

“Lo más importante para tener una IA exitosa en salud es hacer una correcta gestión de datos”



Foto: Cortesía - archivo personal

Oswaldo Solarte Pabón

Docente e investigador de la Universidad del Valle

La IA y los datos están muy relacionados, y se sirven el uno al otro. La IA sin datos no es posible; se alimenta de los datos. Entre mejores y más datos tenemos, obtendremos mejores modelos de IA. Y para aprovechar esos datos necesitamos de buenos algoritmos de IA, o sea que es una relación mutua”. Esta es la orientación de Oswaldo Solarte Pabón, doctor en Sistemas y Computación de la Universidad Politécnica de Madrid y experto en aplicación de tecnología e IA en el sector salud.

Solarte Pabón señala que cuando se van a asumir proyectos de IA es muy importante seguir una metodología que parta de entender los datos, de prepararlos y modelarlos, porque muchas veces se modelan y utilizan

algoritmos de IA sin antes entender y preparar los datos, y por ello no se obtienen buenos resultados al evaluar la precisión del modelo.

Explicó que el sector salud es el que genera más datos en el mundo, según diversos análisis. La salud produce más datos que otros sectores como la industria y los servicios financieros, la educación, el turismo y el entretenimiento, y al originar tantos datos estos deberían aprovecharse para beneficio de los pacientes, los profesionales y las instituciones de salud.

De esos datos producidos por el sector salud, el 90 % son no estructurados y el 10 % estructurados, pero no se aprovecha el enorme valor de aquellos no estructurados. Entre los datos generados en salud hay textos clínicos, datos administrativos y de diagnóstico, codificación de procedimientos, encuestas clínicas, ensayos clínicos, datos genómicos e imágenes médicas. Según el experto, en esa *big data* de salud lo más importante es el valor que le podamos sacar a esos datos y la IA nos ayuda a aprovecharlos.

Los datos no estructurados son los que realmente deben aprovecharse,

porque brindan más información y le aportarían más capacidad a la IA. A su vez, la IA ayuda a procesar esos datos. Por ejemplo: el área de oncología de una clínica quiere estudiar el comportamiento del cáncer de mama en los últimos 5 años en su ciudad. Los oncólogos solo conocen sus 10 pacientes, pero si se quiere hacer un estudio epidemiológico de los últimos 10 años que involucra a 2000 pacientes, la única opción es leer todos esos datos manualmente. Eso no es posible técnica ni económicamente; un especialista tardaría tres años en leer toda esa información y además esto representaría un costo altísimo. En este caso la IA puede apoyar a los médicos al sacar datos e información que para ellos es muy difícil obtener de otra forma.

Si queremos aplicar la IA, tenemos que trabajar con todo este tipo de datos, integrarlos, sacarles el máximo provecho. Hoy en día las únicas organizaciones que están beneficiándose de estos datos son las *big tech*, grandes tecnológicas como Amazon y Facebook, porque en Colombia las empresas no están sacando ventaja de esos datos. ¿Qué hacer entonces para gestionar y aprovechar estos datos?

La IA depende de la calidad de los datos suministrados. Hay universidades y organizaciones en el mundo preparando datos de calidad: la Universidad de Harvard y la de California tienen sitios con *datasets* públicos de salud, anotados por expertos, que pueden aprovecharse. Como esos datos fueron hechos con pacientes de Norteamérica, la pregunta es si estos *datasets* sirven para un hospital de Cali. Ahí es donde debe intervenir el criterio médico y clínico, porque no todo lo va a hacer la IA; la IA ayuda, pero al final el que toma las decisiones es el profesional de la salud, indica el experto. Por eso, si queremos avanzar en IA en salud,

Si queremos aplicar la IA, tenemos que trabajar con todo este tipo de datos, integrarlos, sacarles el máximo provecho. Hoy en día las únicas organizaciones que están beneficiándose de estos datos son las big tech, grandes tecnológicas como Amazon y Facebook, porque en Colombia las empresas no están sacando ventaja de esos datos.

también debe haber iniciativas de las universidades y los gobiernos locales y nacionales para sacarle el máximo provecho.

Un requisito para hacer buenos modelos de IA es la “anotación de datos en salud”. Hay datos públicos en bases de datos de universidades de Norteamérica, pero para hacer modelos propios de IA en Colombia primero debe hacerse anotación de datos, un proceso complejo; de hecho, un solo experto médico no anota los datos, tienen que ser al menos dos expertos y, al final, debe mirarse “la consistencia en los datos”. Dice Solarte Pabón: “En experimentos observamos que cuando los médicos anotan datos de manera separada, sin conocerse, hay inconsistencia en los datos. Si no hay consistencia entre los expertos humanos, todavía no le podemos pasar esto a la IA para hacer modelos. Por ello, para avanzar en IA en salud es crucial anotar bien los datos”.

En el texto clínico hay dos grandes retos: la IA debe detectar diagnósticos, medicamentos, es decir las variables de interés médico; también debe reconocer si se afirma o se niega un diagnóstico, considerar la negación y la incertidumbre en las notas clínicas. De ahí que trabajar con texto clínico de manera automática sea algo complejo. Explica Solarte Pabón que hay dos problemas principales al entrenar modelos de IA, que constituyen las fuentes más comunes de errores: el



sesgo de los datos (diferencia entre la predicción y su objetivo-*underfitting*) y la varianza (un modelo con alta varianza se ajusta a los datos de entrenamiento, pero funciona mal fuera de la muestra-*overfitting*).

“El sesgo se presenta cuando creo un modelo con datos que no son de calidad y por tanto el modelo no aprende. Y la varianza es cuando entreno un modelo con datos que al ponerlo en práctica no se comporta bien. Estas dos cosas no las queremos en IA; queremos modelos que generalicen bien. El principal problema en la IA se presenta en los datos, no en los algoritmos. Cuando entrenen un modelo y les dé baja precisión, no le echen la culpa al algoritmo; busquen problemas en los datos. Con lo que debemos tener más cuidado en la IA es con los datos, ellos mandan en la IA”, recalcó.

Los retos en la gestión de datos

Las instituciones de salud se enfrentan a varios desafíos como la modernización de la infraestructura tecnológica para aprovechar la IA. La arquitectura tecnológica tiene que ver con manejar bases de datos, integración de las historias clínicas mediante aplicaciones y estándares como API y HL7/Fast Healthcare Interoperability Resource (FHIR), almacenamiento en la nube con seguridad reforzada, utilización de lagos de datos (*data lake*) para análisis avanzados y de bodegas de datos para analítica de datos, repositorios donde se han integrado datos.

Para crear modelos más inteligentes de IA, es necesario integrar datos de diferentes fuentes como historias clínicas, datos de medicamentos y demográficos, datos de facturación y de reclamos. De ahí que dicha integración sea uno de los primeros problemas a los que se enfrentan las instituciones de salud en nuestro país, como afirma Solarte Pabón.

El experto agregó que otro reto es que los sistemas o modelos que hagamos tienen que estar integrados con estándares y protocolos internacionales: “Hoy cada clínica tiene su propio sistema de información y cuando quiere compartir datos con otra clínica, debe imprimirlos para que el paciente los lleve a la otra institución. Eso no puede ser posible. Si queremos hacer IA, si usted se va a Bogotá y le pasó algo, allá el especialista debe tener



La calidad de los datos define la calidad de los modelos. Unos datos incompletos o inconsistentes afectan los modelos creados. Es necesario tener mecanismos de validación automática e integración con sistemas de verificación y control, y validar indicadores de calidad (completitud, consistencia, precisión, relevancia, accesibilidad).

acceso a su historia clínica. Y eso se hace con estándares como HL7 o FHIR, tecnologías para integrar”.

Otro desafío tiene que ver con la seguridad y la privacidad en los datos. Dice Solarte Pabón: “Trabajar con datos en salud es bien complejo, sobre todo por la parte de la privacidad, porque no queremos enfrentar una demanda; una de las primeras cosas que debe hacerse si va a implementar IA es anonimizar los datos”. Señala que en la Unión Europea está la Ley General de Protección de Datos (GDPR), en Estados Unidos la Health and Human Services (HIPAA) para proteger datos y, en Colombia, la Ley 1581 del 2012 o Ley de Protección de Datos Personales. Sin embargo, en el sector salud se requiere más especialización por tener un rezago más grande en manejo de datos; por ello siempre debe hacerse cifrado, anonimización y control de accesos, auditoría de accesos y trazabilidad.

El experto señala que las instituciones de salud tienen que implementar la gobernanza de los datos para contar con datos de calidad. “La IA depende de la calidad de los datos. Lo más importante en la IA son los datos. La calidad de los datos define la calidad de los modelos. Unos datos incompletos o inconsistentes afectan los modelos creados. Es necesario tener mecanismos de validación automática e integración con sistemas de verificación y control, y validar indicadores de calidad (completitud, consistencia, precisión, relevancia, accesibilidad). Las

instituciones de salud tienen que implementar departamentos que se encarguen de la calidad de sus datos. Muy pocas empresas y hospitales en Colombia tienen una política de gobernanza de datos”.

Al respecto, indicó Solarte Pabón que las organizaciones podrían implementar *frameworks* para la gestión institucional de datos; por ejemplo, el *framework* Data Management (DAMA) define los principios fundamentales, las mejores prácticas y las funciones esenciales de la gestión de datos.

El experto reitera que lo más importante para que la IA tenga éxito es la correcta gestión de datos: un estudio registró que el 80 % de las organizaciones afirmaban que la IA generativa tendrá un impacto en sus negocios, pero en el 2027 el 60 % de todos los proyectos hechos con IA van a fracasar porque no contaron con un manejo apropiado de los datos. “Todo el mundo quiere hacer IA, pero no han resuelto el problema de hacer una buena gestión de los datos. Si no se hace, es posible que fracase el proyecto de IA. Los datos siempre van de la mano con la IA y mucho más en el sector salud”, agregó.

También advirtió la necesidad de cuidar nuestros datos: “Hay médicos que suben la historia clínica al ChatGPT y le piden un resumen sin anonimizarla. Ahí hay un peligro, porque se alimenta una herramienta de IA de otro país, privada, con datos de clínicas y hospitales de nuestro país. Es dar un enorme poder a estas empresas al tener nuestras historias clínicas. El llamado es: no utilizar herramientas como ChatGPT y demás IA sin antes resolver la anonimización de datos, porque las hacemos más poderosas. Utilizar ChatGPT no está mal, el problema es con qué datos lo uso”.



Foto: Cortesía - archivo personal

Oswaldo Solarte Pabón

Recalca que no deberían hacerse diagnósticos con lo que dice Chat-GPT sin preguntarse antes con qué datos fue entrenado el modelo: “Hay muchos datos en repositorios públicos, hechos con pacientes de otros países. Hay que adquirir una cultura del dato. Los médicos, hospitales y profesionales de la salud tienen que cuidar sus datos. Eso tiene el máximo valor”.

Solarte Pabón agrega que debe tenerse mucho cuidado al crear modelos de IA basados en los modelos grandes de lenguaje: “Si usted le sube una historia clínica, él se queda con esos datos. Los protegen, sí, pero el modelo los va a utilizar a su favor; los usa para volverse más inteligente. Las clínicas y los hospitales pueden desarrollar sus modelos de IA propios, protegiendo sus datos, sin que sus historias clínicas salgan del hospital. Y para eso también tenemos capacidades, hay otras tecnologías. Imagínense que un hospital o una ciudad como Cali tenga su propio ChatDoc de salud, especializado, pero que nuestros datos de la ciudad de Cali no los tomen otros. ¡Eso sería genial!”.

Algunas recomendaciones para trabajar con IA en instituciones de salud

Comenta Solarte Pabón que el objetivo principal de trabajar con IA en instituciones prestadoras de servicios de salud (IPS) debe ser el bienestar de los pacientes a nivel clínico, pero también se puede lograr beneficio económico al reducir costos y mejorar procesos para todos. “Es buscar el bienestar de los pacientes, pero también el bienestar de la institución, de los médicos y del personal asistencial, tiene que ser un beneficio final. Y si se reducen costos y esa reducción se invierte en el bienestar de los empleados, es ideal”, indicó.

Para que clínicas y hospitales desarrollen modelos propios de IA, protegiendo los datos y aprovechando capacidades locales, el reto más importante es capacitar a todo el personal que apoye el proceso, como expertos en datos e IA, médicos que entiendan el lenguaje de la IA y representantes de los comités de ética. Una vez se cuente con el personal calificado, se debe invertir en tecnología y computadores que permitan crear modelos inteligentes propios. Tiene que ser un trabajo interdisciplinario del equipo clínico y asistencial, administrativo y tecnológico, que integre además la parte ética.

Para gestionar y utilizar los datos en pro del paciente y la institución, explicó Solarte Pabón, lo primero es garantizar la seguridad de estos; es decir, que la gente sepa que hay una

estructura segura de salvaguarda de la información. Debe garantizarse una estrategia de gobernanza de datos con unos responsables de su manejo, para verificar la calidad de esos datos y poder modelarlos.

El experto afirma que en el sector salud los datos son muy importantes porque se trata de información real de los pacientes: “Esos datos reales son los que tienen que aprovechar clínicas y hospitales para hacer estudios que apoyen la labor médica. Por ejemplo: un especialista, un oncólogo, un cardiólogo, conoce muy bien sus 10 o 20 pacientes, pero no conoce toda la población de la ciudad o de otras ciudades. Si aprovechamos la IA, ese doctor tendría acceso a todo el conocimiento existente en las historias clínicas. Es darles a los médicos la oportunidad de acceso a un conocimiento más grande”.

Asimismo, recomienda que una de las primeras acciones de las instituciones de salud es integrar sus bases de datos en una bodega, para tener una visión más global. En los hospitales tienen varios sistemas con bases de datos separadas que no se comunican entre sí. “Para que un hospital empiece a hacer modelos de IA, sobre todo modelos predictivos, el primer paso es integrar todos esos datos, de manera que se pueda conocer el perfil de un paciente desde un solo acceso, para tener una visión global del paciente. En general, muy pocas empresas en Colombia integran datos; tienen sistemas dispersos, en los últimos 20 años crearon



sistemas dispersos que no se hablan”, indicó.

Solarte Pabón señala que el error más común al integrar datos es querer integrar todo el sistema de la organización en un solo intento: “Los datos se integran por procesos hospitalarios, misionales, críticos. Se podría empezar a integrar datos de Urgencias, luego de Consulta Externa y luego de Hospitalización. Tiene que hacerse de forma progresiva; mirar el hospital como un conjunto de procesos separados y tomar cada uno de estos. Están separados pero interconectados. Y así deben quedar los datos, procesos independientes pero interconectados con los demás”.

Advierte que a la IA no se le pueden aplicar datos del exterior para tomar decisiones en hospitales colombianos. Por ejemplo, un médico le puede pedir un consejo a ChatGPT, que seguramente tiene muy poca información de Colombia, donde los pacientes tienen características diferentes a



Para implementar e integrar la IA a la gestión hospitalaria, es necesario tener conocimiento técnico avanzado en uso de la IA, en gestión y seguridad de datos. interdisciplinaria.

los de Norteamérica, África o Asia. Entonces, si el primer paso fue integrar los datos, el segundo es adaptar los módulos de IA a las condiciones propias de nuestro país.

El tercer paso es establecer criterios de seguridad en los datos, alerta Solarte Pabón: “No se deberían crear modelos predictivos o sistemas inteligentes mediante los cuales nuestros datos salgan de los hospitales y se vayan a la Internet, donde no hay forma de controlar qué pasa con esos datos. Los datos son lo más importante y los debemos resguardar. Al igual que la vida de los pacientes es lo más importante, garantizar la seguridad de los datos de esos pacientes es lo más importante. Hay que establecer una estructura que garantice que los datos sean seguros y aprovechables, y formar a las personas de la organización en gobernanza y seguridad de los datos”.

Además, debe impartirse formación en uso responsable de la IA y en ética: “Hay que formar en ética y responsabilidad social con los datos. Puede darse que un médico por tener mucho trabajo no lea una historia clínica, se le pasa a ChatGPT y le pide su opinión, y con esa opinión hace un diagnóstico y formula recomendaciones. Eso no debe pasar; la IA solo debe ser apoyo, al final quien toma la decisión debe ser el médico porque ese paciente puede demandar al médico o a la institución y no al ChatGPT. Hay que formar en ética sobre uso y peligros de la IA”, señala el experto.

Para implementar e integrar la IA a la gestión hospitalaria, es necesario tener conocimiento técnico avanzado en uso de la IA, en gestión y seguridad de datos. También se requiere que las personas tengan formación en ética, en responsabilidad social con esos datos y

habilidades de trabajo en equipo para resolver problemas desde una visión interdisciplinaria. Indica Solarte Pabón: “Se necesitan habilidades de ética en el sentido de buscar el bienestar de los otros, porque puedo hacer modelos inteligentes que funcionen, pero debe importarme que no haga daño, porque la IA es una tecnología poderosa que si cae en manos equivocadas puede causar mucho daño”.

Alertó, además, sobre el peligro de la deuda cognitiva: “Si las personas le delegamos todo a la IA, en diez años nuestro cerebro se habrá dormido y la IA nos quitará el trabajo si perdimos todas las capacidades cognitivas, y será más inteligente la máquina que la persona”.

Riesgos y mitigación en el manejo de datos con IA en salud

Según Solarte Pabón, se presentan varios riesgos con la implementación de la IA en el sector: (1) que terceros se apoderen de los datos de los pacientes del hospital o la clínica y los aprovechen para su beneficio particular; (2) tomar decisiones teniendo en cuenta solo lo que dice la IA, sin considerar el criterio médico y delegando la responsabilidad a una IA, y (3) que las personas pierdan habilidades cognitivas y habilidades de pensamiento crítico.

Para mitigar estos riesgos es preciso capacitar a las personas, tanto en herramientas de IA como en sus riesgos y peligros y cómo evitarlos. Las instituciones deben gestionar y generar lineamientos del uso responsable de la IA, para que las personas sepan qué cuidados deben tener. Dice el experto: “En el sector salud es más relevante tomar precauciones porque los datos son muy sensibles. Los médicos toman decisiones de vida o muerte, y una mala decisión puede representar la muerte de un paciente. No se puede dejar que una IA tome las decisiones sobre un paciente”^[1]