el monitoreo básicos son suficientes.

6.2.3.1.2 Riesgo moderado

Sistemas de IA que manejan datos de sensibilidad moderada o que influyen en decisiones de cierta importancia, pero donde los errores o el uso indebido tendrían un alcance de daño limitado. Esto incluye situaciones en las que la IA podría tener efectos notorios pero controlados si es incorrecta o mal utilizada, lo que significa que el impacto podría causar inconvenientes, retrasos o interrupciones temporales, pero sin causar daños duraderos a los derechos de las personas, a su seguridad o a su acceso a servicios esenciales. Por ejemplo, un sistema de IA que ayuda a priorizar las solicitudes de mantenimiento de rutina para infraestructura pública

no crítica, en el que un error podría retrasar ciertas reparaciones, pero no comprometería la seguridad pública ni violaría los derechos civiles. Estos sistemas requieren pruebas estándar, supervisión humana, y documentación y controles más completos para asegurar un uso responsable.

6.2.3.1.3 Riesgo alto

Se consideran de alto riesgo aquellos sistemas de IA que puedan tener un impacto significativo en los derechos individuales, la seguridad pública, el acceso a servicios esenciales o la asignación justa de los recursos gubernamentales. Esta categoría incluye sistemas que procesan datos altamente sensitivos, como información personal identificable (PII) o información de salud protegida (PHI), particularmente cuando se utilizan a gran escala. Ejemplos incluyen aplicaciones de IA para determinación de elegibilidad, triaje en servicios de salud, despacho de seguridad pública, priorización de recursos, así como sistemas de visión computarizada o de reconocimiento biométrico utilizados para identificación, autenticación o vigilancia. Los sistemas de alto riesgo requieren controles rigurosos, incluyendo evaluaciones de equidad y sesgo, auditorías regulares y monitoreo continuo para garantizar el cumplimiento. PRITS podrá imponer condiciones o restricciones adicionales según sea necesario.

6.2.3.2 Si un sistema de IA presenta un riesgo inaceptable, lo que significa que es inherentemente incompatible con los estándares legales, éticos o de políticas (como los sistemas diseñados para vigilancia ilegal, puntuación social o casos de uso que violen los derechos fundamentales), su implementación está estrictamente prohibida. Dichos casos de

uso se clasificarán como prohibidos y las Agencias no deberán realizar actividades de adquisición o desarrollo relacionadas con ellos.

6.2.4 Monitoreo continuo y reevaluación periódica

6.2.4.1 La gestión de riesgos para los sistemas de IA debe ser continua y adaptable. Las Agencias deberán implementar mecanismos para monitorear los sistemas de IA a lo largo de su vida operativa a fin de identificar riesgos nuevos o cambiantes. Las actividades de monitoreo deben incluir métricas de rendimiento, tasas de error, informes de incidentes, retroalimentación de los usuarios y cambios en los datos o en el contexto de uso.

6.2.4.2 Las Agencias deberán revisar la clasificación de riesgos y la evaluación de riesgos al menos una vez al año para cada sistema de IA. La reevaluación también es obligatoria antes de cualquier modificación significativa al sistema, tales como:

- La integración de nuevas fuentes de datos,
- la implementación en nuevos contextos operativos,
- actualizaciones importantes del algoritmo o modelo, o
- la ocurrencia de un incidente que involucre al sistema de IA.

6.2.4.3 Si el monitoreo revela que un sistema de IA ya no cumple con los criterios de confiabilidad, como la precisión, la imparcialidad, la confiabilidad o el cumplimiento, se deben tomar medidas correctivas sin demora. Estas medidas pueden incluir el reentrenamiento del modelo, la mejora de la calidad de los datos, el fortalecimiento de las salvaguardas o, en casos graves, la suspensión del uso del sistema hasta que se resuelvan los problemas.

6.2.4.4 Las Agencias deben informar los cambios significativos en el riesgo de IA o cualquier incidente relacionado con la IA a PRITS y a la Oficina de Evaluación de Incidentes Cibernéticos, según los plazos de notificación establecidos por la Política de Respuesta a Incidentes de Seguridad de la Información (PRITS-POL-0002). Esto incluye, pero no se limita a, las

violaciones que involucren sistemas de IA, el acceso o uso no autorizado, o situaciones en las que los resultados generados por la IA hayan causado daños o errores graves.

6.2.4.5 Los incidentes de ciberseguridad que involucren sistemas de IA deben ser reportados puntualmente, no más tarde de 48 horas, según los requisitos estatutarios.

6.3 Gestión de datos y privacidad

6.3.1 Clasificación y protección de datos

6.3.1.1 Todos los datos utilizados por o resultantes de sistemas de IA deben ser manejados siguiendo la Política de Clasificación de Datos para el Gobierno de Puerto Rico (PRITS-POL-0001). Las Agencias deberán:

6.3.1.1.1 Identificar y clasificar todos los conjuntos de datos involucrados en un sistema de IA, incluyendo los datos de entrenamiento, datos de entrada, datos de salida y cualquier dato derivado o almacenado.

6.3.1.1.2 Prohibir la transferencia o la entrada de datos gubernamentales sensibles a servicios de IA externos o públicos a menos que estén explícitamente autorizados y protegidos por salvaguardas adecuadas (como cifrado, obligaciones contractuales y acuerdos aprobados de manejo de datos). Por ejemplo, los datos confidenciales o restringidos no deben ser cargados a un sistema de IA de terceros sin medidas de seguridad aprobadas por PRITS y una revisión legal.

6.3.1.1.3 Aplicar controles de seguridad apropiados para el nivel de clasificación de los datos. Por ejemplo, los datos restringidos utilizados por un sistema de IA deben ser cifrados tanto en reposo como en tránsito, almacenados en entornos seguros aprobados en la nube o en las instalaciones (*on-premises*), y el acceso debe limitarse al personal autorizado,

siguiendo los estándares y regulaciones de seguridad aplicables.

6.3.1.1.4 Si se utilizan sistemas de IA para generar o procesar información pública, asegurar de que cualquier divulgación de datos públicos resultante cumpla con las regulaciones de datos abiertos y no exponga de forma inadvertida información no pública.

6.3.1.2 Cualquier sistema de IA que ingiera o genere imágenes, video o plantillas biométricas (por ejemplo, rostro, iris, marcha) deberá:

6.3.1.2.1 Tratar las imágenes crudas y derivadas como datos sensibles.

6.3.1.2.2 Almacenar las imágenes solo en repositorios cifrados aprobados por PRITS.

6.3.1.2.3 Aplicar el seguimiento de la procedencia del conjunto de datos y el registro a prueba de manipulación (*tamper-evident logging*).

6.3.1.2.4 Eliminar las imágenes crudas no esenciales después del entrenamiento o la inferencia del modelo, a menos que su retención sea un requisito legal.

6.3.2 Privacidad y datos personales

6.3.2.1 Las Agencias deberán adoptar un enfoque de privacidad desde el diseño para los sistemas de IA que procesen datos personales. Como mínimo:

6.3.2.1.1 Realizar una Evaluación de Impacto a la Privacidad (EIP) para cualquier sistema de IA que maneje datos personales de ciudadanos o empleados. La EIP debe documentar cómo se recopila, utiliza, almacena y comparte la información personal, e identificar las medidas para mitigar los riesgos de privacidad asociados.

6.3.2.1.2 Asegurar el cumplimiento de las leyes de privacidad aplicables, incluyendo los requisitos de protección de datos de Puerto Rico, las leyes federales (por

ejemplo, HIPAA para datos de salud) y cualquier regulación de privacidad específica del sector.

6.3.2.1.3 Implementar técnicas que mejoren la privacidad siempre que sea factible, como la minimización de datos (limitar la recopilación de datos a lo estrictamente necesario para el propósito de la IA), la anonimización o desidentificación, y la agregación de resultados para evitar la reidentificación de individuos.

6.3.2.1.4 Proporcionar aviso a los usuarios o al público, según corresponda, cuando los sistemas de IA procesen datos personales. Dicho aviso debe informar a las personas sobre el rol de la IA, los datos que se procesan y, cuando sea aplicable, ofrecer mecanismos para preguntar o impugnar las decisiones impulsadas por la IA.

6.3.2.1.5 Establecer y hacer cumplir políticas de retención y eliminación de datos para los datos relacionados con la IA. Los datos personales no deben ser retenidos por más tiempo del necesario para el propósito autorizado. Los cronogramas de retención para los datos de entrenamiento de IA, datos de entrada/salida y conjuntos de datos derivados que contengan información personal deben definirse formalmente, aprobarse y estar alineados con las normas de gestión de registros de la Agencia y con cualquier requisito regulatorio aplicable.

6.3.3 Calidad e integridad de los datos

6.3.3.1 Dado que la confiabilidad y la credibilidad de los resultados de un sistema de IA dependen de la calidad de los datos subyacentes, las Agencias deben asegurarse de que:

6.3.3.1.1 Los conjuntos de datos sean precisos, estén actualizados y, en la medida de lo posible, libres de errores conocidos, lagunas o sesgos. Antes de

implementar un sistema de IA, todos los conjuntos de datos de entrenamiento y configuración deberán ser revisados y auditados para verificar su calidad, integridad y representatividad.

6.3.3.1.2 Existen procedimientos para monitorear los cambios en los datos subyacentes y para actualizar o reentrenar los modelos de IA según sea necesario. Esto es esencial para prevenir la deriva del modelo o la degradación de la precisión predictiva con el tiempo.

6.3.3.1.3 Los datos sensibles utilizados para el entrenamiento, el ajuste fino o las operaciones de la IA estén protegidos de modificaciones no autorizadas (integridad) y de acceso no autorizado (confidencialidad). Se deben aplicar medidas del programa de ciberseguridad (monitoreo del acceso a los datos, uso de sumas de verificación o firmas digitales para conjuntos de datos críticos, etc.).

6.3.3.1.4 Cuando los sistemas de IA utilizan datos de proveedores externos o de conjuntos de datos de terceros, dichas fuentes deben ser examinadas y aprobadas formalmente. Las Agencias deben confirmar que los datos fueron recopilados de manera legal y ética, y que están autorizados para el uso previsto de la IA. Se deberá dar preferencia a los conjuntos de datos que tengan una procedencia documentada, métodos de recopilación bien definidos y limitaciones identificadas.

6.3.4 Confidencialidad de los resultados de la IA

6.3.4.1 Los resultados generados por la IA deben someterse a una evaluación para determinar si contienen, se derivan de, o podrían usarse para inferir información sensible. Los ejemplos incluyen informes, análisis, modelos predictivos, resúmenes u

otro contenido generado que podría revelar datos clasificados, de identificación personal o protegidos de alguna otra manera.

6.3.4.2 Cuando se determine que los resultados son sensibles, las Agencias deberán aplicar el nivel de clasificación y los procedimientos de manejo apropiados, de acuerdo con las leyes, regulaciones y políticas de seguridad internas aplicables. Esto incluye clasificar los resultados de forma adecuada, almacenarlos en entornos seguros aprobados y restringir su difusión únicamente a los destinatarios autorizados.

6.3.4.3 Las Agencias deben implementar salvaguardas para prevenir la divulgación no intencionada de información sensible en los resultados de la IA, lo que incluye, entre otras cosas:

6.3.4.3.1 Implementar el filtrado de resultados, la censura o el enmascaramiento de detalles sensibles antes de su difusión.

6.3.4.3.2 Revisar los resultados de la IA generados a partir de conjuntos de datos sensibles para asegurarse de que no revelen indirectamente información confidencial mediante inferencia o correlación.

6.3.4.3.3 Aplicar herramientas de detección automatizada, siempre que sea factible, para marcar y revisar el contenido generado por IA que sea potencialmente sensible antes de su publicación.

6.4 Seguridad de los sistemas de IA (integración de la ciberseguridad)

6.4.1 Diseño y desarrollo seguro

6.4.1.1 Todos los sistemas de IA, ya sean desarrollados internamente, adquiridos de proveedores o integrados a partir de soluciones de código abierto (*open-source*), deben seguir los principios del ciclo de vida de desarrollo de *software* seguro (SDLC) adaptados para la IA. Esto incluye:

6.4.1.1.1 *Seguridad desde el diseño*

Definir y documentar los requisitos de seguridad durante la fase de diseño, incluyendo el control de acceso, la autenticación, el cifrado, el registro de auditoría y la resiliencia contra amenazas específicas de la IA.

6.4.1.1.2 *Pruebas previas a la implementación y periódicas*
Realizar evaluaciones de vulnerabilidades, pruebas de penetración y simulaciones de amenazas específicas de la IA (por ejemplo, manipulación de entradas adversarias, envenenamiento de datos, inyección de *prompts*) antes de la publicación y en intervalos definidos.

6.4.1.1.3 *Código seguro y estándares de modelos*
Asegurar que el desarrollo del *software* y de los modelos de IA siga las prácticas reconocidas de codificación segura. Los componentes de código abierto o los modelos preentrenados deben ser verificados en cuanto a su integridad, procedencia y reputación antes de su uso, y se deben actualizar regularmente para abordar las vulnerabilidades.

6.4.1.1.4 *"Hardening" del ambiente*
Proteger los ambientes de entrenamiento, validación y producción para prevenir el acceso no autorizado o la manipulación. Para los sistemas de IA sensibles, implementar aislamiento de red, controles estrictos de identidad y acceso y monitoreo continuo de la actividad del sistema para garantizar la seguridad.

6.4.2 Control de acceso y autenticación

6.4.2.1 El acceso a los recursos del sistema de IA, incluyendo las interfaces de administración, el código fuente del modelo, los

repositorios de datos de entrenamiento y los conjuntos de datos operativos debe restringirse al personal autorizado.

6.4.2.2 La autenticación multifactor (MFA) deberá implementarse en todas las cuentas de usuarios.

6.4.2.3 Se deben aplicar el Principio de Privilegios Mínimos (PoLP) y los Controles de Acceso Basados en Roles (RBAC) para garantizar que los usuarios y los procesos tengan solo los privilegios mínimos necesarios.

6.4.2.4 En línea con el modelo de Arquitectura de Confianza Cero (*Zero Trust*), cada intento de acceso debe ser tratado como potencialmente no confiable hasta que sea verificado a través de una validación continua de identidad, la autenticación contextual y el monitoreo de la actividad.

6.4.2.5 Los sistemas de IA no recibirán acceso privilegiado o administrativo a la infraestructura de TI del Gobierno (incluyendo, entre otros, configuraciones de sistemas, creación/modificación/eliminación de cuentas de usuario o permisos elevados de API). Dichas funciones deben permanecer bajo control humano directo y estar sujetas a las políticas establecidas de ciberseguridad y gestión de acceso.

6.4.2.6 Los registros de acceso de los sistemas de IA deben mantenerse, revisarse periódicamente y protegerse contra la manipulación para respaldar la auditoría, la detección de incidentes y las investigaciones forenses.

6.4.3 Monitoreo y respuesta a incidentes

6.4.3.1 Todos los sistemas de IA y sus ambientes de alojamiento deberán integrarse en los procesos de monitoreo de ciberseguridad y respuesta a incidentes de la Agencia. Esto incluye, como mínimo:

6.4.3.1.1 El registro de las actividades del sistema de IA, incluyendo los eventos del sistema, los resultados de las decisiones (cuando sea técnicamente factible) y el acceso a los datos. Los registros deberán mantenerse siguiendo las políticas de retención,

protegerse contra la manipulación y monitorearse continuamente en busca de indicadores de compromiso o actividad anormal.

6.4.3.1.2 Implementación de herramientas de gestión de información y eventos de seguridad (SIEM) o equivalentes para detectar actividades sospechosas dirigidas a sistemas de IA, como intentos de acceso no autorizado, explotación de *endpoints* de API o patrones de entrada atípicos que sugieran un ataque adversario.

6.4.3.1.3 En caso de que se confirme o se sospeche de un incidente de seguridad que involucre un sistema de IA, como el compromiso de los conjuntos de datos de entrenamiento, la alteración no autorizada de los parámetros del modelo o la manipulación externa de los resultados, las Agencias deben responder de acuerdo con su plan de respuesta a incidentes. Todos los incidentes deben ser reportados a la Oficina de Evaluación de Incidentes Cibernéticos dentro de los plazos establecidos por las regulaciones aplicables, incluyendo la Política de Respuesta a Incidentes de Seguridad de la Información (PRITS-POL-0002).

6.4.3.1.4 Los planes de Continuidad del Negocio y Recuperación ante Desastres (BC/DR) deberán incorporar explícitamente escenarios que involucren fallas, mal funcionamiento o falta de disponibilidad de los sistemas de IA. Las Agencias deberán establecer y probar periódicamente los procedimientos de respaldo que permitan la continuación de los servicios de forma manual o por medios que no sean de IA hasta que el sistema afectado sea restaurado a sus operaciones normales.

6.4.4 Seguridad de terceros y de la cadena de suministro

6.4.4.1 Reconociendo que muchas soluciones de IA dependen de componentes de terceros, incluyendo servicios de IA basados en la nube, modelos preentrenados y bibliotecas de *software*, las Agencias deberán implementar medidas sólidas de gestión de riesgos de la cadena de suministro para prevenir compromisos derivados de dependencias externas. Esto incluye:

6.4.4.1.1 Utilizar solo servicios, componentes o plataformas de IA de terceros aprobados y confiables, que satisfagan los estándares de seguridad y cumplimiento del Gobierno. Cuando el cumplimiento de un proveedor sea incierto, las Agencias deberán consultar a PRITS para su verificación antes de proceder con la adquisición o integración.

6.4.4.1.2 Incorporar requisitos explícitos de seguridad y cumplimiento en todos los contratos de adquisición y acuerdos de nivel de servicio (SLA), según se describe en la Sección 6.6 y el Apéndice II. Para los sistemas de IA basados en la nube, las Agencias deberán exigir que los proveedores tengan las certificaciones adecuadas, como FedRAMP o su equivalente, y que se comprometan formalmente a proteger los datos gubernamentales en tránsito, en reposo y en uso.

6.4.4.1.3 Mantener un conocimiento continuo de las advertencias de seguridad, las actualizaciones de versiones y los parches emitidos por proveedores externos. Las Agencias deben evaluar y aplicar dichas actualizaciones puntualmente para mitigar las vulnerabilidades.

6.4.4.1.4 Cuando se utilicen modelos de IA, conjuntos de datos o bibliotecas de código abierto, las Agencias deberán realizar evaluaciones formales de riesgos que consideren factores como la confiabilidad de la comunidad de desarrollo, la frecuencia de las actualizaciones y el potencial de manipulación

maliciosa (por ejemplo, envenenamiento de datos en repositorios). Se deberán implementar controles técnicos para validar la autenticidad e integridad de todos los componentes de código abierto antes de su implementación en entornos de producción.

6.4.4.2 Integración con NIST CSF 2.0

6.4.4.2.1 Las Agencias deberán alinear la gestión de seguridad de los sistemas de IA con el Marco de Ciberseguridad (CSF) 2.0 del NIST para asegurar un enfoque integral, estructurado y reproducible para la gobernanza de la ciberseguridad de la IA. La alineación deberá abarcar las siguientes funciones:

6.4.4.2.1.1 *Identificar*
Mantener un inventario actualizado de los activos de IA, conjuntos de datos e infraestructura relacionada. Documentar su propósito comercial, riesgos asociados y clasificaciones de criticidad para guiar la toma de decisiones informada sobre el riesgo.

6.4.4.2.1.2 *Proteger*
Implementar y mantener salvaguardas, incluyendo el control de acceso, cifrado, minimización de datos y permisos basados en roles, para asegurar la continuidad de los servicios habilitados por IA y la protección de los datos asociados.

6.4.4.2.1.3 *Detectar*
Desarrollar y mantener capacidades para identificar la ocurrencia de eventos

de ciberseguridad que puedan afectar a los sistemas de IA, incluyendo la detección de anomalías en los resultados del modelo de IA y el monitoreo automatizado de los registros de seguridad.

6.4.4.2.1.4 *Responder*

Establecer y mantener planes de respuesta para abordar los incidentes relacionados con la IA, incluyendo la contención de modelos corruptos, la revocación de credenciales comprometidas y la reversión temporal a procesos manuales o alternativos.

6.4.4.2.1.5 *Recuperar*

Implementar y probar regularmente planes de recuperación para restaurar las capacidades de IA después de una interrupción, incluyendo el reentrenamiento de modelos a partir de conjuntos de datos de respaldo verificados, la validación de la precisión del modelo y la reinstalación de servicios completos.

6.4.4.2.1.6 *Gobernar*

Asegurar que la gobernanza de la IA esté integrada en todas las funciones de ciberseguridad, con mecanismos de supervisión que evalúen y gestionen los riesgos específicos de la IA, como se describe en la Sección 6.1.

6.5 Imparcialidad, ética y transparencia

6.5.1 No discriminación

6.5.1.1 Todos los sistemas de IA implementados o utilizados por el Gobierno deben operar de una manera que defienda los principios de igualdad y no discriminación, evitando resultados que sean injustos, sesgados o discriminatorios hacia cualquier individuo o grupo.

6.5.1.2 Las Agencias deberán evaluar los sistemas de IA en busca de sesgos potenciales, especialmente para aplicaciones de alto riesgo, en particular aquellas que influyen en el empleo, beneficios, aplicación de la ley u otras decisiones con un impacto individual significativo.

6.5.1.3 Cuando se identifique un sesgo (por ejemplo, un rendimiento consistentemente deficiente para datos demográficos específicos), las Agencias deben tomar medidas correctivas como reentrenar los modelos con datos más diversos, ajustar los algoritmos para reducir la amplificación del sesgo o introducir la supervisión humana para validar y, si es necesario, anular los resultados de la IA.

6.5.1.4 Todas las decisiones impulsadas por la IA que afecten a derechos, beneficios u oportunidades deben ser revisadas para verificar el cumplimiento de la legislación de derechos civiles aplicable y de los principios de igualdad de oportunidades.

6.5.2 Transparencia para el público

6.5.2.1 Los ciudadanos tienen derecho a ser informados cuando interactúan con servicios gubernamentales que utilizan IA o cuando son objeto de una decisión asistida por IA. Las Agencias deberán proporcionar avisos claros y accesibles en los sistemas de cara al público que empleen tecnologías de IA (por ejemplo, "Este servicio es apoyado por un asistente de IA").

6.5.2.2 Si la IA contribuye a una decisión de elegibilidad o cumplimiento que afecta a una persona, la Agencia deberá, cuando sea apropiado, informar al individuo que se utilizó una herramienta

de IA y proporcionar una vía para la consulta o apelación con intervención humana.

6.5.2.3 Cuando la IA haya contribuido a la determinación de la elegibilidad, el cumplimiento o la aplicación de la ley, las Agencias deberán, cuando sea apropiado:

6.5.2.3.1 Informar a las personas afectadas sobre la participación de la IA en el proceso de decisión.

6.5.2.3.2 Ofrecer mecanismos de consulta, aclaración o apelación con acceso a revisión humana.

6.5.3 Capacidad de explicar

6.5.3.1 Para los sistemas de IA de alto riesgo, las Agencias deben priorizar la capacidad de explicar para asegurar que los operadores puedan interpretar y transmitir el razonamiento detrás de los resultados de la IA a las personas afectadas y a las entidades de supervisión. Esto podría implicar:

6.5.3.1.1 Emplear modelos inherentemente interpretables.

6.5.3.1.2 Generar informes de explicación *post-hoc* para algoritmos complejos.

6.5.3.1.3 Producir resúmenes en lenguaje sencillo cuando las decisiones tengan un impacto significativo (por ejemplo, la denegación de una solicitud de servicio).

6.5.3.2 Los sistemas de IA de "caja negra" —donde las entradas y salidas son visibles, pero el proceso de toma de decisiones no es transparente— no se deben utilizar en contextos con consecuencias legales, financieras o civiles, a menos que existan mecanismos suficientes para explicar sus decisiones. Se dará preferencia a los modelos que puedan revelar los factores de decisión, la ponderación o las reglas aplicadas.

6.5.4 Supervisión humana

6.5.4.1 Las herramientas de IA deben apoyar, no reemplazar, el juicio humano en la toma de decisiones del Gobierno. Las Agencias deberán mantener mecanismos de supervisión humana,

incluyendo la revisión humana previa a la decisión para determinaciones críticas o mecanismos de apelación posteriores a la decisión que permitan la intervención humana.

6.5.4.2 Los empleados deberán ser capacitados para evaluar de manera crítica los resultados de la IA, particularmente cuando se detecten anomalías o cuando las decisiones conlleven consecuencias legales o personales significativas.

6.5.5 Uso y propósito ético

6.5.5.1 Las implementaciones de IA deben alinearse con el interés público y las protecciones constitucionales. La IA no se usará de maneras que violen los derechos fundamentales, invadan las expectativas razonables de privacidad sin autorización legal o contravengan de otro modo los principios éticos gubernamentales.

6.5.5.2 Los casos de uso de IA que sean novedosos, sensibles o éticamente polémicos (por ejemplo, el reconocimiento facial en espacios públicos) deben someterse a una revisión de nivel superior por parte de PRITS para verificar el cumplimiento de las políticas de gobernanza tecnológica aplicables.

6.5.5.3 Cada Agencia será responsable de llevar a cabo su propia evaluación legal y ética, incluyendo la evaluación de los impactos potenciales en los derechos constitucionales y la privacidad, en coordinación con su asesor legal u otras autoridades competentes, según sea aplicable. Se recomienda encarecidamente la consulta con comités de ética o expertos cualificados para implementaciones sensibles.

6.6 Adquisiciones y servicios de IA de terceros

6.6.1 Aprobación de PRITS en las adquisiciones

6.6.1.1 Según lo dispuesto por la Ley 75-2019, todas las adquisiciones de sistemas o servicios de IA de las Agencias de la Rama Ejecutiva, ya sean *software*, plataformas de IA en la nube, servicios de datos, consultoría o modelos pre-entrenados,

deberán ser coordinadas con PRITS para asegurar el cumplimiento con las mejores prácticas en arquitectura empresarial, requisitos de seguridad y privacidad, y esta Política de IA y los marcos regulatorios relacionados.

6.6.1.2 Las entidades excluidas del alcance de la Ley 75-2019, pero sujetas a la Ley 40-2024, también deberán asegurar el cumplimiento de esta Política de IA y de los requisitos de seguridad, privacidad y gobernanza aplicables, coordinando con PRITS según sea necesario para abordar los riesgos específicos de la IA.

6.6.1.3 Ningún contrato para un producto o servicio de IA se llevará a cabo sin la confirmación documentada de PRITS de que la adquisición se alinea con las políticas aplicables. Esto incluye la obtención de una autorización escrita explícita de PRITS antes de finalizar cualquier contrato relacionado con la IA.

6.6.2 Requisitos para proveedores de IA

6.6.2.1 Los contratos con los proveedores de IA deben incluir cláusulas estrictas y aplicables para proteger los intereses del Gobierno. Estas disposiciones deberán ser consistentes con las leyes, regulaciones y políticas aplicables (para obtener orientación detallada, consulte el Apéndice II). Como mínimo, se deben abordar las siguientes áreas:

6.6.2.1.1 *Estándares se seguridad y privacidad*

Los proveedores deben cumplir con todos los requisitos de seguridad y privacidad establecidos por el Gobierno de Puerto Rico, incluyendo, entre otros, las políticas de seguridad de PRITS aplicables, los controles de seguridad contractuales específicos del proyecto y los estándares reconocidos de la industria (por ejemplo, NIST CSF, ISO/IEC 27001). Los proveedores deben:

- Implementar salvaguardas técnicas, incluyendo cifrado en tránsito y en reposo, prácticas seguras

de desarrollo de *software* y controles de acceso rigurosos.
- Asegurar que los datos personales sean recopilados, procesados, almacenados y transmitidos en estricto cumplimiento de las leyes y regulaciones de privacidad aplicables.
- Comprometerse con la estricta confidencialidad de todos los datos gubernamentales, prohibiendo cualquier divulgación no autorizada o uso secundario.

6.6.2.1.2 *Abastecimiento y propiedad de los datos*

Los contratos deberán establecer claramente que:
- Todos los datos del Gobierno, y cualquier modelo derivado, análisis o conocimiento generado a partir de ellos, son y seguirán siendo propiedad del Gobierno de Puerto Rico.
- Los proveedores tienen prohibido usar datos del Gobierno para fines ajenos al alcance del contrato sin previa autorización escrita.
- Los proveedores deben divulgar cualquier dato de terceros, conjuntos de datos o modelos preentrenados incorporados en el sistema de IA. Deben garantizar que estos han sido adquiridos legalmente y que el Gobierno tiene el derecho de usarlos.
- Al finalizar o expirar el contrato, los proveedores deben devolver o eliminar todos los datos del Gobierno y certificar dicha acción por escrito.

6.6.2.1.3 *Rendimiento y* calidad

Los contratos deben definir métricas de rendimiento objetivas y medibles para el sistema de IA, tales como:

- Umbrales de precisión, puntos de referencia de latencia y tiempo de respuesta y requisitos de disponibilidad.
- Obligaciones para que el proveedor solucione con prontitud los defectos, vulnerabilidades, sesgos algorítmicos o deficiencias funcionales identificadas durante el período del contrato.
- Compromisos para mantener el rendimiento del sistema durante todo el ciclo de vida de la implementación de la IA.

6.6.2.1.4 *Derechos de evaluación y pruebas*

El Gobierno se reservará el derecho de probar, auditar y evaluar de forma independiente el sistema de IA en cualquier momento durante la vigencia del contrato. Los proveedores deben:

- Proporcionar acceso razonable a los componentes del sistema, la documentación y las interfaces de prueba.
- No imponer limitaciones contractuales que conviertan el sistema de IA en una "caja negra" inaccesible para una validación independiente.
- Facilitar las evaluaciones dirigidas por el Gobierno para la seguridad, el sesgo, la imparcialidad y la garantía de rendimiento.

6.6.2.1.5 *Transparencia y documentación*

Los proveedores deberán entregar documentación completa y precisa, suficiente para que el Gobierno pueda operar, mantener y gobernar el sistema de IA. Esto incluye:

- La arquitectura del sistema, los algoritmos o las descripciones del modelo, las características de los datos de entrenamiento (excluyendo datos sin procesar patentados, cuando sea apropiado) y las limitaciones o riesgos identificados.

- Manuales operativos para usuarios finales y administradores.
- Registros de cambios e historiales de versiones para todas las actualizaciones del sistema.

6.6.2.1.6 *Seguridad y cumplimiento ético*

Los contratos deberán estipular que los sistemas de IA se diseñen, desarrollen e implementen en alineación con los principios éticos de la IA, incluyendo:

- Evitar sesgos discriminatorios y resultados perjudiciales.
- La prohibición de código malicioso, funcionalidades ocultas o capacidades de vigilancia no autorizadas.
- El cumplimiento de las prácticas éticas de IA y las mejores prácticas internacionales relevantes.

6.6.2.1.7 *Notificación de incidentes y errores*

El proveedor debe informar de inmediato a la Agencia contratante y a PRITS sobre:

- Cualquier incidente de seguridad confirmado o bajo sospecha, violación de datos o vulnerabilidad que afecte al sistema de IA.
- Cualquier error significativo o comportamiento anómalo de la IA que pueda resultar en daño, interrupción operativa o decisiones incorrectas.
- Las acciones correctivas tomadas y los plazos para la resolución.

6.6.2.1.8 *Indemnización y responsabilidad*

Los contratos deberán incluir cláusulas que:

- Requieran a los proveedores indemnizar al Gobierno de Puerto Rico por cualquier pérdida o reclamo, como los relacionados con la propiedad intelectual o violaciones de privacidad, causados

por el sistema de IA o los datos derivados de él, o por la negligencia del proveedor.
- Establezcan la responsabilidad del proveedor por los daños causados por el incumplimiento de las obligaciones contractuales, los estándares de seguridad o los requisitos legales.

6.6.2.1.9 *Cumplimiento legal y regulatorio*

Los proveedores deberán:
- Cumplir con la Ley Núm. 40-2024 y todas las leyes aplicables.
- Cooperar con las auditorías gubernamentales, revisiones de cumplimiento y actividades de supervisión.

6.6.2.1.10 *Terminación y remediación*

Los contratos deben incluir mecanismos precisos para la suspensión o terminación en casos donde:
- El proveedor no cumpla con los estándares requeridos o las obligaciones legales.
- El sistema de IA presente amenazas imprevistas y de alto riesgo para la seguridad, la privacidad o el bienestar público.

Al finalizar el contrato, los proveedores deberán:

- Asegurar la devolución ordenada o la eliminación segura de todos los datos gubernamentales.
- Proporcionar asistencia en la transición para mantener la continuidad del servicio o habilitar soluciones de reemplazo.

6.6.3 Herramientas de IA de código abierto o gratuitas

6.6.3.1 En algunos casos, las Agencias pueden considerar el uso de *software* de IA de código abierto o servicios de IA gratuitos, en

línea o de acceso público. Si bien estas soluciones pueden ofrecer flexibilidad, rentabilidad e innovación impulsada por la comunidad, están sujetas a las disposiciones de esta Política. Las Agencias deben:

6.6.3.1.1 Realizar una evaluación exhaustiva de cualquier herramienta de IA de código abierto antes de su adopción. Esto incluye:
- Revisar si existen vulnerabilidades conocidas, puertas traseras o código malicioso.
- Asegurar que la licencia de código abierto de la herramienta sea compatible con el uso gubernamental, incluidas las restricciones sobre la modificación, redistribución o aplicación comercial.

6.6.3.1.2 No cargar, procesar o exponer datos confidenciales o sensibles a ningún servicio de IA gratuito o de acceso público sin la aprobación por escrito, como se indica en la Sección 6.3.1. El hecho de que una herramienta sea gratuita no significa que sea segura, que cumpla con la privacidad o que sea adecuada para manejar datos gubernamentales.

6.6.3.1.3 Registrar todos los componentes de IA de código abierto en el inventario de sistemas de IA y gestionarlos como cualquier otro componente de *software* de terceros. Esto incluye:
- Monitorear las actualizaciones, parches y avisos de seguridad de la comunidad de desarrolladores.
- Implementar actualizaciones oportunas para abordar las vulnerabilidades.
- Documentar el historial de versiones y las dependencias para la preparación de auditorías.

6.6.4 Colaboración e intercambio de conocimientos

6.6.4.1 PRITS actuará como facilitador para el intercambio de mejores prácticas, interpretaciones de políticas y lecciones aprendidas

en la adopción de la IA en todas las Agencias. Cuando sea posible, las Agencias deberán:

6.6.4.1.1 Aprovechar los acuerdos de compra centralizados o los contratos negociados por PRITS para servicios de IA, a fin de garantizar la coherencia y maximizar el poder de negociación colectiva.

6.6.4.1.2 Compartir estudios de caso, desafíos y estrategias de mitigación de riesgos a través de foros dirigidos por PRITS, grupos de trabajo interagenciales o plataformas de colaboración digital.

6.7 Capacitación y desarrollo de la fuerza laboral

6.7.1 Capacitación sobre IA y esta Política

6.7.1.1 Todas las Agencias deberán asegurarse de que el personal con responsabilidades relacionadas con la IA reciba capacitación apropiada para su rol sobre el uso responsable de la IA y el cumplimiento de esta Política. Esto incluye:

6.7.1.1.1 Ejecutivos y gerentes
Deberán entender las implicaciones estratégicas de la IA, los fundamentos de la gestión de riesgos y su responsabilidad bajo esta Política.

6.7.1.1.2 Equipos técnicos (desarrolladores, científicos de datos, personal de TI)
Deberán ser capacitados en prácticas de desarrollo seguro de IA, la adhesión a los principios del Marco de Gestión de Riesgos de IA (AI RMF) del NIST, las obligaciones de privacidad de los datos y la implementación de controles de imparcialidad, transparencia y seguridad.

6.7.1.1.3 Usuarios finales o responsables de la toma de decisiones

(Por ejemplo, el personal que usa las recomendaciones generadas por la IA) deberán ser capacitados sobre cómo interpretar los resultados de la IA, aplicar la supervisión humana y detectar problemas potenciales como resultados inconsistentes, sesgados o erróneos.

6.7.1.1.4 Personal de seguridad y auditoría
Deberá recibir capacitación especializada en la garantía de los sistemas de IA, incluyendo métodos para auditar sistemas de IA, monitorear anomalías y responder a incidentes de seguridad relacionados con la IA.

6.7.1.2 Cónsono con la Ley 40-2024, que exige capacitación en ciberseguridad para todos los empleados del Gobierno, esta Política extiende el alcance de la capacitación para incluir temas específicos de la IA. PRITS apoyará a las Agencias mediante el desarrollo o la recomendación de programas de capacitación que se centren en la gestión de riesgos de la IA, el uso ético y legal de la IA, y los controles de seguridad de la IA.

6.7.2 Educación continua

6.7.2.1 Dada la rápida evolución de las capacidades de la IA y las amenazas asociadas, el dominio en IA debe ser tratado como un requisito continuo. Se deberán impartir cursos de repaso o módulos de capacitación actualizados al menos anualmente, o antes si son motivados por cambios significativos en las políticas, avances tecnológicos o nuevos vectores de amenaza.

6.7.2.2 PRITS puede emitir guías actualizadas, organizar talleres o proporcionar sesiones informativas específicas sobre los desarrollos de IA, tales como marcos regulatorios, avances en

IA generativa o nuevas categorías de riesgo, para mantener al personal de la Agencia informado.

6.7.3 Talento y pericia en IA

6.7.3.1 Se alienta a las Agencias a desarrollar experiencia interna en IA mediante:

- La designación de especialistas en IA o "campeones de la IA" dentro del departamento de TI para supervisar el cumplimiento y optimizar el uso de la IA.
- El establecimiento de asociaciones con centros académicos, de investigación y proveedores de capacitación para desarrollar un flujo de talentos calificados en IA.
- Capacitación al personal con conocimientos para interactuar con los proveedores, evaluar críticamente los resultados de la IA y contribuir a la gobernanza y al perfeccionamiento de las políticas.

6.8 Monitoreo, cumplimiento e informes

6.8.1 Autoevaluación de la agencia

6.8.1.1 Las Agencias deben realizar autoevaluaciones periódicas para evaluar el cumplimiento de esta Política. Las autoevaluaciones pueden integrarse en las revisiones de control interno existentes, los procesos de gobernanza de TI o las auditorías internas formales.

6.8.1.2 Como mínimo, las autoevaluaciones deberán abordar lo siguiente:

6.8.1.2.1 *Inventario de IA*

¿La Agencia ha identificado y documentado todos los sistemas de IA en uso o en desarrollo?

6.8.1.2.2 *Evaluación de riesgos*

¿Las clasificaciones y evaluaciones de riesgos están actualizadas para todos los sistemas de IA?

6.8.1.2.3 *Aprobación de PRITS*
¿Se han revisado y aprobado por PRITS todos los usos de IA, según lo requerido?

6.8.1.2.4 *Capacitación*
¿Se ha completado y documentado la capacitación requerida del personal?

6.8.1.2.5 *Respuesta a incidentes*
¿Los planes de respuesta a incidentes de la Agencia abordan escenarios relacionados con la IA, incluyendo el posible uso indebido, fallos del sistema y daños relacionados con sesgos?

6.8.2 Informe anual de cumplimiento

6.8.2.1 Cada Agencia deberá presentar un Informe Anual de Cumplimiento de IA a PRITS (Oficina del CISO) a más tardar el 31 de enero de cada año (ver Apéndice III para la plantilla e instrucciones). Como mínimo, el informe anual incluirá:

6.8.2.1.1 *Inventario de sistemas de IA*
Una lista actualizada de todos los sistemas de IA en uso o en desarrollo, incluyendo el propósito, la funcionalidad principal y el estado operativo de cada sistema.

6.8.2.1.2 *Clasificaciones de riesgo*
La clasificación de riesgo actual (bajo, moderado, alto) de cada sistema de IA, señalando cualquier cambio con respecto al año anterior.

6.8.2.1.3 *Resumen de evaluaciones de riesgo*

La confirmación de que cada sistema de IA tiene una evaluación de riesgos actualizada, incluyendo la fecha de la última revisión, los riesgos residuales y cualquier vulnerabilidad identificada.

6.8.2.1.4 *Incidentes y problemas*
Un resumen de cualquier incidente, fallo, uso indebido o casi accidente relacionado con la IA que haya ocurrido durante el período del informe, junto con las medidas correctivas tomadas.

6.8.2.1.5 *Exenciones*
Una lista de cualquier aprobación de exención activa, incluyendo su estado de vencimiento o renovación.

6.8.2.1.6 *Capacitación*
Confirmación del cumplimiento con los requisitos de capacitación de IA, incluyendo la cantidad de personal capacitado y cualquier iniciativa de capacitación notable.

6.8.2.1.7 *Proyectos de IA planificados*
Una previsión de cualquier iniciativa de IA significativa planificada para el próximo año, con el fin de facilitar la supervisión de PRITS, la asignación de recursos y la colaboración entre Agencias.

6.8.2.1.8 Mejoras y desafíos
Un breve análisis de los desafíos enfrentados en la implementación de esta Política (por ejemplo, limitaciones técnicas, restricciones en las pruebas de sesgo, capacidad de la fuerza laboral) y cualquier necesidad de orientación, herramientas o recursos adicionales. Las Agencias también pueden compartir historias de éxito o mejores prácticas que podrían beneficiar a otras Agencias.

6.8.2.2 PRITS revisará estos informes anuales para evaluar las tendencias de cumplimiento, identificar problemas comunes e informar sobre el refinamiento de las políticas. Los resúmenes o hallazgos agregados podrían compartirse con la Oficina del Gobernador u otros organismos de supervisión según corresponda. Estos informes contribuyen a las métricas de madurez de ciberseguridad y al marco de rendición de cuentas más amplios del Gobierno.

6.8.3 Auditorías y revisiones

6.8.3.1 PRITS, en colaboración con la Oficina del Contralor u otras autoridades de auditoría, puede realizar auditorías periódicas a los sistemas de IA de las Agencias para verificar el cumplimiento de esta Política. Las Agencias deberán cooperar plenamente, incluyendo proveer el acceso oportuno a la documentación, los sistemas y las entrevistas con el personal.

6.8.3.2 Las auditorías pueden incluir, pero no están limitadas a:

6.8.3.2.1 La revisión de la arquitectura del sistema y la documentación de configuración.

6.8.3.2.2 La validación de las evaluaciones de riesgo de IA y los planes de mitigación.

6.8.3.2.3 Las pruebas de seguridad del sistema de IA y las medidas de protección de datos.

6.8.3.2.4 La verificación de los registros de finalización de la capacitación.

6.8.3.2.5 Las entrevistas con el personal operativo, técnico y de supervisión.

6.8.3.3 Los hallazgos de incumplimiento requerirán acciones correctivas dentro de los plazos definidos.

6.8.3.4 El incumplimiento grave o reiterado puede ser escalado bajo las disposiciones de ejecución de la Ley 40-2024 y podría resultar en penalizaciones u otras medidas correctivas.

6.8.4 Mejora continua de la política

6.8.4.1 PRITS revisará esta Política al menos una vez al año, o con mayor frecuencia si es necesario para abordar nuevos avances tecnológicos, riesgos emergentes o cambios en los requisitos legales o regulatorios.

6.8.4.2 Las actualizaciones de esta Política se basarán en:

6.8.4.2.1 Las mejores prácticas nacionales e internacionales emergentes, incluyendo futuras revisiones a los marcos de NIST, los estándares ISO y otras guías relevantes.

6.8.4.2.2 Las lecciones aprendidas de los informes de las Agencias, los hallazgos de auditoría y las investigaciones de incidentes.

6.8.4.2.3 La retroalimentación de las partes interesadas, incluyendo las Agencias, socios de la industria, el mundo académico y la sociedad civil.

6.8.4.3 Las Agencias y las partes interesadas pueden enviar comentarios o sugerencias a PRITS para mejorar la Política. El objetivo es mantener la Política actualizada, efectiva y receptiva a la innovación, asegurando que el enfoque de Puerto Rico hacia la IA se mantenga alineado con las mejores prácticas (como las futuras actualizaciones de los marcos de NIST) y apoye la adopción segura de nuevas y beneficiosas capacidades de IA.

6.8.5 Aplicación

6.8.5.1 El cumplimiento de esta Política es obligatorio para todas las Agencias, contratistas y terceros involucrados en el desarrollo, adquisición, implementación u operación de sistemas de IA en nombre del Gobierno de Puerto Rico.

6.8.5.2 Cuando se identifique un incumplimiento, PRITS podrá tomar una o varias de las siguientes medidas, según la gravedad y la naturaleza de la violación:

6.8.5.2.1 Emitir un aviso de incumplimiento y exigir acciones correctivas inmediatas, incluyendo la suspensión o

desactivación del sistema de IA hasta que se logre el pleno cumplimiento.

6.8.5.2.2 Imponer multas, sanciones administrativas o medidas disciplinarias contra las partes responsables, de acuerdo con la Ley Núm. 40-2024.

6.8.5.2.3 Tomar acciones legales o aplicar sanciones contractuales, que pueden llegar hasta la rescisión del contrato.

6.8.5.2.4 Remitir los casos que involucren mala conducta intencional o temeraria a los organismos de supervisión pertinentes o a las autoridades encargadas de hacer cumplir la ley para su investigación y posible acción legal.

6.8.5.3 Las violaciones persistentes, particularmente aquellas que demuestren un desprecio por las obligaciones legales o las directivas de PRITS, pueden ser escaladas a entidades de supervisión para que tomen medidas adicionales.

7. Responsabilidades

Para asegurar la implementación efectiva y la rendición de cuentas bajo esta Política, se definen los siguientes roles y responsabilidades:

7.1 Jefes de Agencia (Secretarios/Directores/Alcaldes)

7.1.1 Tener la máxima responsabilidad por el uso de la IA dentro de la Agencia.

7.1.2 Garantizar el cumplimiento de esta Política en toda la Agencia priorizando el cumplimiento, asignando recursos suficientes (personal, capacitación, presupuesto, herramientas técnicas) y responsabilizando a los subordinados por la implementación.

7.1.3 Revisar y aprobar las evaluaciones de riesgo de la IA, incluyendo la firma final, antes de la presentación a PRITS.

7.1.4 Evaluar las solicitudes de aprobación y firma el Informe Anual de Cumplimiento de IA.

7.2 Principal Oficial de Informática (OPI)

7.2.1 Servir como el principal coordinador para la implementación de esta Política dentro de la Agencia.
7.2.2 Mantener un inventario actualizado de los sistemas de IA y asegurar de que la gobernanza de la IA esté integrada en los procesos de TI.
7.2.3 Supervisar la presentación oportuna a PRITS de la documentación requerida.
7.2.4 Asegurar que los equipos técnicos realicen evaluaciones de riesgo de IA exhaustivas y apliquen los controles técnicos y de procedimiento apropiados.
7.2.5 Servir de enlace con PRITS regularmente, monitorear las actualizaciones de los estándares técnicos de IA y difundir la información dentro de la Agencia.

7.3 Oficial de Seguridad de la Información de la Agencia

7.3.1 Liderar la supervisión de la ciberseguridad de los sistemas de IA dentro de la Agencia.
7.3.2 Proporcionar información sobre las evaluaciones de riesgo de la IA, particularmente en la identificación de amenazas, vulnerabilidades y requisitos de control de seguridad.
7.3.3 Asegurar el cumplimiento de las políticas de seguridad, incluyendo los procedimientos de control de acceso, monitoreo y respuesta a incidentes.
7.3.4 Facilitar y colaborar en la investigación y remediación de incidentes de seguridad relacionados con la IA.
7.3.5 Las Agencias que carezcan de un Oficial de Seguridad de la Información dedicado deberán delegar estas responsabilidades al OPI o a otra autoridad de seguridad formalmente designada.

7.4 Principal Oficial de Seguridad Cibernética y equipo de Política de IA de PRITS

7.4.1 Proporcionar supervisión a nivel de Gobierno para la implementación y aplicación de esta Política.

7.4.2 Revisar las presentaciones de las Agencias (por ejemplo, evaluaciones de riesgo, solicitudes de exención, informes anuales de cumplimiento) y las aprueba, rechaza o solicita correcciones según corresponda.

7.4.3 Mantener y actualizar la Política, emitir guías complementarias y proporcionar plantillas, herramientas y capacitación estandarizadas a las Agencias.

7.4.4 Monitorear las métricas de cumplimiento de la IA en todo el Gobierno e informar al Director Ejecutivo de PRITS y a los organismos de supervisión relevantes.

7.5 Director Ejecutivo de PRITS

7.5.1 Servir como el patrocinador de más alto nivel para las iniciativas de política y gobernanza de IA.

7.5.2 Asegurar que PRITS dedique suficientes recursos para la supervisión de la IA e integre la gobernanza de la IA en estrategias más amplias de transformación digital.

7.5.3 Tomar las decisiones finales sobre los usos de IA de alta sensibilidad o las iniciativas intergubernamentales de IA.

7.5.4 Poder convocar consejos interagenciales o grupos de trabajo sobre la gobernanza de la IA.

7.6 Asesor Legal de la Agencia

7.6.1 Asesorar a la dirección de la Agencia sobre las implicaciones legales de la adopción y el uso de la IA, incluyendo el cumplimiento de las obligaciones de privacidad, derechos civiles y contractuales.

7.6.2 Revisar los contratos de adquisición y los Memorandos de Entendimiento (MOU) para soluciones de IA para asegurar la inclusión de las protecciones contractuales requeridas (ver Apéndice II).

7.6.3 Participar en la revisión de proyectos de IA para evaluar el riesgo legal y en las solicitudes de exención, incluyendo las implicaciones de responsabilidad.

7.7 Oficial de Datos

7.7.1 Asegurar que los sistemas de IA cumplen con todos los requisitos de privacidad aplicables.
7.7.2 Realizar y documentar Evaluaciones de Impacto a la Privacidad (EIPs) cuando se utilizan datos personales.
7.7.3 Monitorear los usos de la IA para verificar el cumplimiento de los acuerdos de consentimiento y las leyes de privacidad.
7.7.4 Gestionar y responder a las consultas o quejas de los interesados en los datos relacionadas con el procesamiento de la IA.

7.8 Gerentes de Programa / partes interesadas de la Agencia

7.8.1 Iniciar y supervisar los proyectos de IA dentro de sus respectivos dominios operativos.
7.8.2 Asegurar la coordinación temprana con los OPIs y los Oficiales de Seguridad de la Información para el cumplimiento y la evaluación de riesgos.
7.8.3 Monitorear el rendimiento del sistema de IA y su impacto ético, asegurando que los resultados se alineen con los objetivos del programa.
7.8.4 Informar sobre incidentes, anomalías o preocupaciones relacionadas con el uso de la IA.

7.9 Todos los empleados y contratistas

7.9.1 Usar los sistemas de IA siguiendo los procedimientos y la capacitación aprobados.
7.9.2 Permanecer atentos a fallos, problemas de seguridad o resultados poco éticos de la IA, e informar de tales eventos de inmediato.
7.9.3 Mantener la confidencialidad, integridad y seguridad de los datos y sistemas de IA.
7.9.4 Abstenerse de cualquier intento de manipular, eludir o hacer un mal uso de las funcionalidades de la IA.

7.10 Auditores internos y externos

7.10.1 Verificar de manera independiente el cumplimiento de esta Política a través de auditorías programadas o *ad-hoc*.

7.10.2 Proporcionar hallazgos y recomendaciones para fortalecer la gobernanza y los controles de la IA.

7.11 Oficina para la Evaluación de Incidentes Cibernéticos

7.11.1 Investigar los incidentes de ciberseguridad relacionados con la IA y coordinar con las Agencias afectadas para garantizar la debida corrección e informes.

7.11.2 Colaborar en el análisis posterior a los incidentes para mejorar las estrategias de prevención de riesgos de la IA.

8. Sanciones

PRITS podrá imponer multas a las Agencias que incumplan con lo estipulado en la Ley 40-2024. Toda divulgación de datos no autorizado, modificación no autorizada o pérdida de datos, cuando medie obstrucción, negligencia, mala fe, temeridad o negativa caprichosa en el manejo o informe de un ciberataque, conllevará sanciones, multas, y en ocasiones despidos, recisión de contrato, e inhabilitación para contratar restitución de dinero entre otras sanciones aplicables bajo la Ley 40-2024.

9. Solicitud de exenciones

Si el cumplimiento de esta Política resulta inviable o técnicamente imposible, o se requiere una exención, el propietario o administrador del sistema debe presentar una solicitud formal por escrito al CISO y al CIIO. Esta solicitud podrá canalizarse a través del Service Desk (https://support.prits.pr.gov/) o, en su defecto, al correo electrónico support@prits.pr.gov, indicando claramente el asunto (por ejemplo, "Solicitud de Exención de Política de IA – [Agencia] – [Nombre del Sistema de IA]").

10. Cláusula Derogatoria

Esta Política deroga cualquier directriz, guía o política anterior sobre el uso de la IA en el Gobierno de Puerto Rico que entre en conflicto con sus disposiciones, incluida la Carta Circular 2023-002 de PRITS. Mientras que los documentos o comunicaciones anteriores proporcionaban recomendaciones, esta Política ahora establece requisitos vinculantes.

11. Vigencia

Esta Política entrará en vigor inmediatamente.

Dentro de los 30 días siguientes a su fecha de entrada en vigor, todas las Agencias deberán:

- Notificar a PRITS, por medio de su Jefe de Agencia u OPI, que han recibido y entendido esta Política.
- Presentar a PRITS un estado de cumplimiento inicial, que debe incluir:
    - Un inventario actualizado de todos los sistemas de IA actualmente en uso, en fase de prueba o en desarrollo.
    - Una evaluación del estado de cada sistema con respecto a los requisitos de esta Política.

Para las Agencias con iniciativas de IA ya en marcha bajo guías anteriores (es decir, la Carta Circular 2023-002), los proyectos existentes deben ser revisados y actualizados para asegurar su total alineación con esta Política. PRITS puede proporcionar asesoramiento y asistencia técnica para facilitar la transición.

Las futuras enmiendas, aclaraciones o guías complementarias se comunicarán a través de Cartas Circulares oficiales de PRITS o de revisiones formales de la Política. Las Agencias son responsables de mantenerse informadas sobre estas actualizaciones y de incorporarlas de manera oportuna en sus operaciones. PRITS puede emitir boletines complementarios que aborden casos de uso de IA emergentes, riesgos o desarrollos regulatorios.

12. Referencias

Número de identificación	Título	Versión
PRITS-POL-0001	Política de Clasificación de Datos del Gobierno de Puerto Rico	1.0
PRITS-POL-0002	Política de respuesta de incidentes de seguridad de la información	1.0
NIST AI 100-1	Artificial Intelligence Risk Management Framework (AI RMF)	1.0
NIST CSWP 29	NIST Cybersecurity Framework (CSF)	2.0

Este espacio se ha dejado intencionalmente en blanco.

13. Apéndice

Apéndice I: Matriz de clasificación de riesgos

La siguiente Matriz de Clasificación de Riesgos de la IA establece los criterios estándar para clasificar los sistemas de IA como de riesgo bajo, moderado o alto. Las Agencias deben usarla al realizar evaluaciones de riesgo de la IA según la Sección 6.2.3.

- *Propósito*
  La clasificación de riesgos asegura que la gobernanza, supervisión y los controles de seguridad apropiados estén en vigor, de manera proporcional al impacto potencial de cada sistema de IA.

- *Determinación de nivel de riesgo*
  Si un sistema de IA cumple con uno o más criterios en una categoría superior, se le asignará el nivel de riesgo más alto aplicable.

- *Flexibilidad*
  Esta matriz es ilustrativa y puede ser adaptada por PRITS con el tiempo para abordar tecnologías emergentes, requisitos específicos del sector o estándares regulatorios cambiantes.

Criterios	IA de riesgo bajo	IA de riesgo moderado	IA de riesgo alto
Impacto en Individuos / Comunidad	Impacto mínimo o nulo en los individuos. Los resultados de la IA no toman decisiones que afecten derechos, servicios o la seguridad. Los errores o fallos solo causan inconvenientes menores o requieren una corrección manual rutinaria.	Impacto moderado. La IA informa decisiones que afectan a individuos o procesos, pero existe una revisión y anulación humana. Los errores pueden causar una pérdida financiera limitada, retrasos cortos en el servicio o inconvenientes moderados, pero no resultan en daños mortales o irreversibles.	Impacto alto o significativo. La IA influye directamente o determina resultados que afectan derechos, seguridad o el acceso crítico a servicios. Los errores podrían causar la denegación de servicios esenciales, violaciones de derechos, pérdidas financieras significativas, lesiones físicas u otros daños graves. Incluye IA

Criterios	IA de riesgo bajo	IA de riesgo moderado	IA de riesgo alto
			utilizada en la respuesta a emergencias, la aplicación de la ley, el *triage* de atención médica, la infraestructura crítica o servicios a poblaciones vulnerables o reconocimiento de imágenes/biométrico (por ejemplo, identificación, autenticación o vigilancia)
Sensibilidad de los Datos	Utiliza solo datos públicos o internos, no sensibles (por ejemplo, estadísticas anónimas). No se involucran datos confidenciales o restringidos.	Puede procesar algunos datos personales (PII) de sensibilidad moderada (por ejemplo, datos que no son públicos pero que causarían un daño limitado si se exponen) o pequeñas cantidades de información confidencial. Incluye casos en los que se utilizan datos altamente sensibles pero totalmente anonimizados.	Procesa datos personales sensibles a gran escala (PII, datos de salud, registros financieros, biométricos) o datos restringidos. Una violación o un uso indebido tendría consecuencias graves.
Función	Principalmente de asesoramiento o informativa con verificación humana obligatoria. Automatiza tareas sencillas y de bajo riesgo (por ejemplo, categorizar información pública,	Puede automatizar un paso de la toma de decisiones, pero con revisión y anulación humana. La IA desempeña un papel importante, pero no es la única autoridad.	Opera de forma autónoma o con una intervención humana limitada debido a los requisitos de escala o velocidad. Incluso con supervisión humana, los resultados de la IA influyen fuertemente en resultados de alto riesgo.

Criterios	IA de riesgo bajo	IA de riesgo moderado	IA de riesgo alto
	estadísticas básicas, análisis de sitios web).		
Ejemplos	IA que sugiere categorías de contenido de sitios web; un *chatbot* que proporciona información general de una base de datos de preguntas frecuentes.	Escaneo de solicitudes de empleo por IA (con revisión humana final); programación de inspecciones por IA basada en el riesgo (con posibilidad de ajustes por parte del personal); mantenimiento predictivo para equipos sin riesgo de seguridad inmediato.	IA que determina la elegibilidad para beneficios significativos o servicios sociales; diagnóstico médico asistido por IA; vigilancia predictiva; IA que controla aspectos de la infraestructura crítica (por ejemplo, ajustes de la gestión de la red eléctrica); sistemas de identificación biométrica que afectan el acceso o la libertad de movimiento; IA de procesamiento de imágenes.

IA de Riesgo Inaceptable

Ciertas aplicaciones de IA están prohibidas debido a una incompatibilidad inherente con los principios legales, éticos o constitucionales. Estas incluyen:

- Usos de la IA que son expresamente ilegales bajo la ley de Puerto Rico o la ley federal de EE. UU.
- IA que viola las protecciones constitucionales, los derechos humanos o la política pública establecida.
- IA para la puntuación social (*social scoring*) de los ciudadanos, que resulte en medidas punitivas o discriminatorias.
- Vigilancia masiva habilitada por IA sin autoridad legal.
- IA que afirme predecir el comportamiento criminal basado en la raza, el origen étnico u otros rasgos protegidos, etc.

Si una Agencia se encuentra o se le pide que implemente un sistema de IA de este tipo, deberá detener inmediatamente la actividad e informar del asunto a PRITS para su investigación y que se tomen medidas de cumplimiento.

Documentación y supervisión del nivel de riesgo

Las Agencias deben mantener una justificación por escrito para la clasificación de riesgo asignada a cada sistema de IA, incluyendo los criterios cumplidos y la razón para la determinación.

Esta matriz se aplicará en conjunto con los procedimientos detallados de evaluación de riesgos de la IA descritos en la Sección 6 de esta Política. PRITS puede revisar o expandir esta matriz para abordar los riesgos emergentes de la IA, los desafíos específicos del sector o los nuevos mandatos regulatorios.

Este espacio se ha dejado intencionalmente en blanco.

Apéndice II: Lista de verificación para cláusulas de contratos con proveedores de IA

Al redactar o revisar contratos, incluyendo órdenes de compra, acuerdos de servicio y Memorandos de Entendimiento (MOUs), con proveedores que suministran sistemas o servicios de IA, las Agencias deben asegurarse de que los siguientes temas se aborden explícitamente a través de cláusulas o requisitos aplicables.

Esta lista de verificación está destinada a ser utilizada por los oficiales de adquisiciones, los gerentes de contratos y el asesor legal de la Agencia durante el proceso de contratación para salvaguardar los intereses del Gobierno de Puerto Rico, asegurar la fiabilidad de la IA y cumplir con las leyes y políticas aplicables.

1. *Alcance del trabajo y propósito*
   - Definir claramente el sistema o servicio de IA a ser proporcionado, incluyendo su funcionalidad, alcance y propósito de uso previstos dentro de la Agencia.
   - Especificar las limitaciones de la funcionalidad para evitar casos de uso no autorizados o no intencionados.

2. *Propiedad y derechos de los datos*
   - El Gobierno de Puerto Rico conserva la plena propiedad de todos los datos que proporciona al proveedor y de todos los datos de salida generados a través del sistema de IA.
   - El proveedor no tendrá derechos de uso secundario sobre los datos del Gobierno fuera del alcance del contrato sin previa aprobación por escrito.
   - Incluir restricciones sobre la transferencia de datos a terceros o el movimiento transfronterizo de datos sin autorización explícita.

3. *Protección y privacidad de los datos*
   - El proveedor debe cumplir con todas las leyes y regulaciones de protección de datos aplicables en Puerto Rico y a nivel federal, incluyendo los requisitos para:
     - El cifrado de los datos en tránsito y en reposo.
     - El almacenamiento seguro dentro de las jurisdicciones aprobadas.
     - La prohibición de extraer, vender o reutilizar datos personales o sensibles.

- Si se procesan datos sensibles (por ejemplo, datos de salud bajo HIPAA), requerir acuerdos aplicables como un *Business Associate Agreement* o su equivalente.
- El proveedor debe certificar el cumplimiento del almacenamiento, la eliminación y las pruebas de sesgo de cualquier imagen o dato biométrico procesado por el servicio de IA.

4. *Confidencialidad*
   - Incluir una cláusula de confidencialidad que prohíba la divulgación de datos del Gobierno o de información sensible del sistema de IA a partes no autorizadas.
   - Esta obligación deberá extenderse más allá de la terminación o vencimiento del contrato.

5. *Cumplimiento de controles de seguridad*
   - El proveedor debe cumplir con los controles mínimos de ciberseguridad, haciendo referencia a los estándares de PRITS y a los marcos reconocidos (por ejemplo, NIST CSF o controles específicos).
   - Las soluciones de IA basadas en la nube deben cumplir con certificaciones de seguridad reconocidas (por ejemplo, FedRAMP, ISO 27001).
   - El proveedor debe implementar parches de seguridad y mantenimiento regulares.
   - La Agencia se reserva el derecho de realizar evaluaciones de seguridad independientes o de exigir pruebas de cumplimiento.

6. *Derechos de auditoría*
   - La Agencia tendrá el derecho de auditar o evaluar el rendimiento y el cumplimiento del proveedor, incluyendo la postura de seguridad de la IA, la gobernanza de los datos y las pruebas de sesgo.
   - Incluir disposiciones tanto para inspecciones in situ como para la revisión remota de la documentación, con un aviso razonable.

7. *Pruebas y validación*
   - El proveedor debe cooperar en las pruebas iniciales y continuas del sistema de IA.

- Si se identifican sesgos significativos, fallos de precisión o vulnerabilidades de seguridad, el proveedor debe corregirlos de inmediato.
- Incluir una cláusula de mitigación de sesgos que requiera que el proveedor proporcione documentación del modelo o resultados de la evaluación suficientes para una valoración de sesgo independiente.

8. *Niveles de servicio*
   - Definir los acuerdos de nivel de servicio (SLA) para la disponibilidad, el rendimiento y los tiempos de respuesta.
   - Especificar los procedimientos para las ventanas de mantenimiento y las actualizaciones, incluyendo las notificaciones anticipadas de los cambios que puedan alterar el comportamiento de la IA.

9. *Gestión de cambios*
   - Requerir una notificación anticipada y, cuando corresponda, la aprobación de la Agencia para los cambios en el *software* o en el modelo de IA que puedan afectar los resultados.
   - Prohibir la implementación de cambios de versiones mayores sin pruebas y la aceptación previa de la Agencia.

10. *Entregables – Documentación*
    - El proveedor debe proporcionar:
      - Manuales de usuario y de administrador.
      - Documentación del modelo (por ejemplo, fichas técnicas *o model cards*) que describa la funcionalidad, las limitaciones y los usos previstos.
      - Esquemas de datos y guías de integración.
    - La documentación debe mantenerse actualizada y reflejar cualquier cambio importante.

11. *Derechos de propiedad intelectual (PI)*
    - Clarificar la titularidad de la PI:
      - El proveedor conserva los derechos sobre la PI preexistente, pero otorga a la Agencia las licencias necesarias o el acceso perpetuo para usar la versión adquirida.
      - Para la IA desarrollada a medida, otorgar a la Agencia los derechos de uso, modificación y mantenimiento de la solución.

- Evitar las cláusulas de dependencia que impidan la comprensión o modificación del sistema.

12. Indemnización

El proveedor deberá indemnizar a la Agencia por:

- Reclamaciones de infracción de PI (es decir, si alguien demanda, alegando que la IA infringe su patente o que los datos de entrenamiento se utilizaron ilegalmente).
- Daños causados por los resultados del sistema de IA, negligencia del proveedor, brechas de seguridad o incumplimiento.
- Incluir indemnizaciones específicas por el uso ilegal de datos o la adquisición de datos de entrenamiento.

13. Limitaciones de responsabilidad

- Asegurar que los límites de responsabilidad no se establezcan tan bajos que carezcan de sentido.
- Excluir las brechas críticas (por ejemplo, violaciones de confidencialidad, infracción de PI, incidentes de seguridad importantes) de los límites estándar.

14. Derechos de terminación

- La Agencia puede rescindir el contrato con causa justificada (por ejemplo, incumplimiento grave, falta de entrega) o por conveniencia con previo aviso.
- Tras la terminación, el proveedor debe garantizar el acceso ininterrumpido a los datos o entregables de la Agencia.

15. Asistencia para la transición

- Requerir al proveedor que proporcione una transición ordenada de los servicios y datos al finalizar o terminar el contrato, incluyendo:
    - La devolución de datos en un formato utilizable.
    - La exportación del modelo (si aplica).
    - La asistencia a los proveedores sucesores sin interrupción del servicio.

16. Cumplimiento de leyes, políticas y regulaciones

- Los contratos se regirán por las leyes de Puerto Rico.
- El proveedor debe cumplir con:

- La Ley de Ciberseguridad de Ciberseguridad del Estado Libre Asociado de Puerto Rico.
- Esta Política de IA.
- Todas las leyes locales y federales aplicables (ciberseguridad, privacidad, controles de exportación, etc.).

17. Penalidades

- Cuando sea apropiado, incluir penalizaciones por incumplimientos del SLA o fallas de seguridad.

18. Gobernanza e informes

- Requerir informes periódicos del proveedor sobre el rendimiento del sistema, los problemas, las actualizaciones, etc.
- Para contratos de varios años, exigir una certificación de cumplimiento anual.

19. Subcontratistas

- El proveedor sigue siendo responsable de los subcontratistas y debe imponerles obligaciones contractuales equivalentes.
- Se requiere la aprobación de la Agencia para los subcontratistas que manejen datos sensibles.

20. Resolución de disputas

- Asegurar que los términos de resolución de disputas no fuercen a la Agencia a jurisdicciones o foros de arbitraje desfavorables (por ejemplo, fuera del Estado).

Las Agencias deberán usar esta lista de verificación al redactar, revisar o negociar contratos relacionados con la IA. Si falta algún requisito, los equipos de adquisiciones y legales deben trabajar para incluir protecciones equivalentes

Este espacio se ha dejado intencionalmente en blanco.

Apéndice III: Plantillas para el informe anual de cumplimiento de IA

El Informe Anual de Cumplimiento de IA de cada Agencia (según la Sección 6.8.2) deberá seguir una estructura estándar para facilitar su revisión. A continuación, se presenta un esquema de plantilla y algunas tablas de ejemplo que las Agencias pueden completar:

[Nombre de la Agencia] – Informe Anual de Cumplimiento de IA – [Año del Informe]

1. Resumen ejecutivo
   Proporcione una breve descripción (un párrafo) de las actividades de IA de la Agencia para el año del informe. Resuma el compromiso de la Agencia con esta Política, los logros principales y cualquier cambio notable en el uso o la gobernanza de la IA.

2. Resumen del inventario de IA
   Cada agencia deberá mantener y presentar un inventario actualizado de todos los sistemas de Inteligencia Artificial (IA) en uso, en desarrollo o que hayan sido retirados durante el año del informe. El inventario deberá incluir tanto proyectos en producción como pilotos, prototipos y pruebas conceptuales. Por ejemplo:

Nombre / ID del Sistema de IA	Descripción / Propósito	Estado (Piloto / Desarrollo / Producción / Retirado)	Nivel de Riesgo (Bajo / Mod / Alto)	Clasificación de los Datos de Entrada (Ej. Público, Interno, Confidencial, Restringido)	Fecha de Aprobación de PRITS	Notas (Ej. Proveedor o Interno)
Ejemplo: "ChatBotGov"	Chatbot de cara al público para consultas generales de servicios.	Producción (desde mar 2025)	Moderado	Públicos	Aprobado por PRITS el 1 de febrero de 2025	Proveedor: ABC Corp, alojado en la nube
"DataScan AI"	Herramienta interna para evaluar la finalización de formularios.	Prueba piloto	Bajo	Internos	Fase piloto (info presentada en jun 2025)	Prototipo interno
"RiskPredict"	Análisis predictivo para identificar casos de alto riesgo para inspección.	Desarrollo (etapa de RFP)	Se espera que sea Alto	Restringido	(Presentación planificada)	Proveedor externo, adquisición en curso

- Los sistemas dados de baja deberán incluirse en un apéndice, indicando expresamente la fecha y motivo del retiro.
- Todas las entradas deben reflejar el estado más reciente a la fecha del informe.
- La clasificación de los datos de entrada deberá ajustarse a la política vigente de clasificación de datos del Gobierno de Puerto Rico y será un elemento obligatorio en la documentación de inventario, con el fin de garantizar un manejo seguro, responsable y conforme a los principios de privacidad y ética en IA.

3. <u>Evaluaciones y mitigaciones de riesgos</u>
   - Para cada sistema de IA (especialmente los de riesgo moderado y alto), proporcione:
     - La fecha de la última evaluación o revisión de riesgos.
     - Un resumen de los riesgos clave identificados, incluyendo el uso indebido de la IA y las medidas de mitigación en vigor.
     - Una confirmación del monitoreo continuo para los sistemas de alto riesgo.
     - Cualquier cambio en la clasificación de riesgos desde el informe anterior, junto con la justificación.

   - Ejemplos:

     - *ChatBotGov:*
       - Evaluación de riesgos: enero de 2025.
       - Riesgos: Posibles respuestas inapropiadas; bajo riesgo de privacidad de datos (no se recopila información de identificación personal).
       - Mitigaciones: Biblioteca de respuestas curada y monitoreo humano.
       - Nivel de riesgo: Moderado.
       - Monitoreo: Se implementaron revisiones semanales del registro de chat.

     - *RiskPredict:*
       - Evaluación de riesgos: Evaluación preliminar durante la Solicitud de Propuestas (RFP).

- Riesgos: Probablemente altos debido al impacto directo en los individuos. Los riesgos incluyen sesgos en la toma de decisiones y el acceso no autorizado a los resultados predictivos.
- Mitigaciones: Revisión humana de los casos marcados, pruebas de sesgo antes de la implementación.
- Monitoreo: Evaluación completa programada para agosto de 2025.

4. Lista de verificación de cumplimiento de la Política
   - Auto certificación de cumplimiento con las disposiciones de esta Política de IA. Las agencias deben proporcionar explicaciones y las correcciones planificadas para cualquier respuesta "No".

Área de la Política	**Cumplimiento (Sí / No / En Progreso)**	**Notas**
Gobernanza: - Líder de IA designado - Comité de gobernanza establecido - Política/procedimientos internos actualizados y alineados con la Política de IA	 Sí Sí En progreso	Líder: María López, OPI
Capacitación: % del personal relevante capacitado en la Política de IA	92 %	Actualización anual planificada para el primer trimestre de 2026
Gestión de Datos: Todos los datos de IA clasificados y protegidos según la política	Sí	N/A
Respuesta a Incidentes: Incidentes relacionados con la IA informados según el protocolo	Sí	Sin incidentes mayores
Exenciones: Exenciones activas aprobadas listadas	Ninguna	__

5. Incidentes o problemas relacionados con la IA
   - Documente cualquier evento relacionado con la IA, incluyendo interrupciones operativas, hallazgos de sesgo o brechas de seguridad:
     - La fecha, la naturaleza del incidente, el impacto y las medidas correctivas tomadas.
     - Si no ocurrió ninguno, indique explícitamente: "No se registraron incidentes relacionados con la IA durante el período del informe".
   - Ejemplos:
     - Julio de 2025 – Interrupción del *chatbot* (4 horas) debido a un fallo en la API del proveedor. Se activó la contingencia de preguntas frecuentes; la resolución del proveedor fue confirmada mediante revisión del análisis de la causa raíz (RCA).
     - Septiembre de 2025 – El modelo de lenguaje tuvo un rendimiento deficiente al procesar formularios en inglés. Se contactó al proveedor para volver a entrenar el modelo con conjuntos de datos bilingües.

6. Mejoras e iniciativas de gobernanza
   - Resuma las medidas tomadas para fortalecer la gobernanza de la IA, tales como:
     - El desarrollo de nuevos formularios o flujos de trabajo internos de revisión de la IA.
     - La participación en grupos de trabajo interinstitucionales de IA liderados por PRITS.
     - La implementación de talleres de pruebas de sesgo o capacitación en ética de la IA.
     - La integración del seguimiento de activos de IA en sistemas más amplios de gestión de activos de TI.

7. Planes a futuro
   - Describa las actividades de IA previstas para el próximo año del informe, incluyendo:
     - Nuevos proyectos de IA en etapas de viabilidad, adquisición o diseño.
     - Actualizaciones, expansiones o desmantelamientos planificados de sistemas de IA.
     - El apoyo u orientación solicitados a PRITS (por ejemplo, capacitación, guías de alojamiento, plantillas de adquisición).

8. Comentarios adicionales (opcional)
   - Incluya cualquier historia de éxito notable, desafíos o lecciones aprendidas. Aquí se pueden destacar las métricas de impacto positivo o los ahorros de costos derivados de las implementaciones de IA.

Firma

- Preparado por: [Nombre, Cargo] – generalmente el OPI de la Agencia o el Coordinador de IA designado.
- Aprobado por: [Director de la Agencia o Ejecutivo Autorizado] – con fecha.

Anejos

- La documentación de apoyo relevante (informes de evaluación de riesgos, evaluaciones de impacto en la privacidad, matrices de cumplimiento, informes de incidentes) debe adjuntarse por separado, con los puntos clave resumidos en el informe principal.

Presentación

Los informes deben ser presentados a PRITS **antes del 31 de enero** de cada año, utilizando el método de presentación oficial.

Este espacio se ha dejado intencionalmente en blanco.

14. Description

This Policy establishes a strategic and operational framework for the responsible adoption, implementation, and oversight of Artificial Intelligence (AI) technologies within the Government of Puerto Rico. It provides the principles to ensure that the integration of AI systems across government entities maintains the highest standards of transparency, accountability, ethics, and public trust.

In line with internationally recognized best practices, including the National Institute of Standards and Technology's (NIST) AI Risk Management Framework (AI RMF 1.0) and Cybersecurity Framework (CSF 2.0), this Policy promotes the development and use of AI systems that are safe, fair, and trustworthy. It also fosters innovation in compliance with applicable laws and regulations, taking into account emerging technological advancements and ethical implications. Through this framework, the Government of Puerto Rico seeks to cultivate a culture of responsible AI use that enhances the quality, efficiency, and equity of public services.

15. Legal Basis

This Policy is issued pursuant to Law No. 75-2019, the “Puerto Rico Innovation and Technology Service Act” (PRITS), Law No. 40-2024, the “Commonwealth of Puerto Rico Cybersecurity Act,” Law No. 151-2004, as amended, the “Electronic Government Act,” and Law No. 122-2019, the “Open Data Act of the Government of Puerto Rico.”

16. Purpose

By integrating principles of risk management, privacy protection, and ethical use, this Policy aims to establish a unified, enforceable governance framework for the responsible, secure, and effective use of artificial intelligence across Government entities in Puerto Rico. It is intended to safeguard individual rights and the integrity of public data, while promoting the development and deployment of AI systems that are transparent, accountable, and aligned with internationally recognized best practices.

17. Scope

The provisions of this Policy apply to the Executive Branch of the Government of Puerto Rico, including any department, board, dependency, commission, bureau, office, agency, administration, or organism, political subdivision, public corporations, and municipalities[3]. It also applies to any natural or legal person who does business or has contracts with the Government of Puerto Rico, including, but not limited to, private persons performing public functions and services, but only with respect to the public functions and services performed; to any exercise of public or private administration in which public funds or resources have been dedicated or invested, either directly or indirectly, or over which the authority of any public servant has been exercised, with respect to the data generated as a result of such activities.
Specifically, this Policy governs:

- All AI systems and services used, developed, procured, or maintained by or on behalf of government organisms. This includes, but is not limited to, machine learning models, algorithms, robotic process automation with AI components, natural language processing (NLP) services, generative AI tools, decision-support systems using AI, and any other systems that incorporate artificial intelligence or machine learning functionalities.

- All stages of AI use, from initial planning, proposal, development, or procurement, through testing, deployment, maintenance, and eventual decommissioning. The Policy also applies to significant modifications, upgrades, or repurposing of AI systems already in use.

- All Agency personnel and contractors involved in the design, implementation, management, oversight, or use of AI systems within government operations. This includes government employees, external contractors or consultants, and third-party vendors or service providers. All such individuals and entities are required to comply with the provisions of this Policy. Furthermore, any contractual agreements involving AI services must include terms that reflect and enforce the requirements of this Policy.

[3] See, Article 2 of Law 40-2024.

18. Abbreviations, Acronyms, Definitions, and Meanings

Abbreviation / Acronym	**Meaning**
AI	Artificial Intelligence
AIA	Algorithmic Impact Assessment
API	Application Programming Interface
CIIO	Chief Information and Innovation Officer of the Government of Puerto Rico
CIO	Chief Information Officer
CISO	Chief Information Security Officer of the Government
CSF	Cybersecurity Framework
DEIA	Diversity, equity, inclusion, and accessibility
ERM	Enterprise Risk Management
FedRamp	Federal Risk and Authorization Management Program
HIPAA	Health Insurance Portability and Accountability Act
IEC	International Electrotechnical Commission
IP	Intellectual Property
ISO	International Organization for Standardization
IT	Information technology
MFA	Multi-factor Authentication
MOU	Memorandum of Understanding
NIST	National Institute of Standards and Technology
NLP	Natural Language Processing
PHI	Protected Health Information
PIA	Privacy Impact Assessment
PII	Personally Identifiable Information
PoLP	Principle of Least Privilege
PRITS	Puerto Rico Innovation and Technology Service
RBAC	Role-Based Access Control
RCA	Root Cause Analysis
RFP	Request for Proposals
RMF	Risk Management Framework
SIEM	Security Information and Event Management
SLA	Service-Level Agreement

Term	Definition
Adversarial attack	Malicious activity where an attacker attempts to corrupt or manipulate an artificial intelligence or machine learning system. This is done by introducing deceptive or malicious data during the training or operational phases to force the system into making incorrect decisions or behaving in an unintended manner.
Agency (or Government organism)	Includes all entities and organisms that comprise the Executive Branch of the Government of Puerto Rico and its municipalities. This includes, but is not limited to, any department, board, dependency, commission, bureau, office, agency, administration, organism, political subdivision, and public corporations. It also encompasses the set of functions, positions, and posts that constitute the entire jurisdiction of an appointing authority of these agencies, regardless of how it is denominated.
Agency AI Inventory	A current list of all AI systems in use or development within an Agency, including key details such as purpose, the data inputs, the vendor or developer, deployment status, and designated risk classification.
AI Governance Board / Committee	A body established within an Agency (or Government-wide) to oversee AI strategy, risk management, and compliance. It may consist of the Agency's technology leaders (e.g., CIO), security officers, data officers, legal advisors, and program stakeholders.
AI Use Case	A specific application or function for which an AI system is employed (for example: chatbot for citizen services, AI-assisted decision support in benefit determinations, predictive analytics for maintenance, etc.).
Algorithmic Impact Assessment (AIA)	A systematic process used to identify, analyze, and mitigate the potential risks and impacts that an artificial intelligence system may have on individuals, society, and fundamental rights. This analysis examines the algorithm's design, development, and deployment to determine its risk level and the necessary safeguards. Its goal is to ensure the AI system is transparent, fair, and accountable, especially when used in critical decisions with legal, financial, or civil consequences.
Artificial Intelligence (AI) System	Any software, platform, algorithm or system that performs tasks which typically require human intelligence, by processing data to make or inform decisions, recommendations, or predictions. This includes machine learning models, expert systems, neural networks, natural language generation and understanding systems, computer vision systems, and other emerging AI technologies.

Term	Definition
Checksum	Value computed from the contents of a data object using a specific function, which is then stored or transmitted along with the object. Its primary purpose is to detect errors, unauthorized alterations, or manipulations of the data by verifying whether the computed checksum matches the original value.
Chief Information Officer	Individual responsible for managing and overseeing a government Agency's technology infrastructure and operations.
Confidentiality	Preserve restrictions on access and disclosure, including means to protect privacy and confidential information.
Critical Infrastructure	Essential services, systems, resources, and assets, whether physical or virtual, whose incapacity or destruction would have detrimental impacts on cybersecurity, health, the economy, the security of Puerto Rico, or any combination of these matters.
Data	Any sequence of one or more symbols that are given meaning by specific acts of interpretation.
Data Classification	The categorization of data according to sensitivity and criticality, as defined in PRITS-POL-0001 (Data Classification Policy). For example, data may be classified as public, internal, confidential, or restricted (or similar tiers defined by PRITS) with corresponding handling requirements.
Data Poisoning	A poisoning attack in which an adversary controls part of the AI system's training data.
Impact	The magnitude of the harm that can be expected as a result of the consequences of unauthorized disclosure, modification, or destruction of information, loss of information, or unavailability of the information system.
Information Technology (IT)	For an Agency, means any interconnected system, or resource, or subsystem of equipment used in the acquisition, storage, analysis, evaluation, manipulation, handling, movement, control, display, switching, exchange, destruction, transmission, or automatic reception of data or information, if the equipment is used by the Agency directly or by a third party under a contract with the Agency that requires the use (i) of that equipment; or (ii) of such equipment to a significant extent for the provision of a service or the supply of a product. It includes computers, ancillary equipment (including imaging peripherals, input, output, and storage devices necessary for security and surveillance), peripheral equipment designed to be

Term	Definition
	controlled by a computer's central processing unit, software, firmware, and similar procedures and services (including support services), and related resources.
Lifecycle (of AI)	All phases in an AI system's existence – including design, data collection, development, testing, deployment, operation, monitoring, maintenance, and decommissioning.
Municipality(-ies)	Any of the 78 municipalities of Puerto Rico. For purposes of this policy, the terms Municipality(-ies) and Agency(-ies) will be used interchangeably.
NIST AI Risk Management Framework (AI RMF)	A framework published by NIST (Version 1.0, 2023) providing a structured approach (with Govern, Map, Measure, and Manage functions) to identify, assess, and mitigate risks of AI systems across their lifecycles. The AI RMF emphasizes integration of trustworthiness characteristics into AI design and deployment.
NIST Cybersecurity Framework (CSF) 2.0	The 2024 updated NIST framework for managing cybersecurity risk, organized into six core functions: Govern, Identify, Protect, Detect, Respond, Recover. CSF 2.0 provides outcomes and controls to protect systems (including AI systems) and data against cyber threats, including a new emphasis on governance.
Personal Data	Any information relating to an identified or identifiable individual (natural person). This includes data that can directly or indirectly identify a person, consistent with definitions in privacy laws and regulations.
Personally identifiable information (PII)	Any representation of information that is readable without the need for a special cryptographic key to access it or that facilitates the tracing of an individual's identity, including the name or first initial and last name of an individual combined with other information that is linked or likely to be linked to a specific individual, such as: - Social Security number - Driver's license number, voter registration card, or other government-issued ID - Bank or financial account numbers of any kind, with or without access codes, that may have been assigned to you - Usernames and passwords for access to public or private computer systems - HIPAA-protected health information - Tax information - Job evaluations

Term	Definition
Principle of Least Privilege	Each module (process, user, or program, depending on the topic) can only access the information and resources necessary for its legitimate purpose.
Privacy Impact Assessment (PIA)	Comprehensive analysis and formal documentation of how personally identifiable information is collected, used, processed, stored, maintained, disseminated, disclosed, and disposed of within an information system. The purpose of a PIA is to: (i) ensure that the handling of such information conforms to applicable legal, regulatory, and policy requirements regarding privacy; (ii) determine the risks and potential effects associated with these activities; and (iii) examine and evaluate protections, safeguards, and alternative processes to mitigate privacy risks. PIAs are conducted throughout the system or program development life cycle to help identify, assess, and reduce potential privacy concerns.
Provider, external provider, service provider, third-party provider	Entity, whether a natural or legal person, public or private, that provides services such as networks, applications, programs, infrastructure, or security means through continuous and regular support, as well as active management services either at the facilities of an Agency, at the Agency's data processing center (hosting), or at a third-party data processing center.
Risk	Refers to any reasonably identifiable circumstance or event that has a possible adverse effect on the security of information networks and resources. It is generally determined by: (i) the adverse impacts that would arise if the circumstance or event occurs; and (ii) the probability of occurrence. Security risks related to the information system are those that arise from the loss of confidentiality, integrity, or availability of information or information systems, and they reflect the potential adverse impacts on the organization's operations (including its mission, functions, image, or reputation), the assets of government entities, individuals, other organizations, and the Government.
Risk appetite	General level and types of risk an organization is willing to accept in achieving its mission, vision, or value creation.
Risk Classification (of AI Systems)[4]	The designation of an AI system's overall risk level – low, moderate, or high – as determined by the potential impact the AI's use could have on individuals, communities, or government operations. In general, a high-risk AI system is one that influences decisions with significant effects on civil rights, safety, access to critical services or

[4] The criteria for each level are provided in Appendix I: AI Risk Classification Matrix.

Term	Definition
	resources, or the well-being of individuals or communities. Low-risk refers to applications with minimal impact or sensitivity, while Moderate-risk lies in between.
Role-Based Access Control	Set of authorizations granted to a user upon assuming a role, either explicitly or implicitly. The permissions of a role can be inherited through a role hierarchy and typically reflect the permissions necessary to perform specific functions within an organization. The same role can be assigned to one or more individuals.
Sensitive Data	Any information that is classified as Confidential, Restricted, or otherwise sensitive per Data Classification Policy for the Government of Puerto Rico (PRITS-POL-0001) – for example, personally identifiable information (PII), protected health information (PHI), financial records, critical infrastructure details, or any data that if improperly accessed or used could harm individuals or the public interest.
Trustworthy AI	AI systems that are demonstrably valid and reliable, safe, secure and resilient, accountable and transparent, explainable and interpretable, privacy-enhanced, and fair (with harmful bias managed). These characteristics are derived from NIST's AI RMF and represent the qualities that AI systems must have to maintain public trust.
Zero Trust Architecture	A security model, a set of system design principles, and a coordinate cybersecurity and system management strategy based on the recognition of the existence of threats both inside and outside the network. This model eliminates implicit trust in any element, component, node, or service, and instead requires continuous and real-time verification of the operational landscape to determine access and other system responses.

Any terms not explicitly defined here shall be interpreted following applicable legal provisions, including Laws 75-2019 and 40-2024, as well as prevailing industry definitions, guidance, and documents, such as those issued by the NIST.

19. Policy

19.1 AI Governance and Accountability

19.1.1 Governance Structure

Each Agency shall establish an internal governance mechanism to oversee the responsible use of AI. The Agency Head (e.g., Secretary or Director) shall assign clear accountability for AI oversight. This responsibility may be delegated to the Chief Information Officer (CIO), another designated senior official, or to an internal AI Governance Committee.

If an AI Governance Committee is established, it shall include representatives from executive leadership, information technology (IT), cybersecurity, legal, and relevant programmatic stakeholders. This body shall be responsible for enforcing this Policy, ensuring risk assessments are performed, and securing all necessary approvals before the use of AI systems.

19.1.2 Policies and Procedures

Agencies shall develop or update internal procedures to ensure the implementation of this Policy. These procedures shall address the proposal of new AI use cases, performance of risk and impact assessments, coordination with PRITS for review and approval, ongoing monitoring of AI systems, and mechanisms for reporting compliance. All procedures shall align with this Policy and any additional technical directives issued by PRITS. Agencies should ensure that processes for identifying, evaluating, and managing AI-related risks are clearly defined, consistently applied, and fully transparent.

19.1.3 Accountability

Ultimate accountability for the appropriate use of AI within each Agency resides with the Agency Head. The individual designated as AI lead, whether the CIO or chair of the Agency's AI governance body, is responsible for day-to-day compliance and risk management. Agencies must maintain thorough documentation of decisions regarding AI system design, procurement, deployment, monitoring, and decommissioning, including the identity of the approving authority. Agencies must also ensure that any AI system influencing public services or individual rights incorporates mechanisms for human oversight, review, and accountability.

19.1.4 PRITS Oversight

Agencies must document the purpose, intended use, and rationale for implementing any AI system. This documentation must be made available to PRITS and other relevant oversight bodies upon request. PRITS, acting through the Office of the Chief Information Security Officer (CISO) and under the leadership of its Executive Director, will oversee the government-wide use of AI. PRITS is responsible for maintaining this Policy, offering guidance, and reviewing Agency submissions related to AI. PRITS may also issue supplemental standards or advisories as needed to address emerging risks or technologies related to AI. Agencies must not deploy AI systems or make significant modifications to them without PRITS' prior written approval. For this Policy, "significant modifications" refer to substantive changes that could alter the system's functionality, risk profile, or compliance obligations. Examples include introducing new datasets, retraining models on different data, changing the system's decision-making scope or purpose, integrating with new external systems, or applying major software or algorithmic upgrades. Routine maintenance, minor software patches, or configuration adjustments that do not affect risk or core functionality are not considered significant modifications.

PRITS may impose conditions, mandate modifications, or deny the use of AI systems where necessary to mitigate risk or ensure alignment with public policy objectives.

19.1.5 Compliance with NIST Governance Principles

Both agency-level and government-wide AI governance shall align with the *Govern* function of the NIST frameworks. This includes establishing clear roles and responsibilities, policies, and risk management strategies; incorporation of principles related to diversity, equity, inclusion, and accessibility (DEIA); and engaging stakeholders in a meaningful manner. Governance mechanisms should ensure that AI initiatives support the public interest and proactively address, rather than react to, potential harms and unintended consequences.

19.2 Risk Management and Assessment

19.2.1 AI Risk Management Framework

19.2.1.1 Agencies shall apply the NIST AI Risk Management Framework (AI RMF 1.0) to guide the risk management process across all stages of AI system procurement, development, deployment, and maintenance. By following AI RMF principles, Agencies ensure that trustworthiness characteristics (such as validity, security, etc.) are considered at each stage. The AI RMF is structured around four core functions that Agencies must implement:

19.2.1.1.1 *Govern*
As outlined in Section 19.1, Agencies must establish an oversight mechanism that ensures continual review and accountability in AI risk management.

19.2.1.1.2 *Map*
Identify and contextualize the AI system's intended purpose, stakeholders, and potential impacts. Agencies must clearly define the context, scope, and objectives for each AI system, including legal requirements and expectations.

19.2.1.1.3 *Measure*
Assess and analyze the AI system's risks and performance against defined metrics and benchmarks. Agencies shall evaluate AI systems for their accuracy, reliability, robustness, fairness, privacy, security, and alignment with expected outcomes.

19.2.1.1.4 *Manage*

Implement risk mitigation strategies and make informed decisions regarding the use, restriction, or modification of AI systems considering the associated risks. Agencies must prioritize high-risk scenarios and establish ongoing monitoring and controls.

19.2.1.2 AI risk management should not be siloed. Agencies must integrate it into their overall Enterprise Risk Management (ERM) and cybersecurity risk management programs. Decisions regarding AI-related risks must align with the Agency's risk appetite and the broader cybersecurity posture.

19.2.2 Pre-Deployment Risk Assessment

19.2.2.1 Before any new AI system or implementing significant modifications to an existing one, Agencies must conduct a thorough AI risk and impact assessment. This assessment must be performed before production deployment and submitted to PRITS for review and approval, per Section 19.1.4.

19.2.2.2 The risk and impact assessment shall include, at a minimum, an evaluation of the following areas:

19.2.2.2.1 *Potential Impacts*

Agencies must analyze how the AI system may affect individuals, communities, and government operations. This includes potential for unintended harm, misuse of the AI system, or disruption of essential services. Misuse may consist of the AI system beyond its intended scope, manipulation for unauthorized purposes, or deployment without proper safeguards. Special attention shall be given to the risk of discriminatory outcomes,

infringement on civil rights, erosion of public trust, and impacts on privacy, public health, and safety.

19.2.2.2.2 *Data Sensitivity*

Agencies must review the types of data the AI system will use, including input, training, and output data. All data must be classified according to the Data Classification Policy for the Government of Puerto Rico (PRITS-POL-0001) or any other PRITS- issued regulation, and managed according to its sensitivity level, using appropriate safeguards such as encryption, access controls, and audit logging.

19.2.2.2.3 *Bias and Fairness*

Agencies must assess whether the AI system could result in discriminatory or biased outcomes. When AI systems are used for decision-making about individuals (e.g., service eligibility), Agencies must conduct an Algorithmic Impact Assessment (AIA) and implement corrective measures to address identified disparities, ensuring fairness and equity.

19.2.2.2.4 *Transparency and Explainability*

Agencies must evaluate the extent to which the AI system's logic and outputs can be explained to end users and affected individuals. For any high-risk AI system, particularly those affecting individuals, Agencies must ensure that decisions can be explained upon request.

19.2.2.2.5 *Security Risks*

Agencies must assess the cybersecurity risks associated with the AI system throughout its lifecycle. This includes evaluating the system's vulnerability to adversarial attacks (such as data poisoning or model manipulation), as well as the risk of unauthorized access to the AI system itself or a breach of sensitive data it processes. Agencies must ensure alignment with NIST CSF 2.0 functions (Identify, Protect, Detect, Respond, and Recover) and implement safeguards to protect both the system and its data against internal and external threats.

19.2.2.2.6 *Reliability and Safety*

Agencies must evaluate whether the AI system performs as intended under normal and stress conditions. Failure modes must be documented, and fallback procedures established. No AI system shall be deployed if its failure could endanger life, safety, or property (especially for any AI systems that have physical-world effects, like in traffic management, healthcare, etc.).

19.2.2.2.7 Compliance

Agencies must verify that the AI system complies with all applicable laws and regulations. This includes, but is not limited to, local and federal laws, as well as applicable provisions regarding privacy, civil rights, intellectual property, and other relevant areas.

19.2.2.3 All findings must be documented in a formal AI Risk Assessment Report, which will serve as the official record of the Agency's due diligence. The report must include:

19.2.2.3.1 A summary of the risk analysis across the seven domains listed above.

19.2.2.3.2 Identified risk and recommended mitigation strategies.

19.2.2.3.3 A justification for proceeding with, modifying, or halting the AI system's deployment.

19.2.2.4 This report must be submitted to PRITS as part of the approval request before production deployment (per Section 19.1.4). PRITS may issue standardized templates or checklists for such assessments. Agencies should refer to any such supplemental guidance.

19.2.3 Risk Classification

19.2.3.1 As part of the risk assessment, Agencies shall classify each AI system as low, moderate, or high risk based on the nature and severity of potential impacts (refer to Appendix I for classification criteria). This classification determines the rigor of controls and oversight required:

19.2.3.1.1 *Low Risk*

AI systems that pose minimal risk to individuals, services, or operations. These systems use only non-sensitive data and have no significant impact on decisions or services (for example, an AI tool used to summarize or organize publicly available information). Basic documentation, controls, and monitoring are sufficient.

19.2.3.1.2 *Moderate Risk*

AI systems that handle moderately sensitive data or influence decisions of some consequence, but where errors or misuse would have a limited scope of harm. This includes situations where the AI could have

noticeable but contained effects if incorrect or misused, meaning the impact might cause inconvenience, delay, or temporary disruption, but without causing lasting damage to individuals' rights, safety, or access to essential services. For example, an AI system that assists in prioritizing routine maintenance requests for non-critical public infrastructure, where an error might delay certain repairs but would not compromise public safety or infringe upon citizen rights. These systems require standard testing, human oversight, and more comprehensive documentation and controls to ensure responsible use.

19.2.3.1.3 *High Risk*

AI systems that could significantly impact individual rights, public safety, access to essential services, or the fair allocation of government resources are considered high risk. This category includes systems that process highly sensitive data, such as personally identifiable information (PII) or protected health information (PHI), particularly when used at scale. Examples include AI applications for eligibility determinations, healthcare triage, public safety dispatching, resource prioritization, as well as computer-vision or biometric-recognition systems used for identification, authentication, or surveillance. High-risk systems require rigorous safeguards, including fairness and bias assessments, regular audits, and ongoing monitoring to ensure compliance. PRITS may impose additional conditions or restrictions as deemed necessary.

19.2.3.2 If an AI system presents an unacceptable risk, meaning it is inherently incompatible with legal, ethical, or policy standards, such as systems designed for unlawful surveillance, social scoring, or use cases that violate fundamental rights, its deployment is strictly prohibited. Such use cases shall be classified as banned, and agencies shall not conduct procurement or development activities related to them.

19.2.4 Continuous Monitoring and Periodic Reassessment

19.2.4.1 Risk management for AI systems must be continuous and adaptive. Agencies shall implement mechanisms to monitor AI systems throughout their operational life to identify new or changing risks. Monitoring activities should include performance metrics, error rates, incident reports, user feedback, and changes in data or use context.

19.2.4.2 Agencies shall re-evaluate the risk classification and risk assessment at least annually for each AI system. Reassessment is also required before any significant modification to the system, such as:

- Integration of new data sources,
- deployment in new operational contexts,
- major algorithm or model updates, or
- the occurrence of an incident involving the AI system.

19.2.4.3 If monitoring reveals that an AI system is no longer meeting the trustworthiness criteria, such as accuracy, fairness, reliability, or compliance, corrective actions must be taken without delay. These may include retraining the model, improving data quality, strengthening safeguards, or, in severe cases, suspending the system's use until issues are resolved.

19.2.4.4 Agencies must report significant changes in AI risk or any AI-related incidents to PRITS and the Office for Cyber Incident Evaluation, by the reporting timelines established by the Information Security Incident Response Policy

(PRITS-POL-0002). This includes, but is not limited to, breaches involving AI systems, unauthorized access or use, or situations where AI-generated outputs caused harm or material errors.

19.2.4.5 Cybersecurity incidents involving AI systems must be reported promptly, no later than 48 hours, per statutory requirements.

19.3 Data Management and Privacy

19.3.1 Data Classification and Protection

19.3.1.1 All data used by or resulting from AI systems must be handled following the Data Classification Policy for the Government of Puerto Rico (PRITS-POL-0001). Agencies shall:

19.3.1.1.1 Identify and classify all datasets involved in an AI system, including training data, input data, output data, and any derived or stored data.

19.3.1.1.2 Prohibit the transfer or input of sensitive government data into external or public AI services unless explicitly authorized and protected by adequate safeguards (such as encryption, contractual obligations, and approved data handling agreements). For example, confidential or restricted data must not be uploaded to a third-party AI system without PRITS-approved security measures and legal review.

19.3.1.1.3 Apply security controls appropriate to the data classification level. For instance, restricted data used by an AI system must be encrypted both at rest and in transit, stored in approved secure cloud or on-premises environments, and access limited to authorized personnel,

following applicable security standards and regulations.

19.3.1.1.4 If AI systems are used to generate or process public information, ensure that any resulting public data releases comply with open data regulations and do not inadvertently expose non-public information.

19.3.1.2 Any AI system that ingests or generates images, video, or biometric templates (e.g., face, iris, gait) shall:

19.3.1.2.1 Treat raw and derived imagery as sensitive data.

19.3.1.2.2 Store imagery only in encrypted repositories approved by PRITS.

19.3.1.2.3 Apply dataset provenance tracking and tamper-evident logging.

19.3.1.2.4 Purge non-essential raw images after model training or inference, unless retention is legally required.

19.3.2 Privacy and Personal Data

19.3.2.1 Agencies shall adopt a privacy-by-design approach for AI systems that process personal data. At minimum:

19.3.2.1.1 Conduct a Privacy Impact Assessment (PIA) for any AI system handling personal data of citizens or employees. The PIA must document how personal information is collected, used, stored, and shared, and identify measures to mitigate associated privacy risks.

19.3.2.1.2 Ensure compliance with applicable privacy laws, including Puerto Rico's data protection requirements, federal statutes (e.g., HIPAA for health data), and any sector-specific privacy regulations.

19.3.2.1.3 Implement privacy-enhancing techniques where feasible, such as data minimization (limiting data collection to what is strictly necessary for the AI's purpose), anonymization or de-identification, and the aggregation of outputs to avoid re-identification of individuals.

19.3.2.1.4 Provide notice to users or the public, as appropriate, when AI systems process personal data. Such notice must inform individuals about the AI's role, the data being processed, and, where applicable, offer mechanisms to inquire or challenge AI-driven decisions.

19.3.2.1.5 Establish and enforce data retention and deletion policies for AI-related data. Personal data must not be retained longer than necessary for the authorized purpose. Retention schedules for AI training data, input/output data, and derived datasets containing personal information must be formally defined, approved, and aligned with Agency records management rules and any applicable regulatory requirements.

19.3.3 Data Quality and Integrity

19.3.3.1 Since the reliability and trustworthiness of AI system outputs depend on the quality of the underlying data, Agencies must ensure that:

19.3.3.1.1 Datasets are accurate, up-to-date, and to the extent possible, free from known errors, gaps, or biases. Before deploying an AI system, all training and configuration datasets shall be reviewed and audited for quality, completeness, and representativeness.

19.3.3.1.2 Procedures are in place to monitor for changes in the underlying data and to update or retrain AI models as needed. This is essential to prevent model drift or degradation in predictive accuracy over time.

19.3.3.1.3 Sensitive data used for AI training, fine-tuning, or operations is protected from unauthorized modification (integrity) and unauthorized access (confidentiality). Measures from the cybersecurity program (monitoring data access, using checksums or digital signatures for critical datasets, etc.) should be applied.

19.3.3.1.4 When AI systems use data from external providers or third-party datasets, such sources must be formally vetted and approved. Agencies must confirm that the data was collected lawfully and ethically, and is authorized for the intended use of AI. Preference shall be given to datasets that have documented provenance, well-defined collection methods, and identified limitations.

19.3.4 Confidentiality of AI Outputs

19.3.4.1 AI-generated outputs must undergo an assessment to determine whether they contain, derive, or could be used to infer sensitive information. Examples include reports, analytics, predictive models, summaries, or other generated content that could reveal classified, personally identifiable, or otherwise protected data.

19.3.4.2 When outputs are determined to be sensitive, Agencies shall apply the appropriate classification level and handling procedures consistent with applicable laws, regulations, and internal security policies. This includes classifying outputs appropriately, storing them in approved

secure environments, and restricting dissemination to authorized recipients only.

19.3.4.3 Agencies must implement safeguards to prevent unintentional disclosure of sensitive information in AI outputs, including but not limited to:

19.3.4.3.1 Implementing output filtering, redaction, or masking for sensitive details before dissemination.

19.3.4.3.2 Reviewing AI outputs generated from sensitive datasets to ensure they do not indirectly reveal confidential information through inference or correlation.

19.3.4.3.3 Applying automated detection tools, where feasible, to flag and review potentially sensitive AI-generated content before release.

19.4 Security of AI Systems (Cybersecurity Integration)

19.4.1 Secure Design and Development:

19.4.1.1 All AI systems, whether developed in-house, procured from vendors, or integrated from open-source solutions, must follow secure software development life cycle (SSDLC) principles adapted for AI. This includes:

19.4.1.1.1 *Security-by-design*
Define and document security requirements during the design phase, including access control, authentication, encryption, audit logging, and resilience against AI-specific threats.

19.4.1.1.2 *Pre-deployment and periodic testing*
Conduct vulnerability assessments, penetration testing, and AI-specific threat simulations (e.g., adversarial input manipulation, data poisoning, prompt

injection) before release and at defined intervals.

19.4.1.1.3 *Secure code and model standards*
Ensure AI software and model development follow recognized secure coding practices. Open-source components or pre-trained models must be verified for integrity, provenance, and reputation before use, and updated regularly to address vulnerabilities.

19.4.1.1.4 *Environment hardening*
Secure training, validation, and production environments to prevent unauthorized access or tampering. For sensitive AI systems, implement network isolation, strict identity and access controls, and continuous monitoring of system activity to ensure security.

19.4.2 Access Control and Authentication

19.4.2.1 Access to AI system resources, including management interfaces, model source code, training data repositories, and operational datasets, must be restricted to authorized personnel.

19.4.2.2 Multi-factor authentication (MFA) shall be implemented for all user accounts.

19.4.2.3 The Principle of Least Privilege (PoLP) and Role-Based Access Controls (RBAC) must be enforced to ensure users, and processes have only the minimum necessary privileges.

19.4.2.4 Consistent with the Zero Trust Architecture model, every access attempt must be treated as potentially untrusted until verified through continuous identity validation, context-aware authentication, and activity monitoring.

19.4.2.5 AI systems shall not be granted privileged or administrative access to government IT infrastructure (including but not limited to system configurations, creation/modification/deletion of user accounts, or elevated API permissions). Such functions must remain under direct human control and be subject to established cybersecurity and access management policies.

19.4.2.6 Access logs for AI systems must be maintained, reviewed regularly, and protected from tampering to support auditing, incident detection, and forensic investigations.

19.4.3 Monitoring and Incident Response

19.4.3.1 All AI systems and their hosting environments shall be integrated into the Agency's cybersecurity monitoring and incident response processes. This includes, at a minimum:

19.4.3.1.1 Logging of AI system activities, including system events, decision outputs (where technically feasible), and access to data. Logs shall be maintained following retention policies, protected against tampering, and continuously monitored for indicators of compromise or abnormal activity.

19.4.3.1.2 Implementation of Security Information and Event Management (SIEM) or equivalent tools to detect suspicious activities targeting AI systems, such as unauthorized access attempts, exploitation of API endpoints, or atypical input patterns suggestive of adversarial attack.

19.4.3.1.3 In the event of a confirmed or suspected security incident involving an AI system, such as compromise of training datasets, unauthorized alteration of model parameters, or external manipulation of outputs, Agencies must respond per their incident response plan.

All incidents must be reported to the Office for Cyber Incident Evaluation within the timeframes established by applicable regulations, including the Information Security Incident Response Policy (PRITS-POL-0002).

19.4.3.1.4 Business Continuity and Disaster Recovery (BC/DR) plans shall explicitly incorporate scenarios involving AI system failure, malfunction, or unavailability. Agencies shall establish and periodically test fallback procedures that enable continuation of services manually or via non-AI means until the affected system is restored to normal operations.

19.4.4 Third-Party and Supply Chain Security

19.4.4.1 Recognizing that many AI solutions rely on third-party components, including cloud-based AI services, pre-trained models, and software libraries, Agencies shall implement robust supply chain risk management measures to prevent compromise through external dependencies. This includes:

19.4.4.1.1 Using only approved trustworthy third-party AI services, components, or platforms that meet the Government's security and compliance standards. When vendor compliance is uncertain, Agencies shall consult PRITS for verification before proceeding with procurement or integration.

19.4.4.1.2 Embedding explicit security and compliance requirements in all procurement contracts and service-level agreements (SLAs), as outlined in Section 19.6 and Appendix II. For cloud-based AI systems, Agencies shall require that providers have adequate certifications, such

as FedRAMP or equivalent, and that they formally commit to safeguarding government data in transit, at rest, and in use.

19.4.4.1.3 Maintaining continuous awareness of security advisories, version updates, and patches issued by third-party providers. Agencies must evaluate and apply such updates promptly to mitigate vulnerabilities.

19.4.4.1.4 When leveraging open-source AI models, datasets, or libraries, Agencies shall conduct formal risk assessments that evaluate factors such as the trustworthiness of the development community, frequency of updates, and potential for malicious tampering (e.g., data poisoning of repositories). Technical controls shall be implemented to validate the authenticity and integrity of all open-source components before deployment in production environments.

19.4.5 Integration with NIST CSF 2.0

19.4.5.1 Agencies shall align the security management of AI systems with the NIST Cybersecurity Framework (CSF) 2.0 to ensure a comprehensive, structured, and repeatable approach to AI cybersecurity governance. Alignment shall encompass the following functions:

19.4.5.1.1 *Identify*
Maintain an up-to-date inventory of AI assets, datasets, and related infrastructure. Document their business purpose, associated risks, and criticality classifications to guide risk-informed decision-making.

19.4.5.1.2 *Protect*

Implement and maintain safeguards, including access control, encryption, data minimization, and role-based permissions, to ensure continuity of AI-enabled services and the protection of associated data.

19.4.5.1.3 *Detect*

Develop and maintain capabilities to identify the occurrence of cybersecurity events that could affect AI systems, including anomaly detection on AI model outputs and automated monitoring of security logs.

19.4.5.1.4 *Respond*

Establish and maintain response plans to address AI-related incidents, including containment of corrupted models, revocation of compromised credentials, and temporary reversion to manual or alternative processes.

19.4.5.1.5 *Recover*

Implement and regularly test recovery plans to restore AI capabilities following a disruption, including retraining models from verified backup datasets, validating model accuracy, and reinstating full services.

19.4.5.1.6 *Govern*

Ensure that AI governance is embedded across all cybersecurity functions, with oversight mechanisms that assess and manage AI-specific risks, as described in Section 19.1.

19.5 Fairness, Ethics, and Transparency

19.5.1 Non-Discrimination

19.5.1.1 All AI systems deployed or used by the Government must operate in a manner that upholds the principles of equality and non-discrimination, avoiding outcomes that are unjust, biased, or discriminatory toward any individual or group.

19.5.1.2 Agencies shall assess AI systems for potential bias, especially for high-risk applications, particularly those influencing employment, benefits, law enforcement, or other determinations with significant individual impact.

19.5.1.3 When bias is identified (e.g., consistent underperformance for specific demographics), Agencies must take corrective measures such as retraining models with more diverse data, adjusting algorithms to reduce bias amplification, or introducing human oversight to validate and, if necessary, override AI outputs.

19.5.1.4 All AI-driven decisions affecting rights, benefits, or opportunities must be reviewed for compliance with applicable civil rights legislation and principles of equal opportunity.

19.5.2 Transparency to the Public

19.5.2.1 Citizens have the right to be informed when engaging with AI-enabled government services or when subject to an AI-assisted decision. Agencies shall provide clear and accessible disclosures in public-facing systems that employ AI technologies (e.g., “This service is supported by an AI assistant”).

19.5.2.2 If AI contributes to an eligibility or compliance decision affecting a person, the Agency shall, where appropriate, inform the individual that an AI tool was used and provide an avenue for inquiry or appeal with human intervention.

19.5.2.3 Where AI contributed to determinations of eligibility, compliance, or enforcement, Agencies shall, when appropriate:

19.5.2.3.1 Inform affected individuals of AI involvement in the decision process.

19.5.2.3.2 Offer mechanisms for inquiry, clarification, or appeal with access to human review.

19.5.3 Explainability

19.5.3.1 For high-risk AI systems, Agencies must prioritize explainability to ensure operators can interpret and convey the rationale behind AI outputs to affected individuals and oversight entities. This might involve:

19.5.3.1.1 Employing inherently interpretable models.

19.5.3.1.2 Generating post-hoc explanation reports for complex algorithms.

19.5.3.1.3 Producing plain language summaries when decisions have a significant impact (e.g., denial of service application).

19.5.3.2 Black-box AI systems—where inputs and outputs are visible but the decision-making process is not transparent—shall not be used in contexts with legal, financial, or civil consequences unless sufficient explainability mechanisms are in place. Preference shall be given to models that can reveal the decision factors, weighting, or rules applied.

19.5.4 Human Oversight

19.5.4.1 AI tools must support, not replace, human judgment in government decision-making. Agencies shall maintain human oversight mechanisms, including pre-decisional human review for critical determinations or post-decisional appeal mechanisms that allow for human intervention.

19.5.4.2 Employees shall be trained to critically evaluate AI outputs, particularly when anomalies are detected or when decisions bear significant legal or personal consequences.

19.5.5 Ethical Use and Purpose

19.5.5.1 AI deployments must align with the public interest and constitutional protections. AI shall not be used in ways that violate fundamental rights, intrude upon reasonable expectations of privacy without lawful authorization, or otherwise contravene the governmental ethical principles.

19.5.5.2 AI use cases that are novel, sensitive, or ethically contentious (e.g., facial recognition in public spaces) must undergo a higher-level review by PRITS for compliance with applicable technology governance policies.

19.5.5.3 Each Agency shall be responsible for conducting its own legal and ethical assessment, including the evaluation of potential impacts on constitutional rights and privacy, in coordination with its legal counsel or other competent authorities, as applicable. Consultation with ethics committees or qualified experts is strongly encouraged for sensitive implementations.

19.6 Procurement and Third-Party AI Services

19.6.1 PRITS Approval in Procurement

19.6.1.1 As mandated by Act 75-2019, all Executive Branch Agencies' procurements involving AI systems or services, whether software, cloud AI platforms, data services, consulting, or pre-trained models, shall be coordinated with PRITS to ensure compliance with best practices in enterprise architecture, security and privacy requirements, and this AI Policy and related regulatory frameworks.

19.6.1.2 Entities excluded from the scope of Act 75-2019, but subject to Act 40-2024, shall also ensure compliance with this AI Policy and applicable security, privacy, and governance requirements, coordinating with PRITS as necessary to address AI-specific risks.

19.6.1.3 No contract for an AI product or service shall be carried out without documented PRITS confirmation that the procurement aligns with applicable policies. This includes obtaining explicit written clearance from PRITS before finalizing any AI-related contract.

19.6.2 AI Vendor Requirements

19.6.2.1 Contracts with AI vendors must include stringent and enforceable clauses to protect the government's interests. These provisions shall be consistent with applicable laws, regulations, and policies (for detailed guidance, reference Appendix II. At a minimum, the following areas must be addressed:

19.6.2.1.1 *Security and Privacy Standards*
Vendors must comply with all security and privacy requirements established by the Government of Puerto Rico, including but not limited to applicable PRITS security policies, contractual security controls specific to the project, and recognized industry standards (e.g., NIST CSF, ISO/IEC 27001). Vendors must:

- Implement technical safeguards, including encryption in transit and at rest, secure software development practices, and rigorous access controls.
- Ensure that personal data is collected, processed, stored, and transmitted in strict compliance with applicable privacy laws and regulations.
- Commit to strict confidentiality of all government data, prohibiting any unauthorized disclosure or secondary use.

19.6.2.1.2 *Data Sourcing and Ownership*

Contracts shall clearly state that:

- All government data, and any derivative models, analytics, or insights generated therefrom, are and shall remain the property of the Government of Puerto Rico.
- Vendors are prohibited from using government data for purposes outside the contract's scope without prior written authorization.
- Vendors must disclose any third-party data, datasets, or pre-trained models incorporated into the AI system. They must guarantee that these have been lawfully acquired and that the Government holds the right to use them.
- Upon contract termination or expiration, vendors must return or delete all government data and certify such action in writing.

19.6.2.1.3 *Performance and Quality*

Contracts must define objective and measurable performance metrics for the AI system, such as:

- Accuracy thresholds, latency and response time benchmarks, and availability requirements.
- Obligations for the vendor to promptly remediate defects, vulnerabilities, algorithmic biases, or functional deficiencies identified during the contract period.
- Commitments to maintain system performance over the full lifecycle of the AI deployment.

19.6.2.1.4 *Evaluation and Testing Rights*

The Government shall reserve the right to independently test, audit, and evaluate the AI system at any time during the contract term. Vendors must:

- Provide reasonable access to system components, documentation, and testing interfaces.
- Not impose contractual limitations that render the AI system a "black box" inaccessible to independent validation.
- Facilitate government-led evaluations for security, bias, fairness, and performance assurance.

19.6.2.1.5 *Transparency and Documentation*

Vendors shall deliver complete and accurate documentation sufficient for the Government to operate, maintain, and govern the AI system. This includes:

- System architecture, algorithms, or model descriptions, training data characteristics (excluding proprietary raw data, where appropriate), and identified limitations or risks.
- Operational manuals for end users and administrators.
- Change logs and version histories for all system updates.

19.6.2.1.6 *Safety and Ethical Compliance*

Contracts shall mandate that AI systems are designed, developed, and deployed in alignment with ethical AI principles, including:

- Avoidance of discriminatory bias and harmful outcomes.

- Prohibition of malicious code, hidden functionalities, or unauthorized surveillance capabilities.
- Compliance with ethical AI practices and relevant international best practices.

19.6.2.1.7 *Incident and Error Reporting*

Vendor must promptly report to the contracting Agency and PRITS:

- Any suspected or confirmed security incident, data breach, or vulnerability affecting the AI system.
- Any significant error or anomalous AI behavior that could result in harm, operational disruption, or incorrect decisions.
- Corrective actions taken and timelines for remediation.

19.6.2.1.8 *Indemnification and Liability*

Contracts shall include provisions that:

- Require the vendors to indemnify the Government of Puerto Rico for any losses or claims, such as those related to intellectual property or privacy violations, caused by the AI system or data derived from it, or the vendor's negligence.
- Establish the vendor's liability for damages caused by failure to meet contractual obligations, security standards, or legal requirements.

19.6.2.1.9 *Legal and Regulatory Compliance*

Vendors shall:

- Comply with Law 40-2024 and all applicable laws.

- Cooperate with government audits, compliance reviews, and oversight activities.

19.6.2.1.10 *Termination and Remediation*

Contracts must provide precise mechanisms for suspension or termination in cases where:

- The vendor fails to meet the required standards or legal obligations.
- The AI system poses unforeseen, high-risk threats to security, privacy, or public welfare.

Upon termination, vendors shall:

- Ensure orderly return or secure deletion of all government data.
- Provide transition assistance to maintain service continuity or enable replacement solutions.

19.6.3 Open-Source or Free AI Tools

19.6.3.1 In some instances, Agencies may consider the use of open-source AI software or free, online, or publicly accessible AI services. While these solutions can offer flexibility, cost efficiency, and community-driven innovation, they remain subject to the provisions of this Policy. Agencies must:

19.6.3.1.1 Conduct a thorough evaluation of any open-source AI tools before adoption. This includes:

- Review for known vulnerabilities, backdoors, or malicious code.
- Ensure that the tool's open source license is compatible with governmental use, including any restrictions on modification, redistribution, or commercial application.

19.6.3.1.2 Not upload, process, or expose confidential or sensitive data to any free or publicly accessible AI service without written approval, as outlined in Section 19.3.1. The fact that a tool is free does not mean it is safe, privacy-compliant, or suitable for handling government data.

19.6.3.1.3 Register all open-source AI components in the AI systems inventory and manage them as any other third-party software components. This includes:

- Monitoring for updates, patches, and security advisories from the developer community.
- Implementing timely updates to address vulnerabilities.
- Documenting version history and dependencies for audit readiness.

19.6.4 Collaboration and Knowledge Sharing

19.6.4.1 PRITS shall serve as a facilitator for the sharing of best practices, policy interpretations, and lessons learned in AI adoption across Agencies. When possible, Agencies shall:

19.6.4.1.1 Leverage centralized purchasing agreements or contracts negotiated by PRITS for AI services, to ensure consistency and maximize collective bargaining power.

19.6.4.1.2 Share case studies, challenges, and risk mitigation strategies through PRITS-led forums, interagency working groups, or digital collaboration platforms.

19.7 Training and Workforce Development

19.7.1 Training on AI and this Policy

19.7.1.1 All Agencies shall ensure that personnel with AI-related responsibilities receive role-appropriate training on the responsible use of AI and compliance with this Policy. This includes:

19.7.1.1.1 Executives and Managers
Shall understand the strategic implications of AI, the fundamentals of risk management, and their accountability under this Policy.

19.7.1.1.2 Technical Teams (Developers, Data Scientists, IT staff)
Shall be trained in secure AI development practices, adherence to NIST AI RMF principles, data privacy obligations, and the implementation of fairness, transparency, and security controls.

19.7.1.1.3 End-Users or Decision Makers
(For example, staff using AI-generated recommendations) shall be trained on how to interpret AI outputs, apply human oversight, and detect potential issues such as inconsistent, biased, or erroneous outputs.

19.7.1.1.4 Security and Audit Personnel
Shall receive specialized training in AI system assurance, including methods for auditing AI systems, monitoring for anomalies, and responding to AI-related security incidents.

19.7.1.2 Consistent with Law 40-2024, which mandates cybersecurity training for all government employees, this Policy extends the training scope to include AI-specific topics. PRITS will support Agencies by developing or recommending training curricula focusing on AI risk

management, ethical and lawful AI use, and AI security controls.

19.7.2 Continuous Education

19.7.2.1 Given the rapid evolution of AI capabilities and associated threats, AI competency must be treated as an ongoing requirement. Refresher courses or updated training modules shall be delivered at least annually, or sooner if prompted by significant policy changes, technological advances, or new threat vectors.

19.7.2.2 PRITS may issue updated guidance, host workshops, or provide targeted briefings on AI developments, such as regulatory frameworks, advances in generative AI, or novel risk categories, to keep Agency personnel informed.

19.7.3 AI Talent and Expertise

19.7.3.1 Agencies are encouraged to develop internal AI expertise by:

- Designating AI specialists or “AI champions" within the IT department to oversee compliance and optimize AI usage.
- Establishing partnerships with academic, research centers, and training providers to develop a skilled AI workforce pipeline.
- Enable knowledgeable staff to engage with vendors, critically assess AI outputs, and contribute to governance and policy refinement.

19.8 Monitoring, Compliance, and Reporting

19.8.1 Agency Self-Assessment

19.8.1.1 Agencies must conduct periodic self-assessments to evaluate compliance with this Policy. Self-assessments

may be integrated into existing internal control reviews, IT governance processes, or formal internal audits.

19.8.1.2 At a minimum, self-assessments shall address the following:

19.8.1.2.1 *AI Inventory*
Has the Agency identified and documented all AI systems in use or under development?

19.8.1.2.2 *Risk Assessment*
Are risk classifications and assessments current for all AI systems?

19.8.1.2.3 *PRITS Approval*
Have all AI uses been reviewed and approved by PRITS as required?

19.8.1.2.4 *Training*
Has the required training of personnel been completed and documented?

19.8.1.2.5 *Incident Response*
Do the Agency's incident response plans address AI-related scenarios, including potential misuse, system failures, and bias-related harms?

19.8.2 Annual Compliance Reporting

19.8.2.1 Each Agency shall submit an Annual AI Compliance Report to PRITS (Office of the CISO) by January 31 of each year (see Appendix III for template and instructions). At a minimum, the annual report will include:

19.8.2.1.1 *Inventory of AI Systems*

An updated list of all AI systems in use or under development, including the purpose, core functionality, and operational status of each system.

19.8.2.1.2 *Risk Classifications*
The current risk classification (low, moderate, high) of each AI system, noting any changes from the previous year.

19.8.2.1.3 *Summary of Risk Assessments*
Confirmation that each AI systems has an up-to-date risk assessment, including the date of last review, residual risks, and any identified vulnerabilities.

19.8.2.1.4 *Incidents and Issues*
A summary of any AI-related incidents, failures, misuse, or near-misses that occurred during the reporting period, along with corrective actions taken.

19.8.2.1.5 *Exemptions*
A list of any active exemption approvals, including expiration or renewal status.

19.8.2.1.6 *Training*
Confirmation of compliance with AI training requirements, including the number of staff trained and any notable training initiatives.

19.8.2.1.7 *Planned AI Projects*
A forecast of any significant AI initiatives planned for the next year to facilitate PRITS oversight, resource allocation, and inter-agency collaboration.

19.8.2.1.8 *Improvements and Challenges*
A brief analysis of challenges faced in implementing this Policy (e.g., technical limitations, bias testing constraints, workforce capacity) and any needs for additional guidance, tools, or resources. Agencies may also share success stories or best practices that could benefit other Agencies.

19.8.2.2 PRITS will review these annual reports to evaluate compliance trends, identify common issues, and inform policy refinement. Summaries or aggregated findings might be shared with the Governor's Office or other oversight bodies as appropriate. These reports contribute to the Government's broader cybersecurity maturity metrics and accountability framework.

19.8.3 Audits and Reviews

19.8.3.1 PRITS, in collaboration with the Office of the Comptroller or other audit authorities, may conduct periodic audits of Agencies' AI systems to verify compliance with this Policy. Agencies shall fully cooperate, including granting timely access to documentation, systems, and interviews with personnel.

19.8.3.2 Audits may include, but are not limited to:

19.8.3.2.1 Review of system architecture and configuration documentation.

19.8.3.2.2 Validation of AI risk assessments and mitigation plans.

19.8.3.2.3 Testing of AI system security and data protection measures.

19.8.3.2.4 Verification of training completion records.

19.8.3.2.5 Interviews with operational, technical, and oversight staff.

19.8.3.3 Findings of non-compliance shall require corrective actions within defined timelines.

19.8.3.4 Serious or repeated non-compliance may be escalated under Act 40-2024 enforcement provisions and could result in penalties or other corrective measures.

19.8.4 Continuous Policy Improvement

19.8.4.1 PRITS shall review this Policy at least once annually, or more frequently if needed to address new technological advancements, emerging risks, or changes in legal or regulatory requirements.

19.8.4.2 Updates to this Policy will be driven by:

19.8.4.2.1 Emerging national and international best practices, including future revisions to NIST frameworks, ISO standards, and other relevant guidance.

19.8.4.2.2 Lessons learned from Agency reports, audit findings, and incident investigations.

19.8.4.2.3 Stakeholder feedback from Agencies, industry partners, academia, and civil society.

19.8.4.3 Agencies and stakeholders are welcome to provide feedback or suggestions to PRITS for improving the Policy. The goal is to keep the Policy current, effective, and responsive to innovation, ensuring that Puerto Rico's approach to AI remains aligned with best practices (such as future updates to NIST's frameworks) and supports safe adoption of beneficial new AI capabilities.

19.8.5 Enforcement

19.8.5.1 Compliance with this Policy is mandatory for all Agencies, contractors, and third parties engaged in the development, procurement, deployment, or operation of AI systems on behalf of the Government of Puerto Rico.

19.8.5.2 Where non-compliance is identified, PRITS may take one or more of the following actions, as appropriate to the severity and nature of the violation:

19.8.5.2.1 Issue a notice of non-compliance and require immediate remedial actions, including suspension or deactivation of the AI system until full compliance is achieved.

19.8.5.2.2 Impose fines, administrative sanctions, or disciplinary measures against responsible parties, in accordance with Law No 40-2024.

19.8.5.2.3 Take legal action or enforce contractual penalties, up to and including contract termination.

19.8.5.2.4 Refer cases involving willful or reckless misconduct to relevant oversight bodies or law enforcement authorities for investigation and potential legal action.

19.8.5.3 Persistent violations, particularly those demonstrating disregard for statutory obligations or PRITS directives, may be escalated to oversight entities for further action.

20. Responsibilities

To ensure effective implementation and accountability under this Policy, the following roles and responsibilities are defined:

20.1 Agency Head (Secretaries/Directors/Majors)

20.1.1 Hold ultimate responsibility for AI use within the Agency.

20.1.2 Ensure Agency-wide compliance with this Policy by prioritizing adherence, allocating sufficient resources (personnel, training, budget, technical tools), and holding subordinates accountable for implementation.

20.1.3 Review and approve AI risk assessments, including final sign-off, before submission to PRITS.

20.1.4 Evaluate approval requests and sign the Annual AI Compliance Report.

20.2 Agency Chief Information Officer (CIO)

20.2.1 Serve as the primary coordinators for this Policy's implementation within the Agency.
20.2.2 Maintain an up-to-date inventory of AI systems and ensure AI governance is integrated into IT processes.
20.2.3 Oversee the timely submission of required documentation to PRITS.
20.2.4 Ensure technical teams conduct comprehensive AI risk assessments and implement appropriate technical and procedural controls.
20.2.5 Liaise with the PRITS regularly, monitor updates to AI technical standards, and disseminate information within the Agency.

20.3 Agency Information Security Officer

20.3.1 Lead the cybersecurity oversight of AI systems within the Agency.
20.3.2 Provide input on AI risk assessments, particularly in identifying threats, vulnerabilities, and security control requirements.
20.3.3 Ensure compliance with security policies, including access control, monitoring, and incident response procedures.
20.3.4 Facilitate and collaborate in the investigation and remediation of AI-related security incidents.
20.3.5 Agencies lacking a dedicated Information Security Officer shall delegate these responsibilities to the CIO or to another formally designated security authority.

20.4 Chief Information Security Officer (CISO) and PRITS AI Policy Team

20.4.1 Provide government-wide oversight for this Policy's implementation and enforcement.
20.4.2 Review Agency submissions (e.g., risk assessments, exemption requests, annual compliance reports) and approve, reject, or request remediation as appropriate.

20.4.3 Maintain and update the Policy, issue supplemental guidance, and provide standardized templates, tools, and training to Agencies.

20.4.4 Monitor government-wide AI compliance metrics and report to the PRITS Executive Director and relevant oversight bodies.

20.5 PRITS Executive Director

20.5.1 Serve as the highest-level sponsor for AI policy and governance initiatives.

20.5.2 Ensure PRITS dedicates sufficient resources for AI oversight and integrates AI governance into broader digital transformation strategies.

20.5.3 Make final determinations on high-sensitivity AI uses or intergovernmental AI initiatives.

20.5.4 May convene inter-agency councils or working groups on AI governance.

20.6 Agency Legal Counsel

20.6.1 Advise Agency leadership on legal implications of AI adoption and use, including compliance with privacy, civil rights, and contractual obligations.

20.6.2 Review procurement contracts and Memoranda of Understanding (MOUs) for AI solutions to ensure the inclusion of required contractual protections (see Appendix II).

20.6.3 Participate in reviewing AI projects for legal risk and on exemption requests, including liability implications.

20.7 Data Officer

20.7.1 Ensure AI systems comply with all applicable privacy requirements.

20.7.2 Conduct and document Privacy Impact Assessments (PIAs) where personal data is involved

20.7.3 Monitor AI uses for compliance with consent agreements and privacy laws.

20.7.4 Manage and respond to data subject inquiries or complaints related to AI processing.

20.8 Agency Program Managers/Stakeholders

20.8.1 Initiate and oversee AI projects within their respective operational domains.
20.8.2 Ensure early coordination with CIOs and Information Security Officers for compliance and risk assessment.
20.8.3 Monitor AI system performance and ethical impact, ensuring outputs align with program objectives.
20.8.4 Report incidents, anomalies, or concerns related to the AI use.

20.9 All Employees and Contractors

20.9.1 Use AI systems following approved procedures and training.
20.9.2 Remain alert to AI malfunctions, security concerns, or unethical outputs, and report such events promptly.
20.9.3 Maintain confidentiality, integrity, and security of AI data and systems.
20.9.4 Refrain from any attempt to manipulate, bypass, or misuse AI functionalities.

20.10 Internal and External Auditors

20.10.1 Independently verify compliance with this Policy through scheduled or ad-hoc audits.
20.10.2 Provide findings and recommendations to strengthen AI governance and controls.

20.11 Office for Cyber Incident Evaluation

20.11.1 Investigate AI-related cybersecurity incidents and coordinate with affected Agencies to ensure proper remediation and reporting.
20.11.2 Collaborate in post-incident analysis to improve AI risk prevention strategies.

21. Penalties

PRITS may impose fines on Agencies that fail to comply with the provisions of Law 40-2024. Any unauthorized disclosure, unauthorized modification, or loss of data, when there is obstruction, negligence, bad faith, recklessness, or willful refusal in the handling or reporting of a cyberattack, shall result in sanctions, fines, and in some cases, dismissal, contract termination, disqualification from contracting, restitution of funds, and other applicable sanctions under Law 40-2024.

22. Exemption Request

If compliance with this Policy proves unfeasible or technically impossible, or if an exemption is required, the system owner or administrator must submit a formal written request to the CISO and the CIIO. This request may be channeled through the Service Desk (https://support.prits.pr.gov/) or, failing that, to the email address support@prits.pr.gov, clearly indicating the subject (e.g., 'AI Policy Exemption Request – [Agency] – [AI System Name]').

23. Derogatory Clause

This Policy supersedes any prior directives, guidelines, or policies on AI use in the Government of Puerto Rico that conflict with its provisions, including PRITS Circular Letter 2023-002. Where earlier documents or communications provided recommendations, this Policy now provides binding requirements.

24. Effective Date

This Policy shall take effect immediately.

Within 30 days of its effective date, all Agencies must:

- Acknowledge receipt and understanding of this Policy through their Agency Head or CIO to PRITS.
- Submit an initial compliance status to PRITS, which must include:

- An updated inventory of all AI systems currently in use, in testing, or under development.
- A status assessment of each system against the requirements of this Policy.

For Agencies with AI initiatives already underway under previous guidance (i.e., Circular Letter 2023-002), existing projects must be reviewed and updated to ensure complete alignment with this Policy. PRITS may provide advice and technical assistance to facilitate the transition.

Future amendments, clarifications, or supplemental guidance will be communicated via official PRITS Circular Letters or formal Policy revisions. Agencies are responsible for staying informed of these updates and incorporating them promptly into their operations. PRITS may issue supplemental bulletins addressing emerging AI use cases, risks, or regulatory developments.

25. References

Identification Number	Title	Version
PRITS-POL-0001	Data Classification Policy for the Government of Puerto Rico	1.0
PRITS-POL-0002	Information Security Incident Response Policy	1.0
NIST AI 100-1	Artificial Intelligence Risk Management Framework (AI RMF)	1.0
NIST CSWP 29	NIST Cybersecurity Framework (CSF)	2.0

This space has been intentionally left blank.

26. Firmas de aprobación / *Approval Signatures*

Posición / ***Title***	**Acción realizada** / ***Action Taken***	**Nombre en letra de Molde** / ***Print Name***	**Firma** / ***Signature***	**Fecha** / ***Date***
Director Asociado, Soluciones / *Associate Director, Solutions*	Revisado por / *Revised by*	Víctor J. Filpo Román		29/sep/2025
Principal Oficial de Tecnología / *Chief Technology Officer*	Aprobado por / *Approved by*	Poincaré Díaz Peña		29/sep/2025

This space has been intentionally left blank.

27. Appendices

Appendix I: AI Risk Classification Matrix

The following AI Risk Classification Matrix establishes standard criteria for classifying AI systems as Low, Moderate, or High Risk. Agencies must use it when performing AI risk assessments per Section 19.2.3.

- Purpose
  Risk classification ensures that appropriate governance, oversight, and security controls are in place, proportional to the potential impact of each AI system.

- Risk Level Determination
  If an AI system meets one or more criteria in a higher category, the system shall be assigned the highest applicable risk level.

- Flexibility
  This matrix is illustrative and may be adapted by PRITS over time to address emerging technologies, sector-specific requirements, or changing regulatory standards.

Criteria	Low Risk AI	Moderate Risk AI	High Risk AI
Impact on Individuals / Community	Minimal or no direct impact on individuals. AI outputs do not make determinations affecting rights, services, or safety. Errors or failures cause only minor inconvenience or require routine manual correction.	Moderate impact. AI informs decisions that affect individuals or processes, but human review and override are in place. Errors may cause limited financial loss, short service delays, or moderate inconvenience, but do not result in life-threatening or irreversible harm.	High or significant impact. AI directly influences or determines outcomes affecting rights, safety, or critical access to services. Errors could cause denial of essential services, violations of rights, significant financial loss, physical injury, or other severe harm. Includes AI used in emergency response, law enforcement, healthcare triage, critical infrastructure, services to vulnerable populations, or

Criteria	Low Risk AI	Moderate Risk AI	High Risk AI
			image/biometric recognition (e.g., identification, authentication, or surveillance).
Data Sensitivity	Uses only public or internal data, non-sensitive data (e.g., anonymized statistics). No confidential or restricted data involved.	May process some personal data (PII) of moderate sensitivity (e.g., data that is not public but would cause limited harm if exposed), or small amounts of confidential information. Includes cases where highly sensitive data is used but fully anonymized.	Processes large-scale sensitive personal data (PII, health data, financial, biometric data) or restricted data. A breach or misuse would have severe consequences.
Function	Primarily advisory or informational with mandatory human verification. Automate simple, low-stakes tasks (e.g., categorizing public information, basic statistics, website analytics).	May automate a decision-making step, but with human review and override. AI plays an important role but is not the sole authority.	Operates autonomously or with limited human intervention due to scale or speed requirements. Even with human oversight, AI outputs strongly influence high-stakes outcomes.
Examples	AI suggesting website content categories; a chatbot providing general information from an FAQ database.	AI scanning of job applications (with final human review); AI-scheduling inspections based on risk (with staff adjustments allowed); predictive maintenance for equipment with no immediate safety risk.	AI determining eligibility for significant benefits or social services; AI-assisted medical diagnosis; predictive policing; AI controlling aspects of critical infrastructure (e.g., power grid management adjustments); biometric identification systems affecting access or freedom of movement; image-processing AI.

Unacceptable Risk AI

Certain AI applications are prohibited due to inherent incompatibility with legal, ethical, or constitutional principles. These include:

- AI uses that are expressly illegal under Puerto Rico or U.S. federal law.
- AI that violates constitutional protections, human rights, or established public policy.
- AI for social scoring of citizens, resulting in punitive or discriminatory measures.
- AI-enabled mass surveillance without lawful authority.
- AI claiming to predict criminal behavior based on race, ethnicity, or other protected traits, etc.

If an Agency encounters or is asked to implement such an AI system, it shall immediately halt the activity and report the matter to PRITS for investigation and enforcement.

Risk Level Documentation and Oversight

Agencies must maintain written justification for the assigned risk classification of each AI system, including the criteria met and the rationale for the determination.

This matrix shall be applied in conjunction with the detailed AI risk assessment procedures outlined in Section 19 of this Policy. PRITS may revise or expand this matrix to address emerging AI risks, sector-specific challenges, or new regulatory mandates.

This space has been intentionally left blank.

Appendix II: Checklist for AI Vendor Contract Clauses

When drafting or reviewing contracts, including purchase orders, service agreements, and MOUs, with vendors supplying AI systems or services, Agencies must ensure that the following topics are explicitly addressed through enforceable clauses or requirements.

This checklist is intended for use by Agency procurement officers, contract managers, and legal counsel during the contracting process to safeguard the Government of Puerto Rico's interests, ensure AI trustworthiness, and comply with applicable laws and policies.

1. *Scope of Work and Purpose*
   - Clearly define the AI system or service to be provided, including its intended functionality, scope, and purpose of use within the Agency.
   - Specify limitations on functionality to avoid unauthorized or unintended use cases.

2. *Data Ownership and Rights*
   - The Government of Puerto Rico retains full ownership of all data it provides to the vendor and all data outputs generated through the AI system.
   - The vendor shall have no secondary use rights to government data outside the contract's scope without prior written approval.
   - Include restrictions on data transfer to third parties or cross-border data movement without explicit authorization.

3. *Data Protection and Privacy*
   - Vendor must comply with all applicable Puerto Rico and federal data protection laws and regulations, including requirements for:
     - Encryption of data in transit and at rest.
     - Secure storage within approved jurisdictions.
     - Prohibition on mining, selling, or repurposing personal or sensitive data.
   - If processing sensitive data (e.g., health data under HIPAA), require applicable agreements such as a Business Associate Agreement or equivalent.
   - Vendor must certify compliant storage, deletion and bias-testing of any imagery or biometric data processed by the AI service.

4. *Confidentiality*

- Include a confidentiality clause prohibiting the disclosure of government data or sensitive AI system information to unauthorized parties.
- This obligation shall survive contract termination or expiration.

5. *Security Controls Compliance*
   - Vendor must meet minimum cybersecurity controls, referencing PRITS standards and recognized frameworks (e.g., NIST CSF or specific controls).
   - Cloud-based AI solutions must meet recognized security certifications (e.g., FedRAMP, ISO 27001).
   - Vendor must implement regular security patching and maintenance.
   - Agency reserves the right to conduct independent security assessments or require proof of compliance.

6. *Audit Rights*
   - The Agency shall have the right to audit or assess vendor performance and compliance, including AI security posture, data governance, and bias testing.
   - Include provisions for both on-site inspections and remote documentation review, with reasonable notice.

7. *Testing & Validation*
   - The vendor must cooperate in initial and ongoing testing of the AI system.
   - If significant bias, accuracy failures, or security vulnerabilities are identified, the vendor must remediate promptly.
   - Include a bias mitigation clause requiring the vendor to provide model documentation or evaluation results sufficient for independent bias assessment.

8. *Service Levels & Support*
   - Define service-level agreements (SLAs) for availability, performance, and response times.
   - Specify procedures for maintenance windows and updates, including advance notifications of changes that may change AI behavior.

9. *Change Management*
   - Require advance notification and, where appropriate, Agency approval for AI model or software changes that could affect outcomes.

- Prohibit deployment of major version changes without prior Agency testing and acceptance.

10. *Deliverables – Documentation*
    - The vendor must provide:
        - User and administrator manuals.
        - Model documentation (e.g., model cards, fact sheets) outlining functionality, limitations, and intended uses.
        - Data schema and integration guides.
    - Documentation must be kept up-to-date and reflect any material changes.

11. *Intellectual Property (IP) Rights*
    - Clarify IP ownership:
        - The vendor retain rights to pre-existing IP but grants the Agency the necessary licenses or perpetual access to use the procured version.
        - For custom-developed AI, grant the Agency rights to use, modify, and maintain the solution.
    - Avoid lock-in clauses that prevent system understanding or modification.

12. *Indemnification*
    - Vendor shall indemnify the Agency for:
        - IP infringement claims (i.e., someone sues, saying the AI infringes their patent or that the training data was used illegally).
        - Harm caused by the AI system output, vendor negligence, security breaches, or non-compliance.
    - Include specific indemnities for unlawful data use or training data acquisition.

13. *Liability Limitations*
    - Ensure liability caps are not set so low as to be meaningless.
    - Exclude critical breaches (e.g., confidentiality violations, IP infringement, major security incidents) from standard caps.

14. *Termination Rights*
    - Agency may terminate for cause (e.g., major breach, failure to deliver) or for convenience with notice.
    - Upon termination, the vendor must ensure uninterrupted access to Agency data or deliverables.

15. *Transition Assistance*
    - Require the vendor to provide orderly transition of services and data at contract end or termination, including:
        - Data return in usable format.
        - Model export (if applicable).
        - Assistance to successor vendors without service disruption.

16. *Laws, Policies, and Regulations*
    - Laws of the Commonwealth of Puerto Rico shall govern contracts.
    - Vendor must comply with:
        - Commonwealth of Puerto Rico Cybersecurity Act.
        - This AI Policy.
        - All applicable local and federal laws (cybersecurity, privacy, export controls, etc.).

17. *Penalties*
    - Where appropriate, include penalties for SLA breaches or security failures.

18. *Governance and Reporting*
    - Require periodic vendor reports on system performance, issues, updates, etc.
    - For multi-year contracts, require annual compliance attestation.

19. *Subcontractors*
    - Vendor remains responsible for subcontractors and must impose equivalent contractual obligations on them.
    - Agency approval for subcontractors handling sensitive data.

20. *Dispute Resolution*
    - Ensure dispute resolution terms do not force the Agency into unfavorable jurisdictions or arbitration forums (e.g., out-of-state).

Agencies should use this checklist when drafting, reviewing, or negotiating contracts related to AI. If any requirement is missing, procurement and legal teams must work to include equivalent protections.

Appendix III: Annual AI Compliance Report Templates

Each Agency's Annual AI Compliance Report (per Section 19.8.2) shall follow a standard structure to facilitate review. Below is a template outline and some sample tables that Agencies can fill out:

[Agency Name] – Annual AI Compliance Report – [Reporting Year]

1. Executive Summary
   Provide a brief overview (one paragraph) of the Agency's AI activities for the reporting year. Summarize the Agency's commitment to this Policy, major achievements, and any notable changes to AI usage or governance.

2. AI Inventory Summary
   Table: List of all AI systems currently in use, in development, or decommissioned during the reporting year. Include pilots and prototypes. For example:

AI System Name / ID	Description / Purpose	Status (Pilot / Production)	Risk Level (Low / Mod / High)	PRITS Approval Date	Notes (e.g., Vendor or In-house)
Example: "ChatGovBot"	Public-facing chatbot for general service inquiries.	Production (since Mar 2025)	Moderate	Approved by PRITS on Feb 1, 2025	Vendor: ABC Corp, Cloud-hosted
"DataScan AI"	Internal tool to assess form completeness	Pilot testing	Low	Pilot phase (info submitted Jun 2025)	In-house prototype
"RiskPredict"	Predictive analytics for identifying high-risk cases for inspection.	Development (RFP stage)	Expected High	(Planned submission)	Vendor provided, procurement in progress

- Include decommissioned systems in an appendix, noting retirement date and reason.
- All entries must reflect the most recent status as of the reporting date.

3. Risk Assessments and Mitigations
   - For each AI system (particularly moderate- and high-risk), provide:
     - Date of last risk assessment or review.
     - Summary of key risks identified, including misuse of AI, and mitigation measures in place.
     - Confirmation of continuous monitoring for high-risk systems.
     - Any changes in risk classification since the prior report, along with the rationale.

   - Example entry:
     - *ChatGovBot:*
       - Risk Assessment: Jan 2025.
       - Risks: possible inappropriate responses; low data privacy risk (no PII collected).
       - Mitigations: curated response library and human monitoring.
       - Risk level: Moderate.
       - Monitoring: Weekly chat log reviews implemented.

     - *RiskPredict:*
       - Risk assessment: preliminary assessment during Request for Proposals (RFP)
       - Risks: likely high due to direct impact on individuals. Risks include bias in decision-making and unauthorized access to predictive outputs.
       - Mitigations: human review of flagged cases, bias testing before deployment.
       - Monitoring: complete assessment scheduled for August 2025.

4. Policy Compliance Checklist
   - Self-attestation of compliance with provisions of this AI Policy. Agencies must provide explanations and planned remediation for any “No” responses.

Policy Area	Compliance (Yes / No / In Progress)	Notes
Governance: - AI Lead appointed	Yes	

Policy Area	Compliance (Yes / No / In Progress)	Notes
- Governance committee established - Internal policy/procedures updated and aligned with AI Policy	Yes In progress	Lead: María López, CIO
Training: % of relevant staff trained in AI Policy	92%	Annual refresher planned Q1 2026
Data Management: All AI data classified and protected per policy	Yes	N/A
Incident Response: AI-related incidents reported per protocol	Yes	No major incidents
Exemptions: Active approved exemptions listed	None	—

5. <u>AI-Related Incidents or Issues</u>
   - Document any AI-related events, including operational disruptions, bias findings, or security breaches:
     - Date, nature of incident, impact, and corrective actions taken.
     - If none occurred, explicitly state: “No AI-related incidents were recorded during the reporting period.”
   - Examples:
     - July 2025 – Chatbot outage (4 hours) due to vendor API failure. Contingency FAQ activated; vendor resolution confirmed via root cause analysis (RCA) review.
     - September 2025 – The language model underperformed in processing English-language forms. Vendor engaged to retrain the model with bilingual datasets.

6. <u>Governance Improvements and Initiatives</u>
   - Summarize measures taken to strengthen AI governance, such as:
     - Development of new internal AI review forms or workflows.
     - Participation in PRITS-led interagency AI working groups.
     - Implementation of bias testing workshops or AI ethics training.

- Integration of AI asset tracking into broader IT asset management systems.

7. Forward-Looking Plans
   - Outline anticipated AI activities for the next reporting year, including:
     - New AI projects in feasibility, procurement, or design stages.
     - Planned upgrades, expansions, or decommissions of AI systems.
     - Support or guidance requested from PRITS (e.g., training, hosting guidelines, procurement templates).

8. Additional Comments (Optional)
   - Include any notable success stories, challenges, or lessons learned. Positive impact metrics or cost savings from AI implementations may be highlighted here.

Sign-off

- Prepared by: [Name, Title] – usually the Agency CIO or designated AI Coordinator.
- Approved by: [Agency Head or Authorized Executive] – with date.

Attachments

- Relevant supporting documentation (risk assessment reports, privacy impact assessments, compliance matrices, incident reports) should be attached separately, with key points summarized in the main report.

Submission

Reports must be submitted to PRITS **by January 31** of each year, using the official submission method.

This space has been intentionally left blank.