



TECNOLOGÍA, VERDAD Y CONFIANZA EL RUMBO HACIA UNA IA ÉTICA

Este libro blanco explora cómo los creadores y usuarios de soluciones de inteligencia artificial complejas pueden mitigar los numerosos riesgos éticos que representan tanto para las organizaciones como para la sociedad, sin restar valor al profundo potencial de la IA para abordar algunos de los desafíos más urgentes y complejos de nuestra era. A medida que se acelera el ritmo del cambio tecnológico, se vuelve cada vez más crucial encontrar un equilibrio saludable entre las promesas extraordinarias del avance de la IA y la necesidad de controlar los riesgos existenciales que pueden surgir de su uso. En este sentido, la IA encarna perfectamente la metáfora de ser una «espada de doble filo».

De hecho, ambos filos de esa espada se afilaron aún más a raíz del lanzamiento de ChatGPT por parte de OpenAI en noviembre de 2022. Este hito marcó un punto de inflexión en la evolución de la inteligencia artificial, generando un incremento sin precedentes en innovación, inversión y atención pública, al tiempo que intensificó las ya persistentes señales de alarma. Las crecientes preocupaciones se extienden a un amplio abanico de sectores, desde las empresas, los trabajadores y los ciudadanos, hasta el mundo académico, los gobiernos y los organismos reguladores. Dado el ritmo vertiginoso de los avances, estas inquietudes resultan comprensibles. ¿Puede alguien afirmar con certeza que la inteligencia artificial no superará algún día nuestra capacidad de comprenderla o controlarla?

Una vez más, la percepción que las personas tienen de la inteligencia artificial presenta dos caras: mientras que el 44 % de los líderes ejecutivos encuestados por KPMG en 2021 pensaba que su sector estaba avanzando demasiado rápido en la adopción de la IA, casi todos deseaban que su propia organización avanzara aún más rápido¹. Esta contradicción, lejos de disminuir, se ha intensificado con el tiempo.

Basándose en décadas de experiencia colectiva en IA, y en un profundo conocimiento de los riesgos y requisitos de nuestros clientes, Nuix reconoce que el principal factor del dilema de la IA es la confianza. Hasta hace poco, la desconfianza en la IA era un tema propio de películas y novelas apocalípticas. Nuestras interacciones cotidianas con esta tecnología se limitaban a enrutar llamadas telefónicas, corregir errores ortográficos o sugerir películas. Sin

embargo, a medida que empezamos a delegar en la IA tareas cada vez más complejas y significativas, la atención que se presta a la credibilidad y precisión de sus resultados aumenta de forma rápida y, sobre todo, justificada.

Explicabilidad: para generar y mantener la confianza en los resultados de la inteligencia artificial, es fundamental abrir la llamada «caja negra» y ofrecer a las personas una visión clara del origen de los datos de entrenamiento utilizados para construir los modelos, los posibles sesgos presentes y las razones, comprensibles y transparentes, detrás de cada predicción. La inteligencia artificial debe servir como apoyo para informar y acelerar la toma de decisiones humanas, no como una herramienta que pretende reemplazarla entregando supuestas respuestas definitivas. La IA de Nuix orienta a los responsables de la toma de decisiones hacia los lugares donde es más probable que encuentren los hechos verificables que guiarán sus acciones.

Accesibilidad: el verdadero poder transformador de cualquier tecnología radica en su facilidad de uso y capacidad de personalización. La IA de Nuix garantiza que el ser humano mantenga siempre el control total y defina claramente los límites de lo que la máquina puede, o no puede, ver y hacer en su nombre, sin depender de algoritmos o códigos inaccesibles. Una inteligencia artificial que empodera en lugar de reemplazar a las personas es el eje central de este principio.

Especificidad: el concepto de especificidad permite una aplicación pragmática y eficiente de la inteligencia artificial que genera resultados medibles y relevantes en el mundo real. Los modelos generalistas tienden a ser más difíciles de explicar y menos transparentes, lo que también los vuelve menos repetibles. La clave está en permitir que los usuarios lideren el proceso, incorporando su conocimiento y experiencia en la creación de modelos centrados en alguno de los muchos casos de uso específicos disponibles. El resultado es una mejora significativa en la precisión y una mayor credibilidad en los resultados adaptados a cada dominio.

Las implicaciones de estos principios y el modo en que Nuix cumple estos requisitos se abordan más adelante en este documento.

¿POR QUÉ LA REGULACIÓN DE LA IA DEBE SER EXHAUSTIVA Y ANALIZARSE CON PRECISIÓN?

Aunque el costo de desarrollar y entrenar la inteligencia artificial ha disminuido drásticamente en pocos años, los recursos computacionales y humanos necesarios siguen siendo significativos. Por ello, los avances más importantes en IA se han concentrado principalmente en empresas privadas. De hecho, el mundo académico solo genera una décima parte de los avances en IA², lo que indica que gran parte del desarrollo y la aplicación de esta tecnología están probablemente más guiados por intereses económicos que por preocupaciones éticas.

Una prueba clara de que la tecnología evoluciona más rápido que la capacidad de la sociedad para comprender sus beneficios y riesgos es que incluso los principales innovadores y defensores de la inteligencia artificial están pidiendo un enfoque más prudente y equilibrado. Sin embargo, aunque hay grupos de líderes empresariales que proponen públicamente y de manera recurrente³ una moratoria en el desarrollo de la IA, nadie quiere ser el primero en dar el paso atrás.

Los gobiernos y organismos reguladores intentan ponerse al día con el ritmo del cambio y abordar la inteligencia artificial de manera integral, en lugar de hacer ajustes superficiales. La Unión Europea, por ejemplo, está cerca de aprobar una de las primeras leyes completas que regulan la IA. La Ley de IA⁴ alcanzó un hito importante en el Parlamento Europeo apenas unos días antes de la publicación de este documento, y podría entrar en vigor a finales de 2023. Su objetivo es «garantizar un desarrollo ético y centrado en el ser humano» de la inteligencia artificial. Un aspecto clave de esta legislación propuesta es la «capacidad de entender qué modelos se utilizaron para obtener los resultados, cómo se construyeron esos modelos (datos de entrenamiento) y si las personas involucradas pueden ser las responsables finales de los resultados».

De la misma manera que la Unión Europea lideró en materia de privacidad de datos con la introducción del Reglamento General de Protección de Datos (RGPD)⁵, ahora se posiciona a la cabeza en las normativas sobre inteligencia artificial. Tomando el pasado como referencia, la Ley de IA de la UE es un indicador probable de cómo se desarrollará la regulación de la inteligencia artificial a nivel global. De hecho, ya forma parte de la planificación futura de algunas organizaciones. Cuando un cliente de Nuix (una institución de seguridad pública en Europa), solicitó información sobre nuestra inteligencia artificial, Nuix ofreció garantías detalladas por escrito, confirmando que nuestros modelos y capacidades ya cumplen con todos los requisitos establecidos.

En un enfoque más específico para mitigar los posibles riesgos de la inteligencia artificial, un juez federal de Texas implementó recientemente una nueva norma que exige a los abogados que comparecen ante el tribunal presentar un certificado en el que se afirme que ninguno de los documentos presentados fue generado por inteligencia artificial generativa (como ChatGPT, Harvey.AI o Google Bard), o bien que cualquier contenido asistido por IA ha sido revisado cuidadosamente por una persona para verificar su precisión.⁶

En resumen, aprovechar plenamente el inmenso potencial de la inteligencia artificial requerirá tanto una reflexión seria como acciones cuidadosas y coordinadas. Como señalaron dos académicos australianos especializados en IA, «está claro que nos

«Lo más triste de la vida ahora mismo es que la ciencia acumula conocimientos más rápido de lo que la sociedad adquiere sabiduría».

Isaac Asimov

encontramos en un punto de inflexión: debemos pensar con seriedad y urgencia sobre los riesgos y efectos negativos que está revelando el uso creciente de la inteligencia artificial⁷.

LA VISIÓN DE NUIX

En Nuix llevamos años trabajando con inteligencia artificial y desarrollando sus capacidades. Sabemos que la IA tiene el poder de transformar profundamente a nuestros clientes, e incluso de redefinir economías enteras. Sin embargo, también somos conscientes de que la humanidad podría tener una sola oportunidad para gestionar correctamente la transición hacia un futuro impulsado por la inteligencia artificial. La regulación inteligente, la educación, la disciplina y el debate honesto son de vital importancia para garantizar que la IA se construya y se aplique de forma reflexiva, ética y transparente. En definitiva, una inteligencia artificial ética es sinónimo de una inteligencia artificial fiable.

En palabras de Stephen Stewart, Director Tecnológico de Nuix: «La realidad es que aprovechar la inteligencia artificial y el aprendizaje automático en medio de la increíble diversidad de desafíos relacionados con los datos no es una tarea sencilla. No existe una solución única que funcione para todos los casos de datos no estructurados. La IA de Nuix representa un ejemplo concreto de cómo aplicar los conceptos más avanzados y las innovaciones tecnológicas en inteligencia artificial para dar respuesta a necesidades muy específicas de nuestros clientes, con una eficiencia notable y con información respaldada y defendible».

Nuestro enfoque para desarrollar y aplicar una inteligencia artificial fiable se basa en tres principios fundamentales de diseño tecnológico: explicabilidad, accesibilidad y especificidad. Estos principios nos mantienen firmemente alineados con nuestro propósito de ser una fuerza para el bien, al descubrir la verdad contenida en los datos digitales. Además, nos garantizan la creación de una inteligencia colaborativa que potencie la eficiencia y la efectividad humanas, permitiendo que las personas se concentren en el trabajo de alto valor, aquel que requiere juicio humano, pensamiento crítico matizado y conciencia ética.

Estos principios positionan a Nuix como un proveedor de inteligencia artificial responsable, reconocido y ético. Además, también actúan como una guía reflexiva que orienta nuestros esfuerzos continuos de desarrollo, y garantizan que sigamos alineados con las necesidades de nuestros clientes y con los estándares éticos globales en constante evolución para la IA.

1. EXPLICABILIDAD

Como líderes globales en software de inteligencia digital, nuestros clientes confían en nosotros para procesar, filtrar e interpretar enormes volúmenes de datos con el fin de encontrar información basada en hechos concretos, es decir, la verdad. Para establecer y mantener esa confianza en la inteligencia artificial, nuestros clientes necesitan tener visibilidad sobre los datos utilizados para construir nuestros modelos de IA, así como comprender las explicaciones que sustentan los resultados que dichos modelos generan. Este principio, cada vez más reconocido en la industria, se conoce como «explicabilidad», y nuestra tecnología de inteligencia artificial fue diseñada desde sus cimientos con ese propósito.

Otro elemento fundamental de la explicabilidad es que nuestra IA no pretende tomar decisiones en lugar de las personas. Por el contrario, proporciona datos tamizados, pertinentes y priorizados que respaldan una toma de decisiones humana más eficaz y eficiente⁸ por parte de nuestros clientes. Lo esencial aquí es que el ser humano tiene acceso directo a cada etapa del proceso y conserva siempre el control final sobre las decisiones. Nuestra inteligencia artificial ofrece cuatro capas analíticas aplicadas a datos textuales:

1. Tipo de documento: por ejemplo, formulario fiscal, documento de salud, expediente académico, patente.
2. Tema del contenido: por ejemplo, derecho, medioambiente, finanzas, política.

3. Extracciones: por ejemplo, información personal identificable (PII), información de salud protegida (PHI), entidades con nombre.
4. Prioridad: por ejemplo, nivel de relevancia, importancia o riesgo de cualquiera de los elementos anteriores, según reglas definidas por el cliente.

La solución de Nuix orienta a los responsables de la toma de decisiones hacia los puntos donde es más probable encontrar los hechos incuestionables que guiarán sus acciones y mejorarán tanto la precisión como la eficiencia en sus procesos de decisión. Cada paso del proceso queda registrado y puede reproducirse con consistencia.

2. ACCESIBILIDAD

Hasta hace poco, la inteligencia artificial estaba reservada a un grupo reducido de expertos altamente especializados en aprendizaje automático, programación de software y ciencia de datos. Estas condiciones dejaban fuera del proceso de creación de modelos a los especialistas en los distintos dominios del conocimiento, y los obligaban a depender en exceso de equipos técnicos que, en muchos casos, no contaban con el conocimiento profundo necesario en esas áreas específicas.

Para maximizar el valor de la inversión en inteligencia artificial, y al mismo tiempo cumplir con el imperativo ético de ampliar el acceso a las oportunidades y beneficios que esta tecnología ofrece, es fundamental brindar a las personas sin conocimientos técnicos la posibilidad de acceder a estas capacidades de manera intuitiva y fácil de usar.

Nuix está impulsando una verdadera revolución en la experiencia del usuario, con una interfaz intuitiva diseñada para que los expertos en cada materia puedan crear, optimizar y validar modelos de inteligencia artificial lingüística altamente precisos de forma rápida, mediante una sencilla interfaz de apuntar y hacer clic, sin necesidad de escribir una sola línea de código. Más allá de los avances en flujo de trabajo y productividad que esto representa, este nivel de accesibilidad refuerza la confianza en que los resultados de la inteligencia artificial han sido verificados y están bien alineados con el conocimiento especializado para el que fueron diseñados.

Unas interfaces bien diseñadas, con poco o ningún código, y unos flujos de trabajo integrados, contribuyen a democratizar el poder de la inteligencia artificial, permitiendo que una mayor diversidad de personas pueda interactuar directamente con ella para realizar los ajustes necesarios, probar y optimizar modelos, corregir errores y eliminar sesgos. En definitiva, permite que el control permanezca siempre en manos del usuario de la inteligencia artificial, y no en quien la programó.

3. ESPECIFICIDAD

En el contexto de aplicaciones reales, más grande no siempre significa mejor. Aunque ChatGPT está respaldado por GPT4, el modelo de lenguaje más grande jamás creado⁹, su inmensidad y su intento de «ser todo para todos» limita mucho su utilidad práctica en entornos empresariales reales. Por lo tanto, está claro que sería poco práctico, e incluso imposible, alojar ChatGPT en un entorno seguro propio. Además, subir datos sensibles para usarlo tal cual en la web representa un riesgo significativo de exposición. Por esta razón, muchas empresas y organizaciones han prohibido a sus empleados el acceso a ChatGPT y a otros grandes proveedores de inteligencia artificial generativa. Estos enormes modelos de lenguaje no están bien adaptados para dar soporte a temas específicos y profundos relevantes para su negocio o área de especialización.

«Un ordenador solo merecería ser llamado inteligente si pudiera engañar a un humano haciéndole creer que es humano».

Alan Turing

En cambio, la IA de Nuix está diseñada para implantarse en los entornos seguros de los clientes y ejecutarse en sistemas asequibles basados en CPU. Además, combinando las ventajas de la explicabilidad y accesibilidad mencionadas anteriormente, los clientes pueden ajustar rápidamente modelos predefinidos o incluso crear nuevos modelos que se adapten a sus requerimientos particulares y casos de uso específicos. Una vez más, los humanos mantienen el control y la inteligencia artificial trabaja para respaldar sus necesidades concretas.

IA QUE CUMPLE LAS EXPECTATIVAS



Los tres principios de diseño tecnológico que hemos mencionado fueron clave para nuestra participación galardonada en el desafío de automatización de herramientas CUI de la Marina de Estados Unidos, anunciado en diciembre de 2022¹⁰.

El objetivo de este desafío fue facilitar el intercambio adecuado, efectivo y fundamental de información sensible pero no clasificada, conocida como «información no clasificada controlada» (CUI), entre departamentos, socios y aliados. Las tareas específicas que se debían gestionar incluyeron:

- > Procesar millones de documentos del Gobierno de Estados Unidos y del Departamento de Defensa en diversos formatos.
- > Analizar el contenido de los documentos para identificar los tipos de CUI presentes en cada archivo.
- > Cuantificar un nivel de confianza que respalte cada resultado.
- > Marcar el encabezado y pie de página de cada documento para mantener la cadena de confianza a medida que distintos usuarios interactúan con los archivos.

El éxito de nuestra solución se construyó sobre los cimientos que proporcionaban nuestros tres principios. La **explicabilidad**, que se logra a través de una interfaz sencilla de apuntar y hacer clic, permite entender cómo se crearon los modelos de inteligencia artificial nuevos y existentes, así como las razones claras detrás de nuestros resultados. La **accesibilidad**, proporcionada por una interfaz sin necesidad de programación, empodera al usuario para construir, entrenar, mejorar o ajustar los modelos y sus ponderaciones relativas con el fin de optimizar y validar el rendimiento de cada modelo. Y la **especificidad** de cada modelo, cuidadosamente ajustada para un tipo específico de CUI, y combinada por expertos en la materia para abordar desafíos únicos y reales en la gestión de datos.

LA IA Y SU ORGANIZACIÓN

La capacidad de utilizar tecnologías avanzadas de forma rápida y efectiva ha sido reconocida desde hace mucho tiempo por los economistas como un factor más crucial para el crecimiento que la propia naturaleza de la tecnología. La famosa frase de Robert Solow lo resume perfectamente: «La era informática se nota en todas partes menos en las estadísticas de productividad».

En Nuix, nos hemos forjado una reputación global al hacer que los datos de nuestros clientes sean altamente accesibles y buscables. Ahora, damos un paso adelante con una plataforma unificada e impulsada por inteligencia artificial llamada Nuix Neo. Nuix Neo combina el motor de procesamiento de datos más poderoso del mundo con inteligencia artificial avanzada para ayudar a resolver los complejos y crecientes desafíos de datos que enfrentan nuestros clientes. Gracias a esta estrategia integrada de innovación, los clientes de Neo pueden aprovechar el valor de la inteligencia artificial sin necesidad de contar con experiencia interna en IA ni de adquirir o desarrollar múltiples soluciones puntuales de inteligencia artificial.



La implementación de soluciones de inteligencia artificial no es una tarea que deba tomarse a la ligera. En un mundo cada vez más invadido por chatbots persuasivos que a veces generan información errónea, y saturado de desinformación, noticias falsas y *deepfakes*, los riesgos son mayores que nunca. Sin embargo, la realidad es que los problemas de hoy y del mañana no se resolverán con las técnicas y tecnologías de ayer. Nos guste o no, creemos que dominar la inteligencia artificial es parte fundamental del camino hacia un futuro próspero y sostenible.

La clave estará en adoptar un enfoque consciente y disciplinado, basado en los principios de explicabilidad, accesibilidad y especificidad. El objetivo final será ofrecer resultados fiables, que sean creíbles, verificables y defendibles para nuestros clientes y para la sociedad en general.

Como pone en relieve el caso práctico anterior, minimizar el riesgo de una brecha de datos es fundamental. Sin embargo, si se produce una brecha, las organizaciones deben saber cómo responder. Para muchas, esto implica contar con la capacidad de desplegar tecnologías avanzadas y potencia de procesamiento para analizar y comprender la brecha, así como para satisfacer las necesidades de información de clientes, reguladores y medios de comunicación.

Un banco estadounidense que recientemente sufrió una brecha de datos pudo aprovechar el poder de Nuix para impulsar sus esfuerzos de investigación y remediación. El banco utilizó múltiples servidores para buscar en varios terabytes diarios, lo que le permitió evaluar y reportar rápidamente el alcance de la brecha.

¿QUIERE SABER MÁS?



Las tendencias abordadas en este documento continúan en evolución, al igual que las aplicaciones reales que nuestros clientes tienen para las capacidades de inteligencia artificial de Nuix. Si quiere más información sobre cómo estas capacidades pueden aplicarse en su organización, no dude en ponerse en contacto con nosotros a través de los canales que se indican a continuación. Uno de nuestros expertos en inteligencia artificial se pondrá en contacto con usted para ayudarle a comprender cómo Nuix Neo y su inteligencia artificial integrada pueden ayudar a su organización.

REFERENCIAS

- ¹ Krishna, Campana y Chandrasekaran, "Thriving in an AI World: Unlocking the value of AI across seven key industries," KPMG, 2021, <https://advisory.kpmg.us/articles/2021/thriving-in-an-ai-world.html>
- ² Universidad de Stanford, "Chapter 1.2., Research and Development", Fig 1.2.2, Artificial Intelligence Index Report 2023", 2023, <https://aiindex.stanford.edu/report/#individual-chapters>
- ³ Véase <https://www.science.org/content/article/alarmed-tech-leaders-call-ai-research-pause#:~:text=Attracting%20signatures%20from%20the%20likes,potential%20risks%20of%20the%20technology>
- ⁴ Véase <https://www.europarl.europa.eu/news/en/press-room/20230505IPR84904/ai-act-a-step-closer-to-the-first-rules-on-artificial-intelligence>
- ⁵ Véase <https://gdpr.eu/what-is-gdpr/>
- ⁶ Véase <https://www.mitchellwilliamslaw.com/taming-the-ai-beast-judges-set-rules-to-control-use-of-generative-ai-in-their-courts>
- ⁷ Sonenberg y Walsh, "Artificial intelligence is now part of our everyday lives and its growing power is a double edged sword", The Conversation, noviembre de 2021, <https://theconversation.com/artificial-intelligence-is-now-part-of-our-everyday-lives-and-its-growing-power-is-a-double-edged-sword-169449>.
- ⁸ Mientras redactamos estas líneas, Nuix se encuentra en conversaciones con una de las mayores agencias de seguridad de Europa, que busca aprovechar los beneficios de la inteligencia artificial con un enfoque central en la transparencia de los resultados.
- ⁹ Véase <https://neuroflash.com/blog/gpt-4-parameters-rumors-and-forecasts/#:~:text=It%20is%20also%20possible%20that,trillion%20or%2010%20trillion%20parameters>.
- ¹⁰ 'Nuix and Serco NA partner to score prizewinning results in US Navy AI Automation Challenge', Nuix, <https://www.nuix.com/news/nuix-and-serco-na-partner-score-prizewinning-results-us-navy-ai-automation-challenge>

ENCONTRAR LA VERDAD EN UN MUNDO DIGITAL

Obtenga más información sobre Nuix
o póngase en contacto con nosotros
para una demostración gratuita en
www.nuix.com/contact-us



Nuix Limited (ASX: NXL) es un proveedor líder de software de análisis de investigación e inteligencia, que ayuda a organizaciones a proteger, gestionar y maximizar el valor de sus datos mediante el uso de inteligencia artificial. Con una presencia global y más de 20 años de experiencia, Nuix ofrece herramientas avanzadas de análisis de datos diseñadas para sectores que exigen precisión y profundidad, como el eDiscovery, la gestión y revisión legal, el cumplimiento normativo, la gobernanza de datos, la ciberseguridad y las investigaciones forenses. Gracias al uso de procesamiento avanzado, inteligencia artificial y aprendizaje automático, Nuix le ayuda a procesar, normalizar, indexar, enriquecer y analizar grandes volúmenes de datos complejos con rapidez, escalabilidad y una precisión forense incomparable.

Para más información, visite www.nuix.com

APAC

Australia: +61 1300 511 852

EMEA

Reino Unido: +44 203 934 1600

NORTEAMÉRICA

EE. UU.: +1 877 470 6849

Nuix (y cualquier otra marca registrada de Nuix utilizada) son marcas comerciales de Nuix Ltd. y/o de sus filiales, según corresponda. Todos los demás nombres de marcas y productos son marcas registradas de sus respectivos propietarios. Cualquier uso de las marcas comerciales de Nuix requiere la autorización previa por escrito del Departamento Jurídico de Nuix. Puede ponerse en contacto con el Departamento Jurídico de Nuix enviando un correo electrónico a legal@nuix.com. Este material contiene propiedad intelectual de Nuix Ltd. y sus filiales ('Nuix'), incluyendo contenido protegido por derechos de autor, debidamente señalado como tal y/o registrado ante la oficina de derechos de autor de los Estados Unidos. Cualquier reproducción, distribución, transmisión, adaptación, exhibición pública o ejecución pública de esta propiedad intelectual (salvo para fines internos previamente aprobados) requiere la autorización previa por escrito de Nuix.