USO ACEPTABLE DE HERRAMIENTAS DE INTELIGENCIA ARTIFICIAL (IA)

Por: Edwin Ríos – Socio de Risk Advisory Services – RSM Panamá

La IA coexistiendo en los ecosistemas empresariales

Las herramientas de inteligencia artificial (IA) son parte de un grupo de aplicaciones, programas (*software*) o plataformas que utilizan algoritmos avanzados y modelos de aprendizaje automático que aportan en la realización de tareas que normalmente requieren del factor e inteligencia humana. Estas herramientas, entre sus diversas bondades, tienen una amplia capacidad de poder analizar datos, automatizar procesos, tomar decisiones asistidas y/o mejorar la eficiencia en diversas áreas o procesos del negocio.

Si bien, existen distintas categorías de IA, las herramientas de inteligencia artificial en su contexto general pueden clasificarse de la siguiente forma y para distintos usos:

- Procesamiento de lenguaje natural (ej. *chatbots*, traductores automáticos)
- Biometría y visión por computadora (ej. reconocimiento facial, generación y análisis de imágenes)
- Automatización y productividad (ej. automatización robótica de procesos, generación de contenido, asistentes personales)
- Análisis de datos y predicción (ej. ciencia de datos, analítica y predicción financiera)
- Ciberseguridad y detección (ej. detección de anomalías, fraude, estafa, seguridad en redes)
- Desarrollo de *software* y soporte técnico (ej. depuración de código, mantenimiento predictivo en TI), entre otras.

A nivel individual, las IA de mayor uso se están enfocando principalmente en tecnologías relacionadas al procesamiento de lenguaje y/o GTP (*generative pre-trained transformer*) convirtiéndose ya en herramientas de uso cotidiano para el manejo y generación de textos con propósitos personales, laborales y/o profesionales de: *consulta; revisión, corrección o traducción de textos; generación de estructuras de documentos, resúmenes ejecutivos o material de contenido educativo; análisis de interpretación; automatización de respuestas; toma de decisiones asistidas; creación de imágenes, entre otras*.

En el entorno digital actual, el uso de herramientas de IA ofrece importantes beneficios para la eficiencia, la innovación y la toma de decisiones asistidas dentro de la Organización; no obstante, su implementación debe alinearse con principios de seguridad, ética y cumplimiento normativo para mitigar potenciales riesgos asociados a la seguridad, la ciberseguridad, la privacidad, la disponibilidad, la integridad de la información y el uso indebido de datos personales, corporativos o de activos de información.

Consideraciones y sanas prácticas para el uso aceptable de las IA

Como líderes organizacionales y a sabiendas que estas tecnologías se encuentran ya coexistiendo en los ecosistemas empresariales a nivel global, es de gran importancia definir directrices al más alto nivel que promuevan el uso aceptable, ético, seguro y responsable de estas herramientas con el propósito de asegurar a su vez, que se adhieran a diversos estándares en materia de ética y conducta, así como de seguridad, ciberseguridad, privacidad, entre otros componentes de un buen gobierno corporativo y de TI.

A continuación, algunas recomendaciones a ser tomadas en cuenta para el diseño de políticas de uso aceptable de herramientas de IA, sin limitarse:

- Definición clara del objetivo o propósito, así como del alcance y contexto general.
- Tipos de IA autorizadas (adquiridas bajo licenciamiento o gratuitas bajo entornos controlados; por ejemplo, si sólo se aceptan IA relacionadas al procesamiento de lenguaje o GTP).
- Uso adherido a un marco de ética y conducta, así como otros a nivel de seguridad, ciberseguridad, protección y privacidad de los datos e información.
- Respeto a la propiedad intelectual, asegurando que los datos y modelos utilizados cumplan con atributos de derecho de autor.
- Monitoreo de actualizaciones, versiones, soporte técnico y seguridad por parte del fabricante o proveedor, a través de los SLA, ya sean servicios "*On Premise*" o "*SaaS*"
- Responsabilidad y vigilancia permanente de posibles situaciones inusuales, alertas, comunicaciones sospechosas, eventos o incidentes asociados a manipulación de datos, ciber ataques, entre otros riesgos.

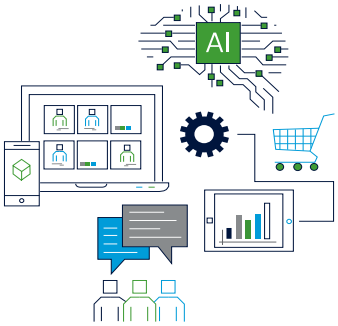
- Asociación con lineamientos, canales o medios para la gestión de incidentes de seguridad o ciberseguridad.
- Reservarse el derecho de restringir, eliminar y/o no autorizar el uso temporal o permanente de las IA cuando existan potenciales riesgos sobre los atributos de seguridad, disponibilidad, integridad, protección o privacidad de la información.
- Definición de límites en su uso para evitar desinformación, confusión o una posible decisión relevante sin la debida supervisión humana. En todo momento, debe prevalecer el sentido lógico, crítico, así como el juicio o criterio humano en el análisis, interpretación o toma de decisiones asistida por la IA.
- Incorporar mecanismos de seguimiento, monitoreo o de auditorías.
- Definición, límites y alcance de las prohibiciones en el uso de las IA.

El estándar IEC/ISO 27001:2022 establece una serie de controles técnicos que son un buen complemento para alinear y documentar lineamientos en una política de uso aceptable de herramientas de IA, por ejemplo, sobre: Política de Seguridad de la Información (A.5.1); Uso Aceptable de la Información y Otros Activos Asociados (A.5.10); Evaluación y Decisión sobre Eventos de la Información (A.5.25); Derechos de Propiedad Intelectual (A.5.32); Uso de Programas de Utilidad Privilegiados (A.8.18); Instalación de Software en Sistemas Operativos (A.8.19); Requisitos de Seguridad de la Aplicación (A.8.26).

Consideraciones y prohibiciones en el uso de las IA

Es importante considerar como parte del marco y políticas de uso aceptable de herramientas de IA, los límites, así como las prohibiciones sobre distintos escenarios (directos o indirectos). A continuación, algunas perspectivas sobre prohibiciones en su uso, sin limitarse:

- Procesamiento indebido de información confidencial, sensitiva o privada.
- Incorporación de nombres, información, archivos, datos de índole sensitiva o no sensitiva de la empresa, así como de clientes, proveedores o aliados estratégicos.
- Recopilación, análisis o compartir datos personales sin el debido consentimiento que pueda atentar contra leyes y regulaciones en materia de protección y privacidad de datos.
- Generación de contenido engañoso, noticias falsas o desinformación que pueda alterar o afectar la opinión o juicio de las personas.
- Actos y conductas indebidas o intimidantes que promuevan acoso en todas sus aristas.
- Generación de contenido con propósitos de sesgo o discriminación por raza, género, religión, política, orientación sexual, discapacidad u otros factores.
- Creación de contenido que promueva (en hecho o apariencia) el desarrollo de ataques cibernéticos o de cualquier índole similar (virus, malware, phishing, cracking, entre otros).
- Suplantación o alteración de identidad.
- Manipulación de datos comerciales, legales, financieros, así como para fines de información regulatoria o de auditorías, para fines de ocultar riesgos, cometer irregularidades o vulnerar mecanismos de control interno para cometer ilícitos (evasión fiscal, fraude, soborno-corrupción, lavado de dinero, blanqueo de capitales, entre otros).
- Inclusión de información ficticia para manipular los algoritmos y métodos de aprendizaje automático de las IA, para obtener información ventajosa o de carácter engañosa.
- Generación de textos basados en consultas provocativas o de índole apocalípticas, relacionadas al desplazamiento, odio o extinción del ser humano.
- Dar por hecho que todo resultado generado a través de una IA es absolutamente verdadero o idóneo para fines personales, laborales, profesionales o empresariales. No depender exclusivamente de la IA para decisiones críticas en distintos entornos.



Todas las consideraciones previamente establecidas, incorporando otras referencias técnicas, deben poder brindar guías generales y complementarias para la elaboración de una Política de Uso Aceptable de Herramientas de IA; no obstante, será responsabilidad de cada organización definir el propósito, contexto, alcance y lineamientos específicos que gobiernen el marco de usabilidad de estas tecnologías emergentes.

Edwin Ríos
Socio de Risk Advisory Services
erios@rsm.com.pa