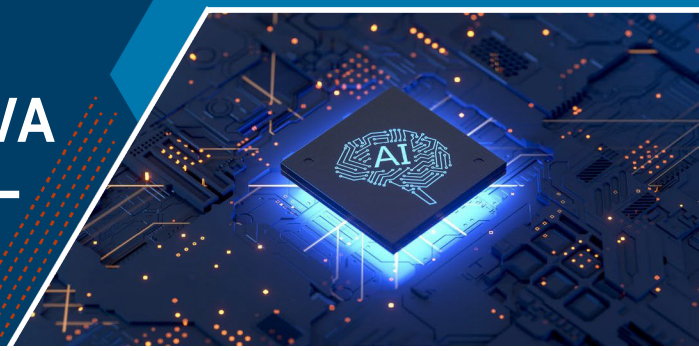
ENFOQUE BASADO EN RIESGOS: IA GENERATIVA Y EL CICLO ELECTORAL DE 2024



SINOPSIS

A medida que las capacidades habilitadas por la Inteligencia Artificial (IA) Generativa (Generative AI, por sus siglas en inglés) se adoptan más ampliamente, es importante que los funcionarios electorales entiendan cómo dichas capacidades podrían afectar la seguridad e integridad de la infraestructura electoral. Las capacidades de IA presentan oportunidades para aumentar la productividad, lo que podría mejorar tanto la seguridad como el manejo electoral. Sin embargo, estas capacidades también conllevan un aumento potencial de mayor daño, ya que los actores maliciosos, incluidos los actores de estados nacionales extranjeros y los ciberdelincuentes, podrían aprovechar estas mismas capacidades con fines nefarios. Para el ciclo electoral de 2024, es probable que las capacidades de IA generativa no introduzcan nuevos riesgos, pero pueden amplificar los riesgos en infraestructura electoral que actualmente existen. Los funcionarios electorales tienen el poder de mitigar dichos riesgos de cara al 2024, y muchas de estas mitigaciones son las mismas prácticas de seguridad recomendadas por los expertos durante años. Este paquete informativo proporciona una descripción general de las capacidades relevantes de la IA generativa, cómo estos actores maliciosos pueden utilizar estas capacidades para atacar la seguridad e integridad de la infraestructura electoral, y las acciones básicas de mitigación tendientes a contrarrestar los elevados riesgos asociados con las capacidades facilitadas por la IA generativa.

TAXONOMÍA DE IA GENERATIVA

La IA generativa es un tipo de aplicación informática. Esta utiliza modelos estadísticos que generalizan los patrones y estructuras de los datos existentes para reorganizar los datos existentes o crear contenido nuevo. Dicho contenido puede abarcar desde la escritura de código informático hasta la creación de nuevo texto y el desarrollo de medios sintéticos como archivos de vídeo, imagen y audio. Algunos ejemplos de cómo los actores maliciosos pueden utilizar las capacidades de IA generativa son:

Tipo de Contenido Sintético y Ejemplos	Tácticas Conocidas
Vídeo ▪ Texto a video ▪ Ultrafalso (*Deepfake*)[i]	▪ Un actor de un estado nacional extranjero utiliza software de texto a video para generar videos falsos de presentadores de noticias reales que informan sobre historias falsas para difundir desinformación como parte de una operación de influencia extranjera. [ii] ▪ Los ciberdelincuentes utilizan vídeos ultrafalsos (*deepfake*) de personas famosas para hacer que público caiga en estafas. [iii]
Imagen ▪ Texto a imagen ▪ Imagen alterada por IA	▪ Los actores de los estados nacionales extranjeros utilizan generadores de texto a imagen para crear imágenes falsas y engañosas con el fin de alterar la percepción pública de los hechos durante una crisis. [iv] ▪ Los actores de los estados nacionales extranjeros crean imágenes sintéticas en perfiles de cuentas falsas utilizadas en operaciones de influencia. [v] ▪ Los actores de los estados nacionales extranjeros alteran imágenes auténticas para apoyar las narrativas de operaciones de influencia. [vi]
Audio ▪ Conversión de texto a voz ▪ Clonación de voz	▪ Los actores de los estados nacionales extranjeros utilizan el audio generado por IA para hacerse pasar por empleados y obtener acceso a información confidencial o con el fin de convencer a las organizaciones para que tomen medidas específicas. [vii] ▪ Los ciberdelincuentes utilizan herramientas de IA generativa para clonar la voz de víctimas desprevenidas como parte de estafas de voz de IA o campañas de desinformación. [vii]
Texto ▪ Texto a texto (Modelos amplios de lenguaje)	▪ Los actores de los estados nacionales extranjeros utilizan texto generado por IA para fortalecer las operaciones encubiertas de influencia extranjera con contenido gramaticalmente correcto en inglés a menores costos marginales. [ix] ▪ Los ciberdelincuentes utilizan chatbots habilitados para IA generativa en sofisticadas campañas de ingeniería social y ciberestafa (*phishing*). [x]

POSIBLES OBJETIVOS RELACIONADOS CON LAS ELECCIONES DEL USO MALICIOSO DE LA IA

Los actores maliciosos pueden utilizar herramientas de IA generativa para reducir los costos y aumentar la escala de los incidentes cibernéticos y las operaciones de influencia extranjera. En el caso de los incidentes cibernéticos, los actores maliciosos pueden utilizar la IA generativa para ayudar a crear variedades de programas maliciosos (*malware*) que puedan evadir las defensas de ciberseguridad. También puede permitir ataques de denegación de servicio distribuido (DDoS) más efectivos, que pueden arruinar sitios web, incluidos los sitios web relacionados con las elecciones, al inundarlos con cantidades masivas de datos. Los actores maliciosos también pueden utilizar la IA generativa para ayudar a implementar técnicas de ciberestafa (*phishing*) e ingeniería social, generar audio real con la voz de cualquier individuo, crear imágenes falsas altamente realistas, impulsar perfiles de redes sociales falsificados y producir ultrafalsos (*deepfakes*) para respaldar las narrativas de las operaciones de influencia. Estas tácticas y tipos de ataques no son nuevos, pero la IA generativa permite a los actores maliciosos emplearlos de manera más rápida y sofisticada y por un costo mucho menor. Algunos ejemplos de posibles ataques maliciosos a las elecciones con IA son:

Procesos Electorales	▪ Los chatbots, las voces generadas por IA o los videos podrían usarse para difundir información falsa sobre la hora, la forma o el lugar de votación a través de mensajes de texto, correo electrónico, canales de redes sociales o medios impresos. ▪ El uso de contenidos y herramientas generados por IA podría aumentar la escala y la capacidad de persuasión de las operaciones de influencia extranjera y las campañas de desinformación dirigidas en contra de los procesos electorales. ▪ Las capacidades de IA podrían utilizarse para generar convincentes registros electorales falsos, tales como registros falsos de votos emitidos e imágenes de boletas electorales falsas.
Oficinas Electorales	▪ Las herramientas de clonación de voz podrían usarse para hacerse pasar como parte del personal de la oficina electoral y obtener acceso a información confidencial o de seguridad de la administración electoral. ▪ El uso de herramientas de IA podría permitir ataques de ciberestafa a lideres (*spearphishing)* de mejor calidad contra funcionarios o personal electoral y así obtener acceso a las redes de las oficinas electorales. ▪ Las herramientas de codificación de IA podrían utilizarse para desarrollar programas maliciosos (*malware*) e incluso malware mejorado que podría evadir más fácilmente los sistemas de detección. ▪ Los comandos (*scripts*) generados por IA y la clonación de voz podrían utilizarse para generar llamadas falsas de votantes con el fin de sobrecargar a los centros de llamadas.
Funcionarios electorales	▪ El contenido generado por IA, como los videos ultrafalsos (*deepfake*) comprometedores, podría ser usado para acosar, suplantar o deslegitimar a los funcionarios electorales. ▪ Las herramientas de IA podrían utilizarse para crear archivos de audio o vídeo haciéndose pasar por funcionarios electorales y difundiendo información incorrecta al público acerca de la seguridad o la integridad del proceso electoral. ▪ Las capacidades de IA podrían utilizarse para mejorar la agregación de datos de información pública para permitir ataques de doxing contra funcionarios electorales.
Proveedores electorales	▪ La tecnología generada por IA permite el uso sofisticado de técnicas de ciberestafa (*phishing*) e ingeniería social. ▪ Las herramientas generadas por IA podrían usarse para crear un video falso de un proveedor electoral haciendo una declaración falsa que ponga en duda la seguridad de las tecnologías electorales.

MEDIDAS DE MITIGACIÓN

Si bien los desarrollos y el uso nefario de las capacidades de IA generativa afectan el panorama de riesgos, los funcionarios electorales están bien preparados para mitigar eficazmente estas amenazas potenciales. Los funcionarios electorales ya están familiarizados con riesgos como la ciberestafa (*phishing*), las operaciones de influencia extranjera y la desinformación que puedan ser amplificados por la IA generativa. Muchas de las mejores medidas de mitigación para las amenazas generadas por IA son las mismas mejores prácticas de ciberseguridad que se han recomendado durante años y que los funcionarios electorales ya pueden haber implementado. Las medidas básicas de mitigación que las partes interesadas del subsector electoral pueden tomar para reducir el riesgo de las amenazas generativas habilitadas por IA incluyen:

Defenderse contra el uso de sofisticadas técnicas de phishing/ingeniería social habilitadas por IA

Implemente controles de seguridad:

- Establezca y aplique protocolos de ciberseguridad sólidos, como la autenticación multifactor (MFA, por sus siglas en inglés), especialmente la MFA resistente a la ciberestafa (*phishing*), como la autenticación *Fast Identity Online* (autenticación FIDO, por sus siglas en inglés), y el software de detección y respuesta de punto final.
- Adopte protocolos de seguridad de autenticación de correo electrónico, como la autenticación, la notificación y la conformidad de mensajes basados en el dominio (DMARC, por sus siglas en inglés), el marco de políticas del remitente (SPF, por sus siglas en inglés) y el correo identificado con claves de dominio (DKIM, por sus siglas en inglés) para protegerse mejor contra la suplantación de identidad del correo electrónico.[xi]
- Mejore las cuentas de redes sociales personales y organizacionales mediante la implementación de cambios como la aplicación de los controles de seguridad y privacidad más estrictos posibles, la desactivación o eliminación de perfiles que ya no estén en uso y la eliminación de cualquier información de identificación personal (PII, por sus siglas en inglés) de dichos perfiles.
- Incorpore los principios de seguridad de confianza nula para evitar el acceso no autorizado a los datos y servicios, y haga que la aplicación del control de acceso sea lo más específica y detallada posible. El Modelo de Madurez de Confianza Nula de CISA ayuda con esta transición al proporcionar una escala de recomendaciones para su implementación.

Limitar las oportunidades de suplantación de identidad y acoso habilitados por IA

Protéjase:

- Considere la posibilidad de hacer que las cuentas personales en las redes sociales sean privadas para que los actores maliciosos tengan menos acceso a su imagen y voz.
- Solicite regularmente que se elimine la información personal de los sitios web de registros públicos.
- Fortalezca las cuentas de redes sociales personales y organizacionales mediante la implementación de cambios como la aplicación de los controles de seguridad y privacidad más estrictos posibles, la desactivación o eliminación de perfiles que ya no estén en uso y la eliminación de cualquier PII de los perfiles de redes sociales.
- Denuncie los casos de acoso a las autoridades competentes.

Proteja la información confidencial:

- Antes de divulgar información confidencial, incluso a nivel interno del personal electoral, confirme las solicitudes a través de canales secundarios y considere implementar la verificación de identidad para las comunicaciones en tiempo real.
- Protéjase contra los intentos de suplantación de identidad virtual adoptando una frase de contraseña continua que solo el personal autorizado conozca, especialmente durante los períodos de votación activos.
- Eduque a los empleados en lo referente a la posibilidad de suplantación de identidad.

Cuente con que el contenido generado por IA exceda el ancho de banda para responder

Comuníquese de manera proactiva:

- Si es elegible, regístrese para obtener el dominio del sitio web .gov para indicar fácilmente su estado como organización gubernamental**.** Asegúrese de que todos los sitios web oficiales y las cuentas de redes sociales sean visibles y fácil acceso para el público.
- Establezca relaciones con los medios de comunicación locales y los líderes comunitarios; organice un equipo de voces confiables para amplificar la información exacta en caso de un incidente.
- Proporcione respuestas a preguntas comunes con antelación a través de múltiples canales de información (por ejemplo, salas de lectura en línea, sitios web, redes sociales, mensajes pregrabados y medios tradicionales).

Implemente controles técnicos para limitar las solicitudes no auténticas:

- Las herramientas de autenticación humana, como las identidades de confianza nula, los CAPTCHA y la verificación física, sirven para diferenciar a los usuarios humanos de los procesos automatizados. La aplicación de estas herramientas en los formularios y las solicitudes de registros abiertos, especialmente las aquellas entregadas en sitios web, puede reducir el volumen de solicitudes no auténticas que recibe una oficina. Revise continuamente las herramientas de autenticación para asegurarse de que son resistentes a las capacidades en constante evolución, para incluir capacidades habilitadas para IA, como el uso de herramientas que emplean tareas "reforzadas con IA" o el uso de acciones de software vinculadas al hardware, tales como la rotación de un teléfono.

Prepárese para las operaciones de influencia extranjera y la desinformación habilitadas por IA

Comunique prácticas electorales seguras de manera proactiva:

- Si es elegible, regístrese para obtener el dominio del sitio web .gov para indicar fácilmente su estado como
- organización gubernamental. Establezca relaciones con los medios de comunicación locales y los líderes comunitarios; construya un equipo de voces confiables para amplificar la información exacta en caso de un incidente.
- Considere la posibilidad de utilizar técnicas de autenticación activas, como marcas de agua, para mostrar que el contenido viene de usted de forma verificable y para identificar cuándo se modificaron los archivos después de aplicar las
- credenciales.
- Hable con los proveedores sobre la adopción de medidas de procedencia y autenticación para los registros electorales. Continúe generando confianza en sus prácticas de seguridad con comunicaciones públicas proactivas y de respuesta, como el desarrollo de puntos de conversación que transmitan hechos basados en evidencias reales referentes al por qué sus votantes deben tener confianza en la seguridad del proceso electoral.

Prepárese para responder:

- Capacite al personal sobre los procedimientos estándar para responder a los medios presuntamente manipulados y entienda los mecanismos para informar acerca de dicha actividad dentro de su organización.
- Considere formas para autenticar la información electoral publicada fuera de su organización (firmas, *hashing*, marcas de agua, etc.).

RECURSOS ADICIONALES:

Esta lista proporciona enlaces a recursos con información adicional actualmente disponible en inglés.

- Avoiding Social Engineering and Phishing Attacks | CISA
- AI Risk Management Framework | NIST
- EI-ISAC membership
- Disinformation Stops with You
- Tactics of Disinformation
- Election Disinformation Toolkit
- Election Cybersecurity Toolkit
- Cyber Incident Detection and Notification Planning Guide for Election Security
- Election Security Services
- Building Trust Through Secure Practices
- Contextualizing Deepfake Threats to Organizations
- No Downtime in Elections: A Guide to Mitigating Risks of Denial-of-Service

[i] Los ultrafalsos o *Deepfakes* son un tipo de contenido sintético, comúnmente generados mediante inteligencia artificial/aprendizaje automático (AI/ML, por sus siglas en inglés), que presentan videos, imágenes, audio o texto plausibles y realistas de eventos que nunca sucedieron.

[ii] How Deepfake Videos Are Used to Spread Disinformation - The New York Times (nytimes.com); Deepfake It Till You Make It (graphika.com)

[iii] Deepfake scams have arrived: Fake videos spread on Facebook, TikTok and Youtube (nbcnews.com)

[iv] Threat Actors are Interested in Generative AI, but Use Remains Limited | Mandiant

[v] How a fake network pushes pro-China propaganda (bbc.com); Facebook finds disinformation and hacking campaigns targeting Ukraine : NPR

[vi] UNC1151 Assessed with High Confidence to have Links to Belarus, Ghostwriter Campaign Aligned with Belarusian Government Interests (mandiant.com)

[vii] Unusual CEO Fraud via Deepfake Audio Steals US$243,000 From UK Company - Noticias de seguridad (trendmicro.com)

[viii] Scammers are now using AI to sound like family members. It's working. - The Washington Post

[ix] Threat Actors are Interested in Generative AI, but Use Remains Limited | Mandiant

[x] Cybercriminals train AI chatbots for phishing, malware attacks (bleepingcomputer.com)

[xi] La suplantación de identidad mediante correo electrónico es una técnica utilizada en los ataques de correo basura (spam) y ciberestafa (phishing) para engañar a los usuarios haciéndoles creer que un mensaje proviene de una persona o entidad conocida y confiable.