ARTÍCULO

Retos de la Inteligencia Artificial para no comprometer la privacidad de los datos

por Miguel Morales Sandoval, Heidy Marisol Marin Castro y Karina Figueroa Mora

Hoy en día existe una gigantesca recopilación de datos a través de casi todos nuestros movimientos en internet. Por ejemplo, mientras compramos, usamos las redes sociales, o realizamos algún trámite en línea, generamos datos digitales como nuestros nombres, direcciones, teléfonos, fotografías o información relacionada con compras en línea (productos, marcas, costos). Los algoritmos de Inteligencia Artificial (IA) generalmente usan estos datos para realizar predicciones u obtener algún conocimiento. Quienes realizan estas tareas de análisis saben que existe un reto enorme sobre el manejo y privacidad de los datos, ya sean personales (número de teléfono) o de carácter privado (salario). Por lo que, el desafío es: ¿cómo entrenamos un modelo de IA sin que esos datos estén vulnerables? El lector podría pensar, "simplemente quitémoslos" y eso sería lo ideal en un escenario de unos pocos registros pero cuando hablamos de cientos de miles de millones es una tarea titánica y prácticamente imposible, ¡primer gran reto! Por supuesto, aunque varía el alcance en cada país, en general cada vez más se desarrollan o actualizan regulaciones en torno a garantizar la privacidad del uso de datos. Por ello, este es un aspecto de gran relevancia para los algoritmos que usan datos para derivar conocimiento, como los de IA [1].

Por otro lado, un entrenamiento para un modelo de IA consiste básicamente en detectar un patrón en los datos. Esta técnica se utiliza para muchas tareas de minería de datos, clasificación, agrupamiento, etc. Por ejemplo, imagine que ud. que se desea realizar una tarea de agrupamiento de salarios de acuerdo con años de experiencia, dentro de un contexto o plano organizacional, para que con estos resultados se puedan realizar más adelante predicciones y una planeación financiera en relación con presupuesto de recursos humanos. Quitar ejemplos de personas con bajo o alto salario, por cuestiones de privacidad, sin duda afectará al modelo de agrupamiento y de predicción, y llevará a una toma de decisiones sin la información correcta. Por otro lado, quien calcule el modelo de agrupamiento o de predicción, estaría comprometiendo la privacidad de los datos de salarios, los cuales deberían ser privados.

En este artículo presentamos el problema relacionado con la privacidad de datos para ser usados en algoritmos de IA y cómo se puede abordar esta problemática desde un enfoque de solución basado en el cifrado homomórfico ("!¿homomó....qué?!", ya le explicaremos). Esta tecnología le permite al propietario de los datos delegar las tareas de procesamiento y de análisis a un tercero, cifrando los datos de entrada sin que éstos pierdan su utilidad. Esto es, se garantiza que dicho algoritmo no aprenderá nada de los datos del propietario (básicamente, no sabrá de quién o de qué estamos hablando). Finalmente, presentamos algunos de los retos relacionados con esta problemática y oportunidades que se tienen en el plano de la investigación y del desarrollo tecnológico sobre este tema.

La confidencialidad, requisito para la privacidad de datos, debe restringir el acceso a datos solo a las entidades autorizadas. El cifrado de datos ha sido una herramienta efectiva para garantizar la confidencialidad.

Cifrado homomórfico de un texto

La criptografía es un campo de las matemáticas y de las ciencias computacionales, cuyos orígenes se remontan a antes de nuestra era. Iniciemos explicando que significa *cifrar* un texto: básicamente es una transformación de dicho texto. Imagine el lector que quiere cifrar el texto "*hola mundo*". El resultado se puede ver cifrado como "*3oij4s823ADe23*". Sin embargo, el texto "*holamundo*" podría ser cifrado como "*aLKQeirjq34*". Note que a pensar de que solo cambió una letra en ambos textos (el espacio), las transformaciones son completamente distintas. En principio, la transformación debe producir texto ilegible, ya que el principal objetivo del cifrado es ocultar el mensaje original (que si es legible). Existen diversas técnicas y algoritmos para cifrar, el más fácil y básico consiste en reemplazar una letra por otra, y nuestros textos cifrados se verían así "*kpñs qimfp*" y "*kpñsqimfp*". Cómo puede darse cuenta, en este ejemplo se ha usado un algoritmo débil pues se puede intuir que es casi el mismo mensaje. Actualmente, los cifradores son una pieza fundamental en la seguridad de la información digital. Algunos ejemplos de su amplio uso son: cuando

nos comunicamos a través de aplicaciones de mensajería instantánea, donde se nos informa que los mensajes intercambiados están cifrados; cuando realizamos compras en línea y accedemos a una página web que se nos dice que es segura porque la conexión está cifrada; o cuando pagamos usando una tarjeta bancaria. Los cifradores modernos son mucho más complejos que el que hemos mencionado en el ejemplo, pero el principio de funcionamiento es el mismo. Todos los cifradores, desde los antiguos hasta los modernos, hacen uso de una *llave* para realizar la transformación. La llave, en este caso, también puede verse como un conjunto de letras y generalmente es de un gran tamaño para evitar que pueda ser adivinada. Entonces, para cifrar se requieren 3 elementos: el mensaje, la llave, y el algoritmo de cifrado. Todo esto se muestra en la Figura 1. Una vez cifrado, un mensaje solo podrá restaurarse a su estado original mediante el descifrado, el cual revierte la transformación realizada por el cifrado.

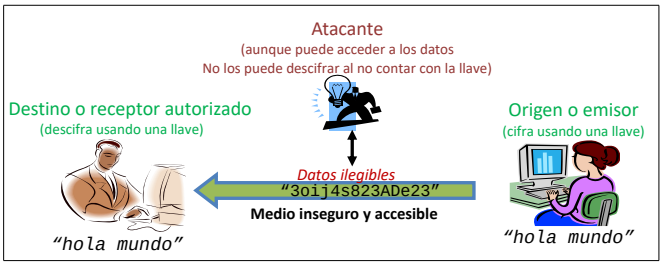

Figura 1. Modelo básico donde la criptografía garantiza la privacidad de los datos que se generan en un origen, se transmiten o se almacenan en un medio inseguro que puede ser accedido por terceros (atacante), pero que solo es accesible por entidades autorizadas que cuente con la llave para descifrarlos.

Comúnmente, el cifrado es visualizado como una caja que resguarda un mensaje, cerrada con un candado bajo llave. Nadie excepto quien cuente con la llave que abre dicho candado podrá abrir la caja y extraer el mensaje. Esta analogía en la práctica se implementa con algoritmos basados en principios matemáticos, particularmente en el álgebra abstracta. La criptografía tiene una relación muy estrecha con el origen de la computación y de la IA. Alan Turing, considerado como el padre de la IA, lideró el equipo que estuvo a cargo de descifrar los códigos de la máquina Enigma, un cifrador usado por los alemanes y sus aliados durante la segunda guerra mundial. Turing y su equipo pudo descifrar los códigos de Enigma gracias a la creación de un dispositivo, que evolucionó después para convertirse en el modelo de computación que se usa en prácticamente todos los dispositivos de cómputo actuales.

El reto es mantener la privacidad: obtener conocimiento sin revelar nada de los datos de entrada.

El cifrado hace ilegible un texto, mensaje o dato, y en principio le quitaría utilidad a los datos en aras de mantener su privacidad. Pero, ¿se podrían seguir aplicando algoritmos de IA sobre los datos cifrados?. La respuesta, estimado lector, es que sí es posible, y se puede lograr usando un cifrado especial, llamado *cifrado homomórfico*.

Para la IA, la relevancia del cifrado homomórfico es que le puede proporcionar la capacidad de seguir realizando operaciones sobre los datos cifrados, de tal forma que el procesamiento resultante es equivalente al cifrado del procesamiento de los datos sin cifrarlos. Esto significa que es posible realizar cálculos sobre los datos sin tener acceso a éstos en claro, y así preservar la privacidad de los mismos. Este no es el caso de las técnicas de cifrado tradicionales, donde el resultado de operaciones que involucran datos cifrados no tiene una interpretación significativa (como nuestro primer ejemplo). Sin embargo, el cifrado homomórfico comparte muchos de los conceptos, términos y notación con la criptografía convencional [2].

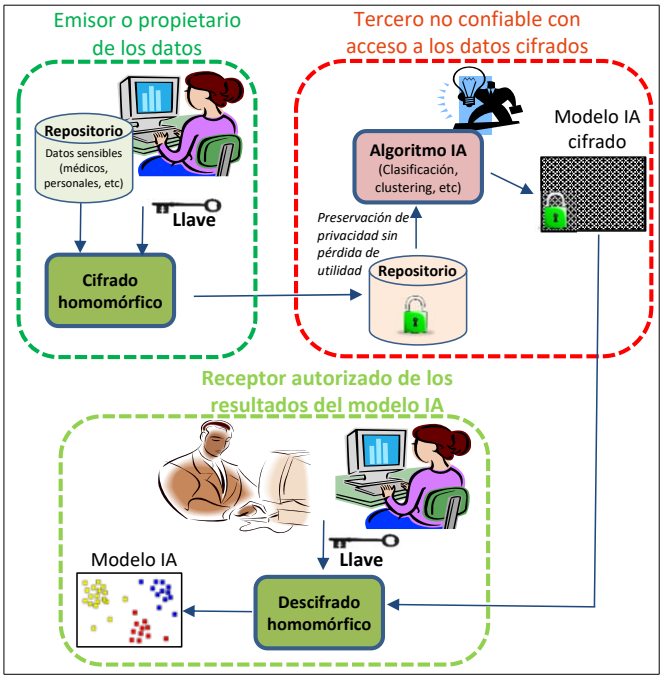

Figura 2. El cifrado homomórfico permite mantener la privacidad de los datos sin pérdida de utilidad para algoritmos de IA.

Al igual que en el cifrado convencional, en un cifrado homomórfico se tienen 3 operaciones: generación de llaves, una para cifrar y otra para descifrar, en algunos

casos pueden ser iguales, cifrado (se usa la llave para cifrar) y descifrado (se deshace el proceso de transformación usando la llave para descifrar). En la Figura 2 se pueden ver estas 3 operaciones. Quien ejecuta el algoritmo IA es el principal atacante a la privacidad de datos que tienen un propietario. El algoritmo de IA debe seguir funcionando aún con los datos cifrados y obtener el modelo para el que está diseñado. Este modelo, cifrado, no tendría ningún significado para quien ejecuta dicho algoritmo: solo podrá ser descifrado por las entidades autorizadas, que puede ser el mismo propietario de los datos o algún otro.

El cifrado homomórfico supone que es posible definir dos operaciones: una que se aplica sobre los datos legibles y otra que se aplica sobre los correspondientes datos cifrados. La operación con los datos cifrados debe implicar la operación de manera indirecta con los datos en claro. Veamos esto a través de un ejemplo. Suponga que los datos legibles son salarios, los cuales no deben accederse en claro. Ahora, suponga que cifrar estos datos consiste en realizar una exponenciación usando una base fija, por ejemplo, la base 3. Así, el cifrado del salario 1,000 sería 3^{1000} (un número muy grande). Por consiguiente, el descifrado de 3^{1000} debería ser 1,000. Evidentemente, un tercero tendría acceso al cifrado del sueldo, 3^{1000} en este caso, pero no tendría acceso al valor real que es 1,000 (ver Figura 2). Ahora, suponga que multiplicamos dos salarios cifrados (esto lo realizaría el algoritmo de IA, quien operaría sobre datos cifrados), por ejemplo 3^{1000} (el cifrado del salario 1,000) y 3^{1500} (el cifrado del salario 1,500), lo cual, usando la propiedad de exponentes sería $3^{1000} \times 3^{1500} = 3^{(1000+1500)}$. Es decir, *una multiplicación* de datos cifrados resultó en *una suma* de los correspondientes datos en claro, y esta suma se obtuvo de manera indirecta, sin tener acceso a esos datos como tal. Para acceder al resultado de la suma (que sería el dato de interés para el propietario de los datos), se debe descifrar el resultado de la multiplicación (como en la Figura 2), que sería el valor (1,000 + 1,500). Esta es la idea básica del cifrado homomórfico. Nuestro querido lector puede corroborar el ejemplo previo dando algunos valores distintos para los salarios y para la base usada.

Desde luego, un aspecto de suma importancia del cifrado homomórfico es la seguridad. En nuestro contexto de la Fig. 2, el cifrado debe ser robusto para evitar que el tercero no confiable (con gran capacidad de cómputo dicho sea de paso) pueda romper el cifrado. La seguridad del cifrado homomórfico está generalmente asociado a la dificultad de resolver un problema matemático, mismo que se usa para la generación de las llaves. Entre más difícil el problema a resolver, más seguro (más difícil de romper) es el cifrado. Uno de los problemas mayormente usados es el cálculo de logaritmos discretos o la factorización de números enteros grandes. Actualmente, estos problemas se consideran difíciles si se usan tamaños de números, al menos, de 2048 bits. Las operaciones de cifrado y descifrado involucran operaciones aritméticas con éstos números. Para darse una idea de lo grande que pueden ser estos números, 2048 bits equivale a un número entero de 620 dígitos decimales.

La seguridad de datos se refiere a garantizar el servicio de confidencialidad de los datos que se delegan a un tercero para análisis, y es un requisito para obtener privacidad.

Retos y oportunidades

En este ámbito, la privacidad se centra en los propietarios de los datos, que pueden ser individuos o grupos, con el objetivo de proteger su información privada mientras sus datos se utilizan por terceros para la construcción de modelos de IA procesados en la nube como *la minería de datos como servicio*, esto es, un tercero accede a un repositorio de datos, posiblemente sensibles, para analizarlos y extraer conocimiento de ellos, etc. Así, cuando la naturaleza de los datos sobre los que se aplicarán algoritmos de minería de datos es sensible, como en el dominio de los ejemplos previos, se requiere resolver desafíos de privacidad.

De manera general se puede decir que si los datos contienen información sensible o de acceso restringido, o si el uso de dichos datos está sujeto a regulaciones de acceso, su uso en algoritmos de IA (de análisis para descubrir conocimiento) puede llevar a comprometer la confidencialidad y por tanto la privacidad de los mismos.

Reto 1: Eficiencia y nivel de seguridad. El principal desafío del cifrado homomórfico es que los esquemas actuales solo admiten un número limitado de operaciones aritméticas, como en el caso del cifrado homomórfico completo, o totalmente homomórfico, donde es posible utilizar dos operaciones aritméticas: la suma y la multiplicación. Por otro lado, el cifrado homomórfico parcial únicamente soporta un tipo de operación, ya sea la suma o la multiplicación. Siendo la suma y multiplicación operaciones elementales, cualquier operación aritmética en algoritmos de inteligencia artificial podría implementarse a través de un cifrador homomórfico, aunque con algunas restricciones o con una penalización en la eficiencia.

Dicha eficiencia es afectada directamente por el nivel de seguridad del cifrador homomórfico, el cual como se

ha comentado está asociado a la dificultad para resolver un problema matemático. Dicha dificultad está en función del tamaño de los números usados para las operaciones aritméticas de cifrado y descifrado. Para números de 2048 bits, como se recomienda actualmente, el impacto del costo de las operaciones del cifrador podría ser muy alto cuando se procesen grandes cantidades de datos.

Oportunidad. El despliegue eficiente de los métodos de cifrado homomórfico puede realizarse haciendo uso de técnicas de paralelismo u otros modelos de aceleración que permitan contar con métodos tanto seguros como eficientes. Es en esta parte donde existe un área de oportunidad de investigación y desarrollo. Por ejemplo, un método de IA con preservación de privacidad de datos puede desplegarse bajo un enfoque hardware/software, donde las partes más demandantes computacionalmente se aceleren con hardware dedicado, o se implementen bajo modelos como el cómputo heterogéneo, el cual explota el poder de cómputo de todos los dispositivos disponibles en la infraestructura de cómputo al mismo tiempo, como núcleos del procesador, GPUs o FPGAs.

Reto 2: Viabilidad y aplicación en un escenario de cómputo cuántico. Los métodos criptográficos homomórficos se definen en estructuras algebraicas donde se establece un problema matemático, como la factorización de enteros o el cálculo de logaritmos discretos. La dificultad para resolver ese problema determina el nivel de seguridad del método criptográfico. Por ello, las estructuras algebraicas deben ser de tamaño suficiente para que el problema sea difícil de resolver. Esta suposición se ha mantenido hasta estos tiempos, ya que no existe un algoritmo eficiente (con complejidad polinomial) que pueda ejecutarse en una computadora convencional, que resuelva el problema en el que el cifrado homomórfico sustenta su seguridad. Sin embargo, ya se ha demostrado que, para el dominio del cómputo cuántico, si existen algoritmos de complejidad polinomial que pueden resolver prácticamente cualquiera de los problemas en los que se sustentan los algoritmos criptográficos actuales. Por ello, los métodos criptográficos homomórficos, sobre los que se construya la protección a la privacidad están amenazados en un escenario postcuántico. De acuerdo con expertos, se prevé que en 2030 se cuenten con computadoras cuánticas capaces de vulnerar cualquier esquema de cifrado actual.

Oportunidad. Desde 2006, se está desarrollado una nueva línea de investigación llamada Criptografía Post-Cuántica [3] (PQC, por sus siglas del Inglés). Esta comprende el desarrollo de algoritmos criptográficos resistentes a ataques de computadoras cuánticas. Actualmente, se han identificado tres grandes familias de enfoques de PQC, y se ha realizado recientemente trabajo de estandarización de esquemas PQC para cifrado [4], y sobre los cuales se pueden construir sistemas homomórficos, poco explorados aún. En octubre de 2022 concluyó la ronda 3 de evaluación y selección de algoritmos PQC por parte del NIST (National Institute for Standadars and Technology), que finalmente recomendó a la criptografía basada en látices como la más adecuada para cifrado de datos. Sin embargo, PQC aún no se ha explorado ampliamente en el contexto del cifrado homomórfico para la minería de datos como servicio. De ahí que ésta sea un área de oportunidad importante de investigación y desarrollo en los próximos años.

Reto 3: Disponibilidad de herramientas. Herramientas disponibles para análisis de datos mediante algoritmos de IA, con preservación de privacidad, aún son escasas. Si fuera posible garantizar la privacidad de los datos, se podrían disponer de herramientas en línea, en la nube, donde los propietarios de los datos puedan subir sus datos, procesarlos y obtener algún modelo de IA de interés (agrupación o clasificación por ejemplo), sin riesgo de que los datos de entrenamiento queden expuestos. Desde el punto de vista de la academia y de la investigación, este tipo de herramientas en la nube también pueden ser usadas para estudio y mejora de métodos de privacidad basados en cifrado homomórfico. Una herramienta como tal podría incluir más métodos que le permitan tener distintos alcances respecto a algoritmos de IA soportados y podría ser configurable para soportar un nivel de seguridad específico. Por ello, para promover la adopción de herramientas de seguridad y privacidad como servicio es necesario construir más herramientas, desde la perspectiva del usuario final y también del científico de minería de datos.

Oportunidad. El cuerpo del conocimiento en relación con el cifrado homomórfico y la criptografía postcuántica, así como de los modelos de aceleración bajo entornos de cómputo en la nube pueden converger para crear modelos de protección de privacidad robustos, eficientes y seguros, así como herramientas flexibles y configurables, que promuevan la adopción de servicios en la nube en los casos de uso más representativos, como algoritmos de agrupamiento o de clasificación.

Conclusiones

La privacidad de datos es un aspecto que ha cobrado relevancia en distintos ámbitos, motivada por la alta disponibilidad y acceso a los mismos, sobre todo cuando dichos datos son sensibles (datos médicos, personales, financieros, etc.). La garantía de privacidad está actualmente siendo incorporada en la legislación de cada país, por lo que se debe preservar la privacidad cuando los datos se generan, almacenan, transmiten o *cuando se usan.* Este último caso es el que compete al dominio de los algoritmos de inteligencia artificial, como aquellos que se usan en minería de datos y en aprendizaje de máquinas. Ante un escenario de cómputo en la nube, es cada vez más común disponer de algoritmos de análisis de datos como servicio. Una herramienta que permite garantizar

la privacidad de los datos usados en escenarios como el del cómputo en la nube es la criptografía homomórfica, la cual permite operar con los datos cifrados y no con los datos en claro. El resultado producido, al descifrarse por el propietario de los datos, será el mismo que se hubiera obtenido usando los datos sin cifrar. De esta forma, se pueden construir herramientas efectivas que salvaguarden la privacidad de los datos.

Sin embargo, no es suficiente que las herramientas en la nube sean efectivas, se requieren que sean eficientes y viables de operar en aplicaciones reales. En este sentido, se requiere aún resolver problemas abiertos en lo que se refiere a eficiencia, seguridad y practicidad de los sistemas homomórficos como servicios. Por ello, los modelos actuales de cifrado homomórfico deben estudiarse usando niveles de seguridad recomendables, sin pérdida de eficiencia y bajo un escenario de cómputo postcuántico, donde los esquemas de cifrado homomórfico basados en criptografía convencional no serán seguros.✳

REFERENCIAS

1. Bartneck C., Lutge C., Wagner A. y Welsh S. (2021) "Privacy Issues of AI". In An Introduction to Ethics in Robotics and AI, Springer International Publishing, Cham, pp. 61-70.
2. Almutairi N., Coenen F., Dures K. (2017) "K-Means Clustering Using Homomorphic Encryption and an Updatable Distance Matrix: Secure Third Party Data Clustering with Limited Data Owner Interaction". In Bellatreche L., Chakravarthy S. (eds) Big Data Analytics and Knowledge Discovery. DaWaK 2017. Lecture Notes in Computer Science, vol 10440. Springer, Cham. https://doi.org/10.1007/978-3-319-64283-3_20.
3. Bernstein D.J. (2009) "Introduction to post-quantum cryptography". In Post-Quantum Cryptography, Springer Berlin Heidelberg, pp. 1-14.
4. Alagic G., Alperin-Sheriff J., Apon D., Cooper D., Dang Q., Kelsey J., Liu Y., Miller C., Moody D., Peralta R., Perlner R., Robinson A. y Smith-Toneand D. (2020) "Status report on the second round of the NIST post-quantum cryptography standardization process". The National Institute for Standards and Technology (NIST).
5. ur Rehman M.H., Yaqoob I., Salah K., Imran M., Jayaraman P.P. y Perera C. (2019) "The role of big data analytics in industrial Internet of Things". Future Generation Computer Systems, Vol. 99, Elsevier, pp. 247-259.
6. Almutairi N., Coenen F. y Dures K. (2021) "Secure third-party data clustering using SecureCL, data and multi-user order preserving encryption". Expert Systems, Vol. 38, No. 7, Wiley, pp. e12581.
7. Qiu G., Gui X. y Zhao Y. (2020) "Privacy-Preserving Linear Regression on distributed data by homomorphic encryption and data masking". IEEE Access, Vol. 8, IEEE, pp. 107601-107613.

SOBRE LOS AUTORES

Miguel Morales Sandoval es profesor-investigador adscrito a la Unidad Tamaulipas del Cinvestav. Es miembro del Sistema Nacional de Investigadores Nivel II. Sus intereses académicos y de investigación se centran en Seguridad de Datos. Recibió el doctorado en Ciencias de la Computación en 2008 del Instituto Nacional de Astrofísica, Óptica y Electrónica de México. Actualmente está enfocado en el desarrollo de esquemas de seguridad de datos con aplicaciones en la Ciencia de datos, la Nube y el Internet de las Cosas.

Heidy Marisol Marin Castro es profesora e investigadora por México-Conacyt adscrita a la Facultad de Ingeniería y Ciencias de la Universidad Autónoma de Tamaulipas. Doctora en Ciencias de la Computación por el Centro de Investigación y de Estudios Avanzados del IPN, con intereses de investigación en Ciencia de los Datos con énfasis en Minería de Procesos y de Datos, así como Aprendizaje Máquina y Gestión de Información Web. Es Miembro del Sistema Nacional de Investigadores Nivel 1.

Karina Figueroa Mora es Doctora en Ciencias de la Computación por la Universidad de Chile, Ingeniera Electricista y Maestra en Ing. Eléctrica opción Sistemas Computacionales por la Facultad de Ingeniería Eléctrica de la Universidad Michoacana de San Nicolás de Hidalgo. Profesora e investigadora de la Facultad de Ciencias Físico Matemáticas "Mat. Luis Manuel Rivera Gutiérrez". Miembro del Sistema Nacional de Investigadores. Perfil Prodep. Especialista en recuperación de información, diseño y análisis de algoritmos, Bases de Datos (de texto y métricas).