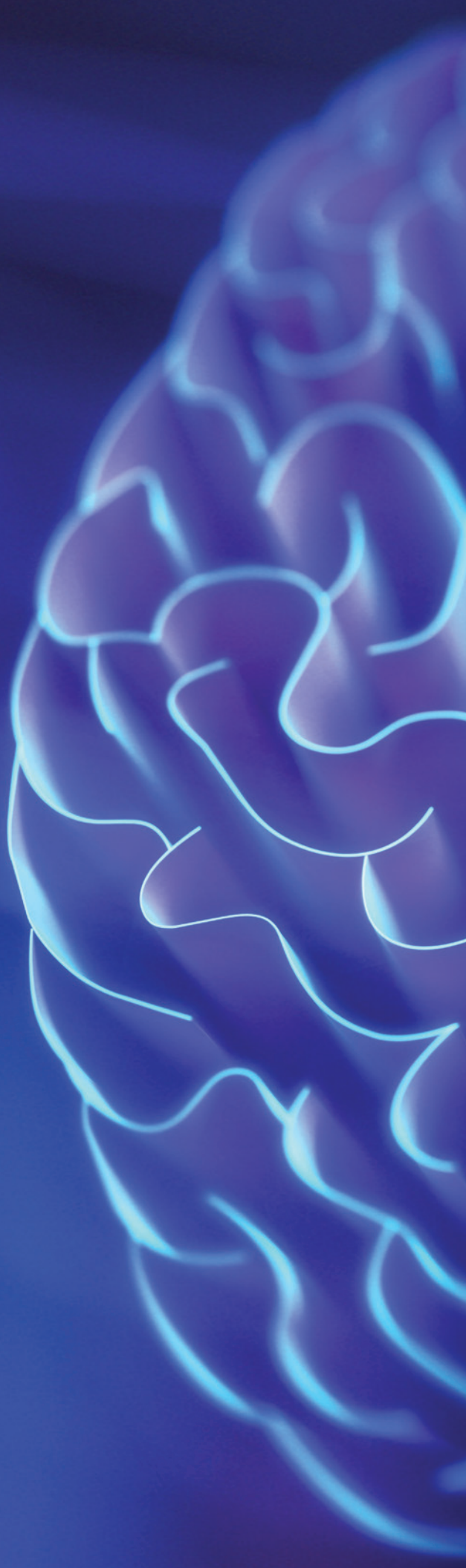
KPMG

¿Cómo controlar la inteligencia artificial (IA)?

El imperativo de la transparencia y la claridad



Octubre 2020

kpmg.com.ar

Nuestros autores

Martin Sokalski
Responsable, Advisory, Riesgos de las Tecnologías Emergentes,
KPMG en Estados Unidos

Martin Sokalski es Líder Global de la práctica Riesgos de las Tecnologías Emergentes de KPMG en Estados Unidos. Asiste a las organizaciones de todo el mundo en la adopción de un enfoque basado en el "arte de lo posible", impulsado por las tecnologías emergentes, tales como la inteligencia artificial, para facilitar el desarrollo de ideas, la innovación y una adopción responsable.

A través de los años, ha colaborado con numerosas organizaciones de diversos sectores en la evaluación, el diseño y la implementación de nuevos modelos operativos y de gobierno digitales, para ayudarlas a alcanzar los resultados comerciales esperados, a la vez que incorporan los imperativos clave en materia de gobierno corporativo, confianza y valor. Regularmente, Martin asiste en calidad de orador a diferentes conferencias y contribuye al liderazgo innovador en inteligencia artificial, transformación digital y tecnologías emergentes.

Martin cree que, en la actualidad, la adopción de la inteligencia artificial a gran escala se ve obstaculizada por la falta de confianza, transparencia, claridad y la existencia de prejuicios, y planea trabajar con los líderes de la industria para superar este desafío.

Professor Dr. Sander Klous
Socio, Líder de Data & Analytics,
KPMG en los Países Bajos

Sander Klous es líder en Data & Analytics en los Países Bajos y profesor en ecosistemas de Big Data para empresas y la sociedad en la Universidad de Ámsterdam. Cuenta con un Doctorado en Física de Altas Energías (HEP) y trabajó por más de diez años en una serie de proyectos para CERN, el instituto de Física más grande del mundo, en Ginebra.

Su libro más vendido, We Are Big Data, obtuvo el segundo lugar en el premio al libro del año sobre gestión en 2015.

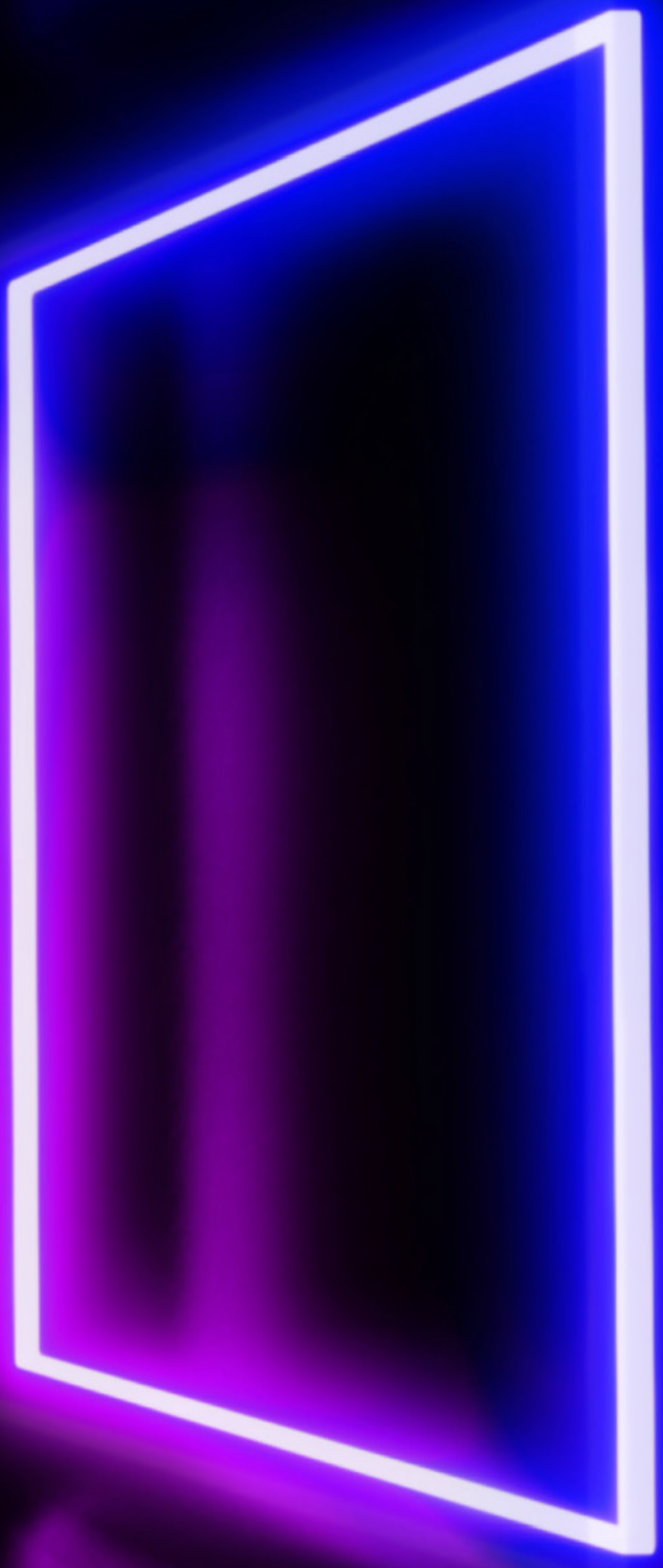
Su nuevo libro, Building Trust in a Smart Society, se encuentra entre los libros con mayor nivel de ventas en los Países Bajos. Sander cuenta con vasta experiencia en computación distribuida a gran escala, sistemas en tiempo real y tecnologías de procesamiento de datos. Actualmente, su trabajo se centra en el uso generalizado de análisis fiables, algoritmos éticos y herramientas confiables de analytics, de una forma que los clientes y la sociedad valoren.

Swami Chandrasekaran
Director Ejecutivo Innovación y Soluciones Empresariales,
KPMG en Estados Unidos

Swami Chandrasekaran es líder en Innovación y Soluciones Empresariales de KPMG y ayuda a liderar la arquitectura, tecnología y desarrollo de la automatización inteligente, incluida la inteligencia artificial y las tecnologías emergentes. Lideró el desarrollo, el diseño y la creación de diversos productos y soluciones de inteligencia artificial, sorteando una amplia gama de desafíos, en áreas como Impuestos y Auditoría, Automatización Industrial, Seguridad Aérea, Centros de Atención al Cliente, Reclamos de Seguros, Servicios de Campo, Enriquecimiento Multimedia, Trabajo Social, Marketing Digital, Fusiones y Adquisiciones Swami también cuenta con experiencia en la automatización de procesos comerciales, así como en la integración de sistemas y datos. Swami es Ingeniero Distinguido de IBM.

Recientemente, publicó Learning to Build Apps Using Watson AI y luego presentar 20 solicitudes de patentes, principalmente en el área de la inteligencia artificial, de las cuales 17 fueron otorgadas, fue reconocido como Inventor Líder de IBM.

Contenido



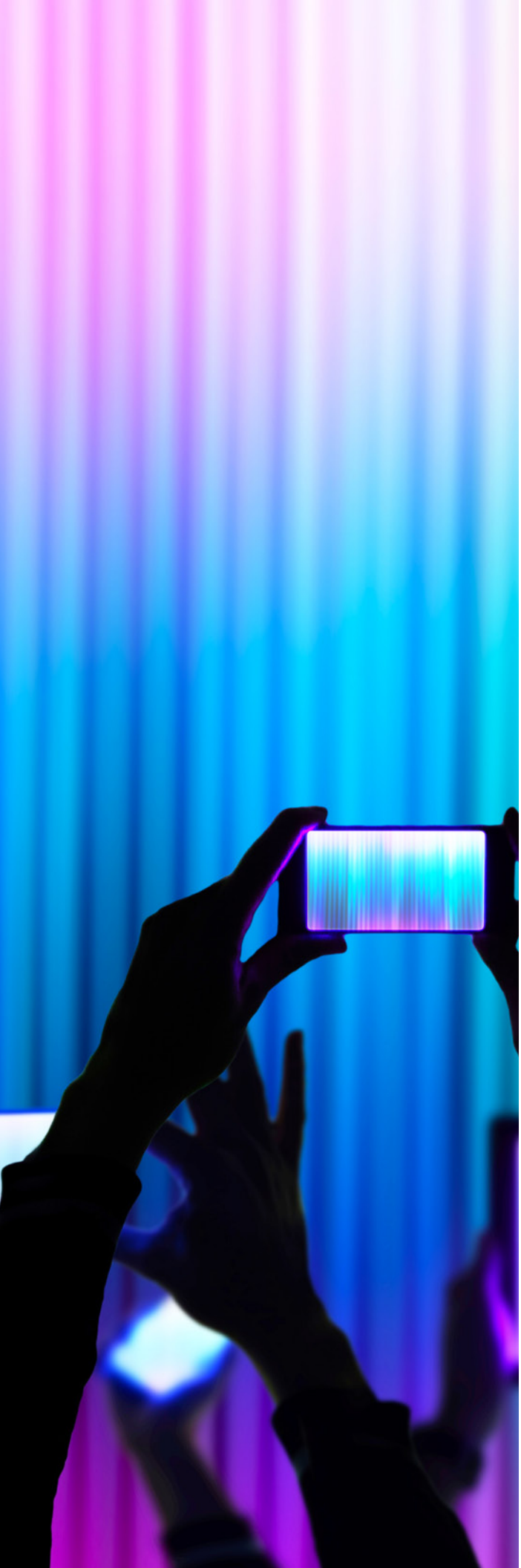
Introducción

Las tecnologías que generaron un cambio en el mundo a lo largo de la historia de la humanidad tienen un elemento en común: el control.

El vapor y la luz, y una larga lista de inventos y tecnologías, surgieron porque pudimos guiar las fuerzas naturales hacia un poder transformador.

La aviación no existiría si no hubiéramos dominado el vuelo.

La IA tiene potencial para cambiar el mundo.



Pero desconocemos todo el impacto que la IA puede generar en el mundo.

Y, como toda otra tecnología transformadora, solo es posible conocer el potencial de la IA si comprendemos y controlamos su composición y funcionamiento. Por esta razón, las compañías deben establecer una política general de gestión de la IA, con un enfoque centrado en aprovechar responsablemente el potencial de estas tecnologías.

La IA revela un mundo oculto detrás de la complejidad. La información obtenida a partir de algoritmos que evolucionan constantemente está cambiando nuestros negocios y nuestras vidas. Muchos científicos proyectan un futuro en el que podrán resolverse algunos de los misterios más profundos y los problemas más difíciles que enfrenta la humanidad. Ya estamos viendo algunos de sus beneficios: desde algoritmos que descubren partículas subatómicas y permiten identificar agujeros negros, hasta su aplicación a nivel empresarial, donde a partir de herramientas sofisticadas de data & analytics, impulsadas por la IA, se toman decisiones críticas que afectan el negocio y la marca, así como la salud y la seguridad de los consumidores.

.

IA en la práctica.

Imaginemos a un responsable de una línea de negocios de préstamos personales en una importante entidad financiera. Se genera una situación que involucra actos de discriminación y prejuicio, junto con uno o dos titulares en las noticias. Durante una reunión del Directorio, los miembros piden explicaciones a este líder y al Director Ejecutivo de Sistemas respecto de la decisión y la lógica de rechazar el otorgamiento de préstamos a aquellos solicitantes de un determinado rango de edad o grupo racial. Lo que está en discusión es un algoritmo de inteligencia artificial que produjo esos resultados o intensificó una decisión de los ejecutivos de préstamos del área. El problema para estos dos líderes es el siguiente: nadie puede explicar exactamente por qué el algoritmo hizo lo que hizo.

Situaciones como estas se están dando en diversas áreas del sector público y privado: selección de personal, transporte, marketing, cuidado de la salud, procesos de admisión a la universidad, vivienda, y gestión de ciudades inteligentes.

Toda organización que desarrolle o adopte tecnologías de avanzada de aprendizaje continuo accede a un nivel de conocimiento y toma de decisiones que supera ampliamente las capacidades de la mente humana. Es una gran oportunidad.

Sin embargo, los algoritmos pueden ser destructivos cuando producen resultados inexactos o sesgados. Esto, sumado a la complejidad inherente a la tecnología, aumenta la preocupación de aquellos líderes que desean confiar en el uso de estos algoritmos. Por esta razón, en medio del gran entusiasmo en torno a la IA, existen dudas respecto a la posibilidad de delegar ciertas decisiones a las máquinas, sin certezas acerca de cómo se toman esas decisiones y si resultan justas y adecuadas. Hay una brecha de confianza.

Generar confianza.

Los grandes beneficios de la IA surgirán cuando los algoritmos puedan explicarse (y, por lo tanto, comprenderse) en un lenguaje sencillo y para todos. La brecha de confianza existe por la falta de transparencia de la IA: hay un temor asociado a los aspectos desconocidos en torno a esta tecnología.

Generar confianza implica, también, comprender y proteger el linaje de los modelos de IA (y a los datos que los forman) de distintos tipos de ataques adversarios y usos no autorizados.

Las decisiones de negocios críticas tomadas mediante IA afectan la marca (y la confianza de los consumidores en la marca), y pueden tener un gran impacto en el bienestar y la seguridad de los consumidores y los ciudadanos. Nadie quiere decir "porque la máquina lo dijo". Nadie quiere malinterpretar a la IA. .

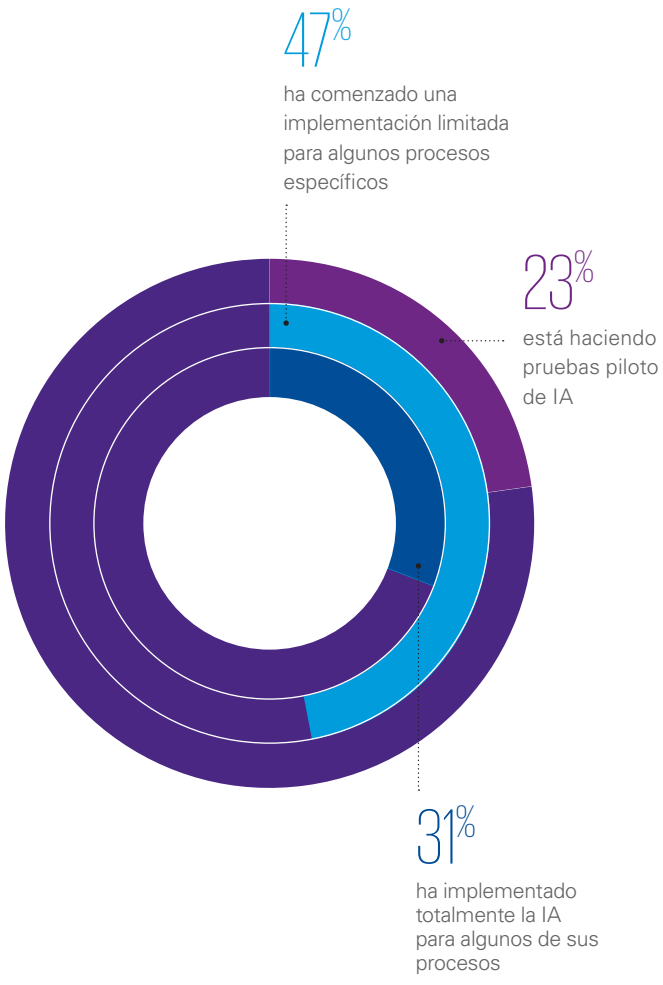
Superar la brecha de confianza.

En la actualidad, alcanzar una IA objetiva y clara es algo más que una tarea difícil para los ejecutivos y el Directorio: es una necesidad.

Por ejemplo, según la edición 2019 del informe CEO Outlook de KPMG1, el 66 % de los líderes encuestados desestimaron información proporcionada por un análisis de datos basado en informática porque era contraria a su experiencia o intuición.

Para la mayoría de las organizaciones, la IA aún está, por así decirlo, en una etapa experimental y se aplica a nivel funcional, pero sin ser una parte integrante del proceso de toma de decisiones en los negocios. Sin embargo, esto está cambiando rápidamente.

Según la edición 2019 del informe CEO Outlook de KPMG en Estados Unidos, las organizaciones se encuentran en diferentes etapas de implementación de la IA.



[1]Informe CEO Outlook 2019 de KPMG: "Agile or Irrelevant: Redefining Resilience", junio de 2019

¿Cuál es la solución?

Para que la IA avance hacia el bien común y que los líderes asuman responsabilidad y rindan cuentas por los resultados, es esencial establecer un marco (impulsado por métodos y herramientas) para facilitar una adopción responsable y a escala de la IA.

Este informe está diseñado para líderes que operan en el mundo de los algoritmos de la inteligencia artificial y el aprendizaje automático.

El imperativo de negocios y cumplimiento para comprender y confiar en las tecnologías de IA ha alcanzado un nivel crítico.

Este documento explica cuál es la urgencia y describe los métodos y herramientas que pueden ayudar a los líderes a gestionar sus programas de IA.

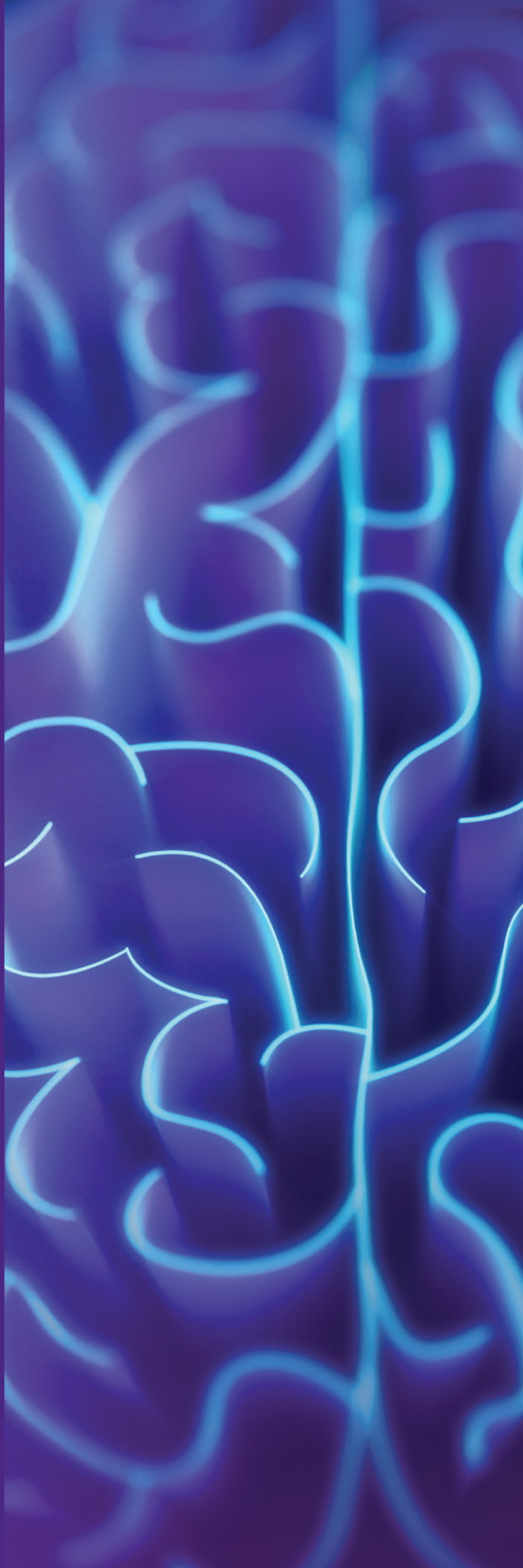
"En lo que respecta a la inteligencia artificial, el verdadero arte de los posible se desplegará cuando exista mayor confianza y transparencia. Para lograrlo, es necesario incorporar los pilares básicos de los programas de IA: integridad, claridad, imparcialidad y resiliencia."

Martin Sokalski
Principal, Advisory, Emerging Technology Risk Services
KPMG in the U.S.

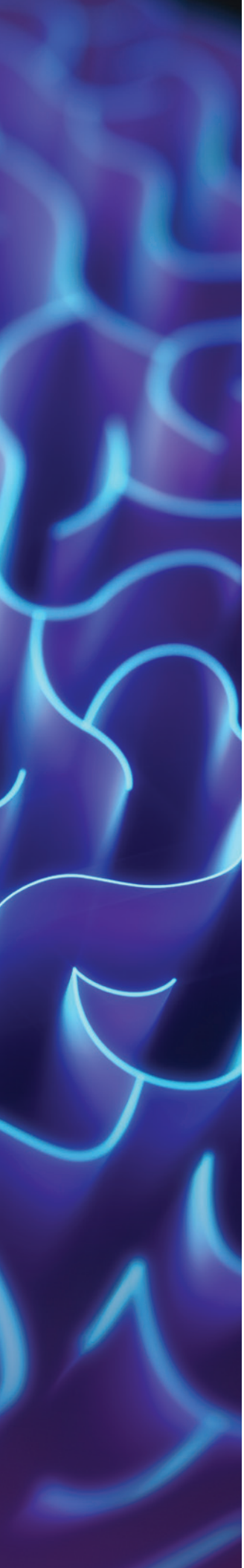
Principales desarrollos

En base a las entrevistas con ejecutivos que impulsan la estrategia de IA en las grandes compañías, hemos recibido un mensaje consistente.

Muchas compañías recién están comenzando a invertir en marcos de control de IA, en comparación con otras prioridades de implementación de IA[2].



[2]Estudio de KPMG de 2019 sobre la adopción de IA a nivel empresarial



Ganar confianza en la IA es uno de los principales objetivos de los líderes.

El 45 % de los ejecutivos encuestados dijeron que fue difícil o muy difícil confiar en los sistemas de IA[3].

Nuevas iniciativas sobre políticas y normas en torno a los datos y la IA marcan el fin de la autorregulación y el surgimiento de un nuevo modelo de supervisión[4].

La mayoría de los líderes no tienen en claro cuál debería ser el enfoque para abordar la IA.

Alrededor del 70 % afirmó no saber cómo manejar algoritmos[5].

Las compañías están teniendo dificultades en decidir quién se responsabilizará por los programas y los resultados de IA.

Durante las entrevistas, escuchamos que la mayoría de las compañías están intentando determinar quién tendrá a su cargo el desarrollo de la IA. Algunas compañías han designado una autoridad central a través de un consejo de IA o Centro de Excelencia; otros han asignado esa responsabilidad a diferentes líderes como el Director de Tecnología o el Director de Información.

Un marco que incluya métodos basados en la tecnología puede ayudar a abordar riesgos inherentes a la IA y cuestiones éticas relacionadas.

El objetivo es ayudar a los usuarios a controlar sus programas de IA al instaurar cuatro pilares de confianza: integridad, claridad, equidad y resiliencia.

[3]Forrester Research, Q2 2018 Global AI Online Survey
[4]AI, Internet Policy Proposals Signal Shift Away From Self-Regulation. Wall Street Journal, WSJ Pro: Artificial Intelligence, 9 de april de 2019
[5]Fuente: KPMG, Why AI Must Be Included in Audits, 2018

La necesidad de saber: pilares de confianza

El costo de equivocarse con el uso de la IA va más allá de las finanzas (pérdida de ingresos, multas por incumplimiento), e incluso pueden surgir implicancias éticas, de marca y de reputación.

Imaginemos que un líder de CDO y LOB intenta explicar el resultado de un solo modelo al directorio. Los niveles de responsabilidad van desde los altos cargos directivos y el dueño de la línea de negocio de tarjetas de crédito de un banco (alguien que asume la responsabilidad por todo lo relacionado con este negocio, incluidos los modelos de IA) hasta el nivel del cliente, con un oficial de préstamos que puede tener que asumir la responsabilidad y que, de muchas maneras, representa la marca.

Las decisiones comerciales clave a escala tienen un efecto determinante en el éxito, por ejemplo:

¿Debería la división aprobar una tarjeta de crédito para un cliente?

Entre las decisiones que se deben tomar por cada cliente se encuentran: la tasa anual, el límite de gasto y una larga lista de otros factores. Los modelos de aprendizaje automático suelen tomar estas decisiones por millones de clientes. Es válido afirmar entonces, dada la escala, que el negocio está en manos de un puñado de científicos de datos inteligentes, y de las máquinas que ellos mismos construyen y entrenan, para que utilicen datos reales creados a partir de datos históricos sobre préstamos.

Algoritmos autónomos: antes y después.

La mayoría de los algoritmos actuales son relativamente simples y deterministas: Producen el mismo resultado a partir de un conjunto predeterminado de estados y un número fijo de reglas. Los enfoques para evaluar su validez e integridad están ampliamente establecidos y aceptados. De hecho, en nuestra estimación, se conocen más del 80% de las prácticas líderes necesarias para mantener su precisión y eficacia.

Pensemos en sistemas expertos en fabricación. Pensemos en la ciencia actuarial que utiliza reglas deterministas o tablas de decisión para seguros. Pensemos en la automatización de procesos en los servicios financieros.

No es tan difícil determinar si las conclusiones a las que llegan son aceptables, y es relativamente fácil lograr una supervisión sólida y escalable.

Estas reglas pueden volverse muy complejas, especialmente cuando aumenta el número de atributos (también conocidos como características o variables) en los datos o cuando la cantidad de registros aumenta.

El aprendizaje automático y el aprendizaje profundo, y demás tipos de inteligencia artificial, son algo totalmente diferente. Están entrenados para aprender de los datos (comúnmente conocidos como la verdad fundamental) en lugar de ser programados explícitamente, lo que significa que pueden “comprender, aprender y descubrir” matices y patrones en los datos, pueden manejar un conjunto muy grande de atributos y, a menudo, son significativamente más complejos en cuanto a cómo hacen lo que hacen.

Pensemos en un modelo de predicción que se entrena a partir de un conjunto de un millón de solicitudes de préstamos, que a su vez utiliza 100 atributos. Pensemos en detectar un tumor a partir de un millón de imágenes de resonancia magnética. Pensemos en clasificar correos electrónicos.

Una vez entrenados y evaluados, estos modelos pueden recibir datos nuevos o no vistos con anterioridad a partir de los cuales pueden hacer predicciones. Son de naturaleza probabilística y responden con cierto grado de confianza.

Si bien todos estos aspectos son buenos, puede que no esté claro qué es lo que están haciendo los modelos: qué aprenden, particularmente cuando emplean técnicas de aprendizaje profundo como las redes neuronales, cómo se comportarán o si desarrollarán prejuicios con el tiempo a medida que continúan evolucionando. De ahí la importancia de comprender qué atributos de los datos de entrenamiento influyen en las predicciones del modelo.

Riesgo algorítmico: Confianza en la máquina

Analicemos en detalle un problema potencial para el CDO y el dueño de la división de préstamos de una gran empresa financiera.

Si un error se escondiera en un algoritmo (o en los datos que alimentan o entrenan al algoritmo), ello puede influir en la integridad y en la equidad de la decisión que la máquina ha tomado. Esto podría incluir datos contradictorios o el enmascaramiento de datos como verdad fundamental. Los líderes empresariales están comprometidos a preservar la reputación de la marca de la empresa, incluso cuando los modelos de IA toman cada vez más decisiones que pueden no ser entendidas o no están en línea con las políticas corporativas, los valores corporativos, las pautas y las expectativas del público. Multipliquemos estos problemas por la cantidad de algoritmos que utiliza la división de préstamos. Es entonces que la confianza se debilita o desaparece.

Se han propuesto varias técnicas, incluidas las basadas en la teoría de grupos de renormalización[6]. A medida que los modelos de IA, como la visión por computadora, el reconocimiento de voz y el procesamiento del lenguaje natural, se vuelven más sofisticados y autónomos, asumen un mayor nivel de riesgo y responsabilidad. Cuando se suspende el entrenamiento por largos períodos de tiempo, las cosas pueden salir mal: el tiempo de ejecución avanza con dificultades, los conceptos quedan a la deriva y problemas como los ataques adversarios pueden comprometer lo que aprenden estos modelos.

Imaginemos resonancias magnéticas comprometidas o semáforos manipulados en una ciudad inteligente.

Los algoritmos de aprendizaje continuo también introducen un nuevo conjunto de consideraciones de ciberseguridad. Los primeros implementadores siguen luchando con la magnitud de los riesgos que presentan estos temas en el negocio.

Entre los riesgos se encuentran los ataques adversarios que afectan la base misma de estos algoritmos al envenenar los modelos o alterar los conjuntos de datos de entrenamiento, lo que potencialmente compromete la privacidad, la experiencia del usuario, la propiedad intelectual y cualquier otro aspecto comercial clave. Consideremos cómo las vidas de las personas o su entorno podrían verse afectados ante un ataque adversario a dispositivos médicos o sistemas de control industrial.

La manipulación de datos podría perturbar las experiencias de los consumidores al proporcionar sugerencias inapropiadas en los servicios minoristas o financieros. Estos ataques podrían, en última instancia, erosionar la ventaja competitiva que los algoritmos pretendían crear.

Con algoritmos complejos de aprendizaje continuo, los seres humanos necesitan conocer más que solo datos o atributos y sus respectivos pesos para darse cuenta de lo que significa que la IA se equivoque o se vuelva deshonesta; necesitan comprender aspectos tales como el contexto y el propósito por el cual se desarrolló el modelo, quién los capacitó, la procedencia de los datos y cambios realizados, y cómo los modelos fueron (y son) servidos y protegidos. Y deben comprender qué preguntas hacer y qué indicadores clave buscar en torno a la integridad, claridad, equidad y resiliencia de un algoritmo.

El desafío No. 1 para los que adoptan la IA es la calidad de los datos. El director de tecnología de una agencia gubernamental indicó específicamente en nuestra encuesta global sobre IA que si no pueden confiar en los datos, no pueden utilizar la IA[7].

[6]An Exact Mapping Between the Variational Renormalization Group and Deep Learning, Pankaj Mehta, David J. Schwab. 2014

[7]Forrester Research, Q2 2018 Global AI Online Survey

El pilar de la confianza.

Cuando se analizan todas las acciones y capacidades necesarias para garantizar la confianza en los algoritmos y modelos, y por lo tanto en la marca, KPMG cree que surgen cuatro dimensiones.

Integridad del algoritmo.

Se trata de inspeccionar las "estructuras" del algoritmo en busca de fallas en su integridad. Lo que los líderes necesitan saber es: la procedencia y el linaje de los datos utilizados para el entrenamiento, los controles sobre el entrenamiento y la construcción del modelo, las métricas de evaluación del modelo, el mantenimiento de principio a fin, y la verificación de que ningún cambio comprometa el objetivo o la intención original del algoritmo.

La clave también sería el monitoreo continuo de las métricas de desempeño del modelo, incluida la detección de desviaciones del concepto.

Claridad.

Comprender las razones por las que un modelo hizo una predicción, y ser capaz de interpretarlas, es esencial para confiar en el sistema, especialmente si uno tiene que tomar una acción basada en esos resultados probabilísticos. Esta es una capacidad subjetiva de la IA. Ser capaz de explicar por qué y cómo un modelo que produce un resultado (conocimiento, decisión) depende de la definición de éxito establecida y de la gestión general del algoritmo, a partir de una verdad fundamental que sea limpia, suficiente y apropiada para la evaluación continua de los resultados. Existen varios enfoques, incluido LIME, una técnica de explicación que se centra en los aspectos locales o aislados de la toma de decisiones[8] y el programa de IA explicable (XAI) de la Agencia de Proyectos de Investigación Avanzada de Defensa (DARPA), que tiene como objetivo crear un conjunto de técnicas de aprendizaje automático que generen modelos más explicables, con una interfaz de explicación.

Equidad: ética y responsabilidad.

No es posible confiar en la inteligencia artificial y los algoritmos si éstos no son justos. Para que sean justos, deben estar diseñados y construidos lo más libre de prejuicios posible, y deben mantener la equidad a medida que evolucionan. Los atributos utilizados para entrenar algoritmos deben ser relevantes, apropiados para el objetivo y su uso debe estar permitido. En algunos casos, sin embargo, la información personal es relevante para el modelo, como en la atención médica, cuando el género o la raza pueden ser una parte fundamental de los estudios o del tratamiento. Es preciso supervisarlos y gestionarlos cuidadosamente para asegurarse de que no se utilicen datos sustitutos para entrenar un modelo. Un código postal, por ejemplo, puede ser un sustituto de la etnia o los ingresos y generar inadvertidamente resultados sesgados y riesgos posteriores (solo uno de ellos representa una violación a la normativa). Deben aplicarse técnicas para comprender qué prejuicios son inherentes a ciertos datos, y mitigarlos mediante enfoques como el reequilibrio, la reponderación o la eliminación del prejuicio.

Las herramientas de monitoreo y supervisión continua son esenciales para ayudar a garantizar que los modelos que se entrenan continuamente con datos de uso y retroalimentación no provoquen sesgos durante el tiempo de ejecución.

Resiliencia.

Aquí es dónde hablamos de la solidez y la resiliencia de los modelos o algoritmos que se desarrollan o que se utilizan. Los modelos que se utilizan se exponen generalmente como API o se incorporan en las aplicaciones, y deben ser portátiles y funcionar en diferentes y complejos ecosistemas.

La IA resiliente debería cubrir todos los aspectos de una adopción segura y abordar los riesgos holísticamente a través de una arquitectura de diseño firme y la detección de anomalías mediante el uso de los conceptos de la IA, tales como las redes generativas antagónicas, que enfrentan a los algoritmos entre ellos para obtener mejores y más variados resultados. El objetivo es permitir que todos los componentes estén adecuadamente protegidos y monitoreados. ¿Por qué? Las circunstancias externas pueden generar errores cuando los algoritmos son incapaces de corregir o compensar el dato que es impreciso o anómalo. Proteger los datos de uso y de retroalimentación que podrían ser utilizados para capacitar continuamente a los modelos es crítico. Las acciones básicas incluyen el monitoreo continuo de los límites de los modelos y el control de acceso a los modelos.

Queda por responder una pregunta clave: ¿Quién entre los humanos se responsabiliza de los resultados de la IA?

La responsabilidad es un tema de gobierno clave que debe establecerse en todas las iniciativas de IA, respecto de cada uno de los modelos individuales. Observamos variaciones significativas entre los participantes del *2019 Enterprise AI Adoption Study* de KPMG en torno a la asignación de autoridad y responsabilidad. Algunas organizaciones establecieron una autoridad centralizada como. por ejemplo, el Consejo de IA; otros le asignaron funciones tales como la dirección ejecutiva de tecnología o de información. Pero pocas organizaciones cuentan con prácticas de responsabilidad sólidas; una brecha de liderazgo que puede menoscabar la confianza internamente y en relación con los interesados externos. Una razón importante en relación con esta falencia: la mayor parte de las organizaciones no cuentan con las herramientas y los conocimientos para obtener un entendimiento íntegro e introducir transparencia en sus algoritmos.

[8]Marco Tulio Ribeiro, Sameer Singh, Carlos Guestrin, "Why Should I Trust You? Explaining the Predictions of Any Classifier." 2016

Gobierno y ética de IA

El gobierno y la ética se convierten en el 'cómo' de la adopción responsable de la IA al considerar el riesgo de que los algoritmos complejos puedan corromperse.

Miren las normas y reglas que gobiernan a la industria de la aviación, y las prácticas recomendables internas que dominan los procedimientos de cada aerolínea individual, desde los ejecutivos hasta la cabina del piloto. Miren la confianza que cada uno de los participantes involucrados – la tripulación, los pasajeros y las empresas que transportan activos valiosos por aire – depositan en la experiencia. Esto es a lo que la industria debe apuntar en términos de IA.

Llegamos a un punto clave en términos de la necesidad de un gobierno efectivo y de la adopción responsable y grado de la IA. En muchos casos, las organizaciones desarrollan políticas internas y funciones de gobierno para supervisar los temas que se relacionen con la IA en un esfuerzo por generar confianza y transparencia en toda la empresa y en los grupos de interés externos, incluso los consumidores.

En el Reino Unido y en la Unión Europea, con su Normativa General sobre Protección de Datos, la marea se mueve hacia el establecimiento de la supervisión. Y la oportunidad es importante, ya que las semillas de la IA están en tierra firme y creciendo. El grado todavía no se determina, pero estas tecnologías están preparadas para expandirse dentro de la empresa y en todos los sectores de la industria y asumir mayor autonomía y responsabilidades. Este es el momento de establecer un marco de gobierno y ética en torno a los anclajes de la confianza.

El control de la IA ayudará a lograr una expansión responsable del poder.

Gobierno y ética.

Evaluar y asegurar los anclajes de la confianza de la IA pueden derivar de un conjunto nuevo de prácticas y métodos líderes orientados a mantener el control sobre la IA y los algoritmos de aprendizaje automático. Una estrategia de gobierno efectivo sienta las

bases de la confianza y la transparencia al implementar los mecanismos y las herramientas que medirán a la IA de manera continua. Los líderes podrán tomar decisiones sobre la base de la información y sus organizaciones construirán una cultura de responsabilidad que es más fuerte y más conscientemente representativa de la brújula ética de la organización.

Gobierno.

Una larga lista de preguntas emerge cuando nos adentramos en los trabajos de IA y muchos de ellos son problemáticas humanas. ¿Por qué y cómo se seleccionaron determinados casos de uso como candidatos de IA? ¿Por qué el equipo eligió esas particularidades (y excluyó lo que excluyó? ¿Cómo medimos y demostramos el éxito (o cómo explicamos los fracasos)? ¿Por qué el algoritmo hizo lo que hizo y quién es el responsable del resultado? La expresión: “Porque los algoritmos lo dicen” no funcionará para los líderes y el público en general, ya que los sistemas son cada vez más poderosos e integradores.

La necesidad: Considerar el contexto en general y establecer un rumbo y una cultura claros al comienzo.

Si no cuentan con un modelo de gobierno o un modelo operativo para IA, será difícil lograr los resultados de negocios deseados o tener confianza en la integridad, la posibilidad de explicar, la razonabilidad o la resiliencia de la IA. El gobierno de la IA es lo correcto en términos de confianza y visibilidad. Esto significa mirar los marcos y el gobierno corporativo a través de una nueva lente en torno a la gente, los procesos y la tecnología durante todo el ciclo de vida: desde las primeras etapas del modelo hasta la estrategia, puesta en marcha, monitoreo, capacitación y funcionalidades, y medición continua

.

Ética.

Este es un tema muy importante en relación con la IA tanto en términos de problemas y dilemas que enfrentan las empresas y la sociedad como en relación con las medidas y resguardos necesarios para controlar la IA. La ética y la confianza están entrelazadas. Y ambas son el combustible necesario para que la IA se mueva de manera de beneficiar a la sociedad en su conjunto. Se implementan resoluciones y normativa.

La normativa general sobre protección de datos es un ejemplo claro; y otras están estableciendo el contexto, por ejemplo, las pautas éticas para la inteligencia artificial confiable recientemente emitida por la CE[9].

Otro aspecto de la ética es la autonomía personal. ¿Qué decisiones podemos delegarle a las máquinas y que decisiones deberían permanecer en el ámbito humano? Esta es una parte vibrante de la discusión dentro de la comunidad científica y en los gobiernos. No sorprende que los Ministros de la CE hayan recientemente declarado que la IA y las tecnologías de aprendizaje automático "no deben utilizarse para influir indebidamente o manipular la conducta y la forma de pensar de las personas."[10]

La necesidad: Establecer los resguardos éticos desde las primeras etapas de un programa de IA, que requiere visibilidad –y monitoreo– de todo el ciclo de vida de la IA, desde la estrategia y la ejecución hasta el desarrollo continuo.

[9]Ethics Guidelines for Trustworthy AI. European Commission, High-Level Expert Group on AI, April 2019. https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai

[10]Council of Europe, Committee of Ministers, Declaration by the Committee of Ministers on the Manipulative Capabilities of Algorithmic Processes. February 2019 https://search.coe.int/cm/pages/result_details.aspx?objectid=090000168092dd4b

Control de la IA
Ciudad de Ámsterdam

KPMG está trabajando con la Ciudad de Ámsterdam para evaluar un servicio municipal digitalizado que les permite a los residentes realizar presentaciones en línea para temas tales como residuos en la vía pública. El algoritmo de aprendizaje automático identifica el tipo de problema, la prioridad y el servicio específico al que debería responder.

Los funcionarios de Ámsterdam utilizan el marco de control de la IA para obtener una evaluación efectiva y continua del desarrollo de las aplicaciones de IA para evitar que, inadvertidamente, utilicen patrones de aprendizaje que pudieran terminar en decisiones inadecuadas o sesgadas.

El éxito de las ciudades dependerá mayormente de cuán inteligentes y éticos sean en el manejo de los datos.

El resultado esperado

Mejorar la confianza del público en una ciudad segura y bien mantenida; ayudar a la ciudad en su misión para proteger los derechos digitales de los residentes.

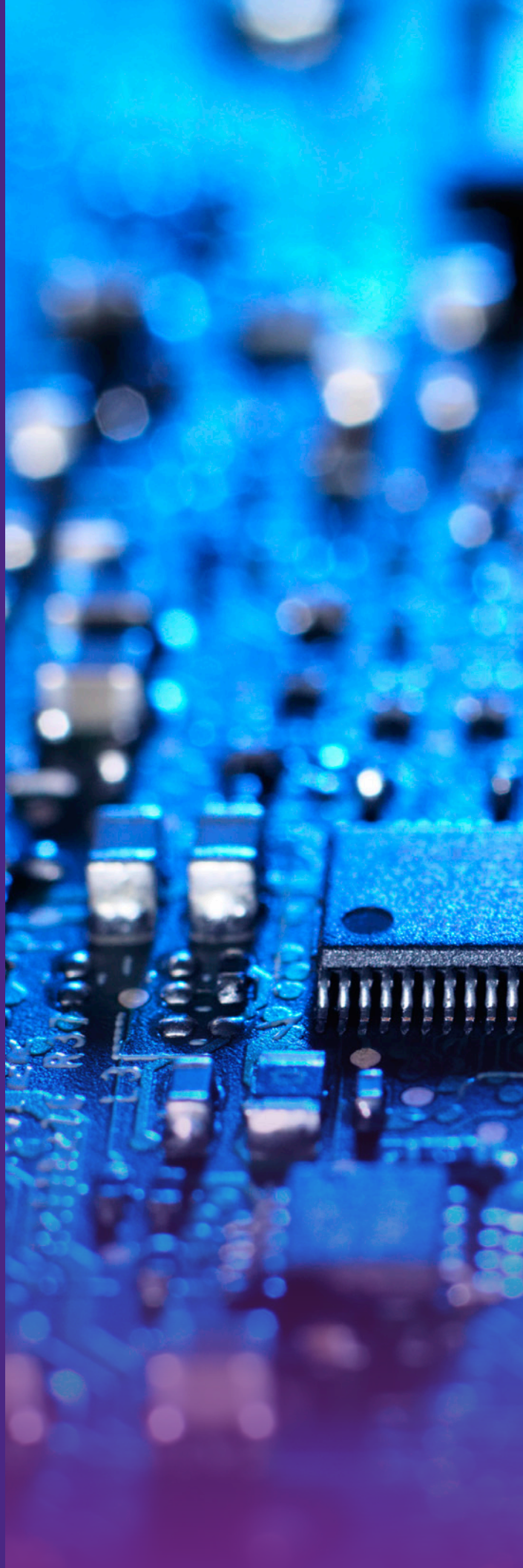
"La ciudad de Ámsterdam pretende proteger los derechos digitales de sus ciudadanos, y tenemos la responsabilidad de ser inclusivos y transparentes acerca de los algoritmos de aprendizaje automático que implementamos para darle soporte a nuestros servicios municipales y programas, por lo tanto, tratamos de desarrollar un enfoque con un socio, KPMG, para que nos ayude a desarrollar el método de clasificación para verificar y aprobar estos algoritmos."

Ger Baron

Director ejecutivo de Tecnología,
Ciudad de Ámsterdam

Claves para gestionar la Inteligencia Artificial: un marco para impulsar la transparencia

La implementación de estas normas colabora con la compañía en el gobierno de los algoritmos de manera transparente y efectiva, a la vez que se genera confianza en la calidad de las decisiones adoptadas por la compañía.



La mayoría de los líderes no saben cómo cerrar la brecha porque no saben cómo gobernar la IA y ver el contexto en el que funciona.

Lo que los líderes necesitan es una mirada a nivel que les revele tanto las métricas clave de desempeño como los indicadores de riesgo. El sesgo en los datos de capacitación es una gran preocupación. El riesgo del modelo en general es otro. Y también lo es el cumplimiento y la seguridad, y una larga lista de otros elementos. Esto nos lleva de vuelta a la escena incómoda del directorio en la introducción de este informe:

¿Cómo pudo un programa generar un resultado equivocado? En algunos casos, puede deberse a errores subyacentes en el código o los datos. Solamente un enfoque riguroso puede ayudar a prevenirlos o detectarlos, despejar toda duda y cerrar la brecha de confianza.

¿Cómo controlar la IA?

Un marco efectivo puede ayudar a las organizaciones a generar confianza en su tecnología de IA. Dicho enfoque debería ahondar profundamente en la IA a nivel del modelo de la compañía y a nivel individual para asegurarse de que las obligaciones de confianza clave queden integradas y controladas. Se debería evaluar y mantener el control sobre los algoritmos sofisticados y en desarrollo al implementar métodos, controles y herramientas que aseguren los anclajes de confianza en todo el ciclo de vida desde la estrategia hasta el desarrollo. También deberían aportar lineamientos claros para que la organización –partes de interés en diferentes funciones de la gerencia y supervisión- puedan gestionar claramente y a conciencia el ciclo de vida de la IA. Los ejemplos de cómo se puede lograr esto y algunas consideraciones son:

Estrategia

El gobierno de la IA comienza al principio. Una estrategia adecuada para establecer la IA corporativa o para un modelo específico comienza con una clara visión y aspiración y con resultados esperados. Aquí nos encontramos con los conceptos de ética y responsabilidad.

Diseño

Ayuda a que el objetivo propuesto de los algoritmos quede claramente definido y que los modelos sean diseñados para lograr ese objetivo propuesto a través de la ingeniería de características, la determinación del sesgo de los datos y la evidencia empírica. El diseño debe alinearse con los principios (valores y ética), con los estándares y las pautas de seguridad y calidad y con los requerimientos de cumplimiento.

Modelo y capacitación

Una vez que se cumplen los criterios de diseño, se inicia la construcción del modelo y la capacitación. En esta fase, para mantener la integridad, razonabilidad y posibilidad de explicación y resiliencia del modelo, se deberán considerar la detección de los sesgos, los híper-parámetros, la procedencia de las características, entre otras variables. Las características y los datos del modelo deben cumplir con los principios organizacionales, las políticas, los requerimientos del negocio y la normativa vigente.

Evaluar

Éste es un paso clave en el ciclo de vida de la IA, y tiene que ver con la capacidad de verificar que los modelos de IA y los resultados que producen cumplan con los requisitos de integridad, equidad, claridad y resiliencia. Se trata de saber qué preguntas hacer, qué indicadores clave de rendimiento y riesgo buscar, y de tener la capacidad para ejecutarlos. La efectividad e integridad de la capacidad de evaluación y monitoreo de la IA afectará directamente la confianza que una organización (o regulador externo) tenga en la IA empresarial.

Desarrollar & Evaluar

Los modelos de IA y ML no son estáticos y continuarán evolucionando, incluso después de haber pasado a producción, a través de la interacción con nuevos conjuntos de datos u otros modelos. Por lo tanto, los factores clave a considerar en esta fase incluyen el monitoreo en tiempo de ejecución y la presentación de informes sobre controles, cumplimiento, indicadores clave de rendimiento y riesgo, y métricas para determinar la precisión, integridad, equidad, claridad y resiliencia de un modelo. La capacidad de la empresa para reaccionar ante estos indicadores (incluida la calibración dinámica del modelo) también es una capacidad necesaria.

Inteligencia Artificial: un marco para entender los algoritmos

La transparencia de un marco sólido de métodos y herramientas es el combustible para una inteligencia artificial confiable, y crea un entorno que fomenta la innovación y la flexibilidad.

Las organizaciones que construyen e implementan tecnologías de IA están aprovechando un poder de conocimiento y toma de decisiones que supera con creces la capacidad humana. Es una gran oportunidad para las empresas y la sociedad en su conjunto. Pero los algoritmos pueden ser destructivos cuando producen resultados inexactos o sesgados. Es por eso que los líderes dudan en delegar decisiones a las máquinas sin saber por qué se tomaron, o si son justas y precisas.

El poder y el potencial de la IA emergerán por completo, solo cuando sea posible entender los resultados de los algoritmos en un lenguaje claro y sencillo. Las empresas que no prioricen la gestión de la IA y el control de los algoritmos probablemente estén poniendo en peligro su estrategia general de IA, al igual que sus iniciativas y potencialmente su marca.

"Es preciso adoptar prácticas líderes que ayudarán a mitigar el riesgo inherente de la IA en lo que respecta a claridad y sesgo."

Martin Sokalski
Principal, Advisory, Emerging Technology Risk Services
KPMG in the U.S.[11]

[11]Wall Street Journal Pro, "AI, Internet Policy Proposals Signal Shift Away From Self-Regulation," April 2019

KPMG desarrolló el marco de Inteligencia Artificial para ayudar a las organizaciones a ganar mayor confianza y transparencia a través de teorías probadas sobre el gobierno de la IA, al igual que métodos y herramientas utilizadas a lo largo del ciclo de vida de la IA, desde la estrategia hasta la evolución. Por su diseño, este marco aborda los riesgos inherentes ya explicados, e incluye algunas de las recomendaciones clave y prácticas líderes para formar un gobierno corporativo de IA, evaluar la IA, y establecer un monitoreo continuo y visualizaciones.

Gobierno de IA

Desarrollar un criterio de diseño de IA y establecer controles en un entorno que fomente la innovación y la flexibilidad.

Diseñar e implementar un gobierno de IA de punto a punto y un modelo operativo lo largo de todo el ciclo de vida: estrategia, construcción, capacitación, evaluación, desarrollo, operación, y monitoreo de la IA.

Evaluar el marco del actual de gobierno corporativo y realizar un análisis de deficiencias para identificar oportunidades y áreas que deben actualizarse.

Diseñar un marco de gobierno corporativo que ofrezca soluciones de IA e innovación through guidelines, templates, tooling, and accelerators to quickly, yet responsibly, deliver AI solutions.

Integrar un marco de gestión de riesgos para identificar y priorizar algoritmos críticos para el negocio e incorporar una estrategia ágil de mitigación de riesgos para abordar las consideraciones de seguridad cibernética, integridad, equidad y resiliencia durante el diseño y la operación.

Diseñar y establecer criterios para mantener un control continuo sobre los algoritmos sin sofocar la innovación y la flexibilidad.

"Primero, debemos asegurarnos de que los datos estén limpios, sean suficientes y adecuados. Luego, debemos asegurarnos de que el algoritmo genere resultados uniformes y no dependa de pequeños cambios en los supuestos iniciales. Finalmente, debemos asegurarnos de que el objetivo general se haya logrado sin consecuencias demasiado negativas para las partes interesadas."

Cathy O'Neil
Consultor y autor de "Weapons of Math Destruction, from the introduction to Building Trust in a Smart Society" (Sander Klous, KPMG Países Bajos)

Evaluación de la IA

AI realizar una revisión diagnóstico del programa de IA empresarial y de su gobierno corporativo, es posible evaluar el estado actual de los elementos existentes de gobierno corporativo y la posibilidad de aplicarlos a la IA, al igual que a los actuales modelos operativos y su nivel de preparación para llevar la IA a escala. Esto incluirá una evaluación de capacidad y madurez, al igual que una hoja de ruta y recomendaciones para ayudar a alcanzar el estado objetivo.

Evaluación de la conducta individual de los algoritmos de IA y ML: prueba de controles, evaluación del diseño, implementación y operación del algoritmo basado en cuatro pilares de confianza: integridad, claridad, equidad y resiliencia.

Monitoreo continuo y panel de control

- **Crear visibilidad completa** de las métricas relacionadas con los imperativos de confianza, entre ellos los indicadores clave de rendimiento y riesgo, como informes a nivel **Directorio, Ejecutivo y Programa**, centrados en los KPI y KRI de la IA relevante.

- **Habilitar el monitoreo continuo de controles y métricas clave** – qué funciona y qué no en sus modelos de IA/ML.

- **Proporcionar una visión de la tendencia ascendente/descendente** durante un período de tiempo, según los controles y las pruebas.

- **Tener la capacidad de responder y corregir los problemas** a medida que surjan. Por ejemplo, existe un sesgo en el modelo de aprendizaje o se utilizan características prohibidas en la toma de decisiones.

- **Evaluar** su(s) modelo(s) de IA o verificar el estado de su programa de IA empresarial más amplio.

Diferenciadores clave del marco de control de IA de KPMG

Independiente de la plataforma

Monitoreo continuo, identificación de sesgos y precisión

Protección y seguridad continuas, incluidos los datos de capacitación, para prevenir ataques cibernéticos adversarios y de otro tipo.

Capacidad para asignar los términos y conceptos de la ciencia de datos a los indicadores clave de riesgo empresarial

Visibilidad total de lo que están haciendo los modelos de IA

Marco completo que gobierna la construcción, implementación y evolución de los modelos

Ayudará a lograr una **mayor aceptación y escala** en toda la empresa

¿Cómo funciona el Control de la IA?

El conjunto básico de componentes incluye:

Marco integral de inteligencia artificial

El marco de IA ayuda a las organizaciones a generar confianza en el desempeño de su tecnología, al manejar algoritmos de manera transparente y efectiva.

Experiencia en mapeo de conocimientos de IA

Es preciso observar el marco general de gobierno corporativo y gestión de la IA y mapearlo de nuevo a las políticas y directrices corporativas desde una perspectiva de riesgo.

Capacidad de arquitectura de prototipos

Un entorno que fomenta un mayor control de la IA, digitalizado y flexible, para medir el riesgo algorítmico.

Visibilidad y panel de gestión de riesgos

Permite que el usuario tenga visibilidad de las distintas métricas relacionadas con los imperativos de confianza.

Descubra todo el potencial de su IA.

Las organizaciones de hoy dependen en gran medida de aplicaciones basadas en algoritmos para tomar decisiones comerciales críticas. Si bien esto abre oportunidades, también plantea preguntas sobre la confiabilidad. A medida que entramos en una era de gobierno por algoritmos, las organizaciones deben pensar en el manejo de los algoritmos para generar confianza en los resultados y alcanzar todo el potencial de la inteligencia artificial.

Es ahí donde entra en juego KPMG. Las firmas miembro de KPMG creen que gestionar la IA es tan importante como gestionar a las personas. Los profesionales de KPMG operan en un entorno que no depende de la tecnología, y sus recomendaciones se basan en lo que es mejor para sus necesidades. Nuestras firmas miembro trabajan para proporcionar un enfoque holístico y de amplio alcance que lo ayudará a lo largo de su viaje de IA y a lograr sus objetivos comerciales presentes y futuros.

read.kpmg.us/Alincontrol

Contactanos

Martin Sokalski
Principal, Advisory, Emerging Technology Risk Services
KPMG in the U.S.

T: +1 312 665 4937
E: msokalski@kpmg.com

Professor Dr. Sander Klous
Partner, Data & Analytics Lead
KPMG in the Netherlands

T: +31206 567186
E: klous.sander@kpmg.nl

Swami Chandrasekaran
Managing Director
KPMG Innovation & Enterprise Solutions
KPMG in the U.S.

T: +1 214 840 2435
E: swamchan@kpmg.com

read.kpmg.us/Alincontrol

CREATE. | CRT114031A | June 2019