BLUECORE RESEARCH

Ciberseguridad / Conceptos

IA y Ciberseguridad

Whitepaper

www.bluecoreresearch.com
info@bluecoreresearch.com

JUNE 26, 2023

¿Cómo la inteligencia artificial, en particular la nueva inteligencia artificial generativa, afecta la ciberseguridad? Este trabajo examina desarrollos en inteligencia artificial, sus posibles implicaciones en ciberseguridad, y las esperadas tendencias en la industria.

Capacidades IA

Antes de discutir ciberseguridad, debemos entender primero la IA y qué puede hacer.

IA intenta imitar a la gente. Y significa mímica sin pensar. Solo actuar en formas que parecen similares. La IA no tiene una agenda, no entiende consecuencias, no aplica la lógica, sentimientos, o incluso sentido común. La IA no entiende qué dice o hace o por qué. Solo busca lo que la gente ha hecho y trata de imitarlo así los resultados lucen similar.

Para conseguir esto, necesitamos entrenar una inteligencia artificial. Entrenar a una IA significa darle muchos ejemplos y dejándole encontrar los patrones. No sabemos usualmente o entendemos que ve la IA en la información con la que la alimentamos, o si al seguir esos patrones conseguiremos los resultados esperados.

Pero desde que no entiende o piensa, podemos asumir errores cuando aplicamos escenarios intrincados. Errores que incluso niños no cometerían porque aún un chico entiende qué está haciendo.

Desafíos en ciberseguridad

Más allá de sus capacidades limitadas, la IA representa una significativa y en aumento amenaza a la ciberseguridad.

Una razón por esta amenaza es que la IA es mucho mejor hoy imitando humanos. Sistemas de seguridad y protocolos que intentan establecer que están hablando con una persona probablemente fallarán eventualmente. El ejemplo más simple es CAPTCHA, pero hay otros casos, incluyendo protocolos de mesa de ayuda, cuyo propósito es establecer que no se están comunicando con una máquina.

Un tema más alarmante es que las imitaciones de la IA aparecen como falto de patrones o, más precisamente, tiene patrones muy complejos de reconocer. Esto significa que la seguridad de los sistemas basados en la firma no será capaz de identificar las firmas de los ataques. Estos incluyen anti-virus, anti-malware, WAF, filtros de email, y más.

Por ejemplo, los virus polimorfos no son nuevos, pero la IA puede reescribir el código de ataque en formas elaboradas, haciéndola imposible de detectar. El mismo aplica a WAF y muchos otros sistemas de seguridad basados en firmas.

Finalmente, la IA moderna a menudo aparece como humana, y no podemos decir cuál es la diferencia. Entonces los ataques de ingeniería social pueden ser automatizados y escalados en volumen y complejidad. Las personas encontrarán más difícil de identificar phishing emails en tanto lucirán como los realistas.

Desde que cada phishing email será completamente diferente, ellos también engañarán spam filters. Así que usen filtros Bayesian u otros algoritmos, será duro reconocer múltiples emails son esencialmente una reescritura del mismo mensaje. Por lo tanto, protección de email será menos efectiva.

Debido a que las capacidades generativas de las IAs modernas las vuelven indistinguibles de los humanos, la mayoría de las defensas de seguridad basadas en IA probablemente se volverán ineficaces contra una IA atacante.

Mientras todo esto suena muy severo, en última instancia es solo otro clavo en la muerte del perímetro. El perímetro sufrió múltiples golpes a través de los años recientes, incluyendo trabajo remoto y BYOD. Junto con la actual alta tasa de éxito de los ataques de phishing, estamos perdiendo la batalla. Es probable que esta batalla empeore cuando los atacantes empiecen a aprovechar la IA.

Asegurar que nadie penetre el perímetro es imposible, y nuestra habilidad para controlarlo probablemente continuará siendo menor.

Diferenciación de actor

Mientras la IA presenta desafíos, muchos actores probablemente no tendrán acceso a ella. Puede cambiar en el futuro, pero el acceso directo a la tecnología y recursos requeridos de informática es poco probable se convierta en un lugar común.

Actor estatales probablemente ya tienen acceso a IA o pronto la tendrán. Ataques de grupos grandes tal vez también obtengan acceso a los recursos necesarios. Pero los script kiddies son el grupo más grande de atacantes y es poco probable que tengan acceso a tales capacidades.

En otras palabras, la diferencia entre los actores grandes y pequeños se volverá más extrema. Grandes actores son probablemente ya capaces de penetrar tu perímetro y encontrarán eso más fácil de hacer en el futuro, mientras pequeños actores continuarán utilizando sus actuales métodos y herramientas.

De cualquier forma, mejorar nuestras defensas contra las amenazas de la IA no solo mitigará los riesgos de la IA pero además significativamente reducirá el riesgo de pequeños actores.

Medidas efectivas

Es importante considerar qué tipo de medidas son efectivas contra la IA. Lo más alto que percibes la amenaza de actores con accesos a la IA, más debes invertir en tecnologías resilientes.

La seguridad más efectiva es la basada en la verdad. Esta reside en saber la respuesta correcta. La respuesta más simple es un usuario & contraseña. Ahí hay solo una respuesta correcta, y más allá de como los humanos- como una IA es, no será capaz de adivinarlo. Lo mismo aplica a autenticación de dos factores, preguntas de seguridad, cartas clave, biométricos, y más. Pero no es necesario invertir en complejidad o seguridad cara para combatir la IA desde que toda la seguridad basada en la verdad es IA resiliente.

Una medida en la que vale invertir es la gente. Más precisamente, pensando en la gente. Las IAs atacantes pueden imitar gente pero no puede pensar, y las IAs defensivas tienen la misma limitación. Trayendo a la gente de vuelta en el juego de la seguridad traerá de vuelta la fortaleza de pensamientos humanos y sentido común.

Pero hay buenas razones para nuestra mayor dependencia de la automatización en lugar de las personas. Estos incluyen el costo de personal, encontrar la persona apropiada, y, más importante, el desafío de humanos controlando volúmenes masivos de actividad.

Traer gente de vuelta a la seguridad requiere armarlos con sistemas detectives que pueden manejar los volúmenes de actividad y dar visibilidad a ella. Como los sistemas de

prevención y detección automática continúan tu declive en efectividad, seguridad basada en el humano se volverá más crítica.

Finalmente, deberíamos volvernos menos predecibles y no confiar tanto en paradigmas estándardes de seguridad y buenas prácticas. Pero personalizando nuestra seguridad y haciendo cosas diferentes, sistemas automatizados no sabrán qué esperar y falta de patrones para seguir o imitar.

Incluso hoy, escaneos de vulnerabilidad es una herramienta efectiva usada por atacantes. Es efectiva porque el software escaneado sabe que esperar y puede testear un número grande de vectores de ataques contra un número grande de sistemas. Lo más único de nuestra seguridad, lo menos probable es que sea por un escaneo de vulnerabilidad o una IA entrenada para asistir a un atacante.

Educación & Entrenamiento

¿Por qué están el desafío del perímetro e IA haciendo las cosas peor? La principal razón nuestros usuarios. Y también es una actitud general que Internet puede ser seguro y usarlo requerirá pequeños entrenamientos y conocimiento.

Nuestros usuarios constantemente interactúan con internet pero carecen de habilidades para protegerse y a la organización. También usan softwares que pretenden ser fáciles pero promueven comportamiento de seguridad pobre.

Por ejemplo, métodos de automatización de emails como SPF, DKIM, y DMARC pueden validar la fuente de email y están aumentando su popularidad. Como sea, la gente no mira la dirección de email o el dominio - miran lo más rápido de leer que es el nombre del display. Y las personas solo tienen parte de la culpa porque la mayoría del software de emails solo muestra este nombre para mostrar en la lista de corre electrónico y, usualmente, luego de abrir el email también.

Para empoderar a la gente a ser consiente de quién les manda el email, debemos hacer más que meramente hacer cumplir DMARC políticas. Debemos exponer a la gente a una dirección de email y nombre de dominios de quien envía y entrenarlos para mirar y entender esto.

Otro ejemplo es la facilidad con la cual la gente puede abrir un adjunto o hacer click en un link en un email. Dichas acciones deberían requerir al menos dos pasos. Por ejemplo, usuarios deberían salvar archivos adjuntos antes de ser capaces de abrirlos. Un link debería ser copiado y pegado en un navegador. Mientras tales pasos extra hacen el uso menos amigable, eso evita el accidental común click y fuerza a los usuarios a ser más consiente de acciones potencialmente peligrosas.

Similarmente, encriptado y certificados pueden validar la máquina remota y asegurar el secreto de la comunicación. Como sea, para esto ser efectivo, la gente necesita mirar el dominio en la URL.

Internet no es un lugar seguro. Esa es la realidad. Son individuos, grupos, y organizaciones que quieren hacernos daño. Con la IA a su lado, nuestra capacidad para fingir que estas entidades nefastas no existe disminuirá.

Centrado en datos

Como el perímetro se vuelve menos efectivo, debemos continuar cambiando nuestro foco a seguridad centrada en datos. Esto es una tendencia en crecimiento, y las amenazas de IA pueden probablemente potenciarlo.

Seguridad centrada en datos va alrededor de proteger los sistemas responsables por procesamiento de datos y almacenamiento. La motivación es que una filtración de perímetro será una filtración de datos fallidos si el atacante no puede comprometer los sistemas internos que manejan los datos.

Proteger las aplicaciones y bases de datos es efectivo contra amenazas internas y no solo las externas. El tipo de seguridad también crea una capa donde los ataques pueden ser detectados por múltiples medidas de seguridad, por lo tanto reduciendo las oportunidades de una filtración exitosa.

Seguridad centrada en datos busca proteger contra la gente, entonces una IA imitando exitosamente el comportamiento humano no es un riesgo en crecimiento. Centrado en datos es además más resiliente a la IA porque estos sistemas internos son usualmente separados de Internet por la menos una capa de protección de red. Esta separación de red hacer que sea más difícil para una IA basada en la nube atacar sistemas internos.

Seguridad detective

La creciente importancia de la seguridad centrada en datos, involucramiento humano, y menos predecible defensas llevan atención para un tipo particular de soluciones de seguridad detectives.

Estas son soluciones que pueden:

- Manejar el volumen extremadamente alto de actividad de bases de datos y aplicaciones.
- Ofrece personalizaciones usando ambientes específicos atributos en alertando, reporteando, análisis, y más.
- Dar al personal de seguridad completa visibilidad sobre todo lo que está sucediendo usando herramientas forenses interactivas y de avanzada.

Core Audit de Blue Core Research hace más fácil conseguir estos objetivos. Desde análisis de anomalías a 360° forenses, y mucho más, Core Audit es un solución integral que puede ayudar a que encuentres y superar tu actual y futura seguridad y desafíos de cumplimiento normativo.

Ideas finales

Es difícil predecir cuando la amenaza de IA de ciberseguridad se materializará o por cuánto tiempo tomará antes que sepamos de una filtración asistida por IA. Hoy, esta amenaza es probablemente limitada por el estado o muy grandes actores de quien las filtraciones de datos son a veces no descubiertas y cuyos métodos pueden mantener desconocidos.

IA es poco probable que cambie los riesgos de pequeños actores, pero IA- defensas resilientes probablemente también será efectivo contra ellos.

El desafío principal continúa siendo mejorar el comportamiento humano. Es probable que la Ingeniería social aumente como riesgo principal, ya que las personas siguen siendo el talón de Aquiles de la seguridad.

IA no parece forzar un cambio en las tendencias de seguridad, solo acelerarlas y demandar atención a cambios ya necesarios. Estos incluyen el valor menguante de las defensas perimetrales de alta gama ya que luchan por la justificación y el cambio de enfoque hacia la seguridad centrada en los datos.