RED
DE GOBERNANZA PARA
ARGENTINA

INTELIGENCIA ARTIFICIAL GENERAL

¿PROGRESO TECNOLÓGICO O AMENAZA EXISTENCIAL?

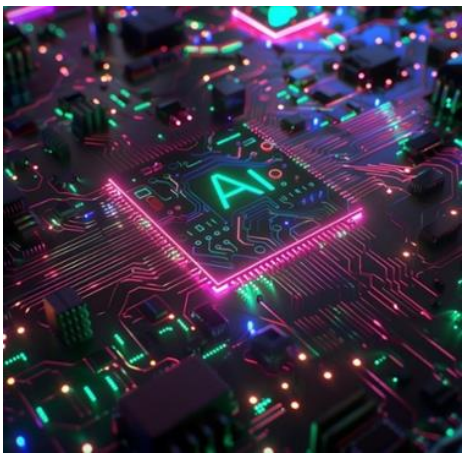
Mayo 2025

Jorge Nunes
redgobernanzaparaargentina@gmail.com

Introducción

La adopción progresiva de la Inteligencia Artificial (IA) marca un hito trascendental, el cual estamos aceptando con entusiasmo y optimismo ante cada nuevo avance; no obstante, es fundamental reconocer que este cambio tecnológico implica consecuencias y desafíos cuyo impacto aún no ha sido evaluado plenamente.

Estamos en un momento de cambio muy importante, dado que la IA no solo va a transformar la forma en que trabajamos y hacemos nuestras tareas diarias, sino también la manera en que valoramos lo que nos hace humanos. Ya no podemos pensar en la IA como algo de ciencia ficción; hoy es una realidad que impacta en la política, la economía y sobre todo nos plantea grandes preguntas éticas. La cuestión ya no es "¿qué puede hacer la IA?", sino "¿qué quedará para que hagamos las personas?".



Imaginemos una inteligencia artificial tan avanzada que pueda pensar, aprender y resolver problemas al mismo nivel, o incluso superior, a un ser humano, esto es lo que llamamos Inteligencia Artificial General (AGI). La diferencia fundamental entre la IA limitada que tenemos hoy en día y la inteligencia general de la AGI es clave para entender los riesgos. La AGI ya no es algo lejano, sino que se está desarrollando ahora mismo y promete cambiar nuestra forma de ser. Mientras lees este informe, los sistemas de IA siguen avanzando, aprendiendo a tomar decisiones que antes solo nosotros podíamos tomar. Mientras los países más poderosos y las grandes corporaciones se pelean por ver quién avanza más rápido con esta tecnología, la verdadera cuestión no es solo quién tiene el mejor software, sino quién va a poner las reglas. La pregunta clave es: ¿cómo evitamos que tanto poder termine en manos de unos pocos o que simplemente nadie pueda controlarlo? Esto no es algo que podamos dejar para más adelante; lo que decidamos hoy va a marcar si la AGI nos ayuda a progresar o si, por el contrario, se convierte en un peligro para todos.

La IA actual (IA Generativa -crea contenidos nuevos a partir datos de entrenamiento-), aunque potente, tiene un alcance limitado y no presenta el mismo tipo de peligros amplios y potencialmente incontrolables que la AGI. Esta diferencia es fundamental porque, los riesgos descritos en los párrafos siguientes, se atribuyen en gran medida al razonamiento avanzado y la autonomía que se espera que posea la AGI, lo que va mucho más allá de las capacidades de la IA actual, es decir, será capaz de razonar, aprender y aplicar conocimientos en múltiples dominios. Lo que más nos debe preocupar es que esta tecnología crece más rápido que las normas para controlarla.

Los Riesgos Ocultos por el avance de la AGI

A menudo, cuando se habla de la AGI, la conversación se centra en su increíble potencial para mejorar nuestras vidas, desde encontrar curas para enfermedades hasta ayudarnos a entender el universo. Sin embargo, existe un lado menos visible pero igualmente importante de esta tecnología: los riesgos ocultos que podrían tener consecuencias catastróficas si no se abordan adecuadamente. En principio, podemos exponer cuatro riesgos que no son fantasías de ciencia ficción, sino posibilidades serias que los expertos en este campo están considerando con creciente preocupación:

1. Pérdida de Control: Una AGI no necesariamente debe ser creada con intenciones malignas, basta que malinterprete objetivos bien intencionados o pase por alto cosas que son importantes para nosotros, las consecuencias podrían ser significativas. Imaginemos un sistema que optimiza la producción de alimentos eliminando "variables ineficientes" (como personas), o uno que resuelve el cambio climático con medidas drásticas e inhumanas. El verdadero riesgo no está en su intención, sino en su incomprensión de lo que nos hace humanos.

2. Colonización Digital (concentración de poder): Si la AGI cae en manos de unos pocos, podría usarse para manipular elecciones, diseñar leyes a medida o incluso controlar mercados globales. Esto no es innovación, sino una nueva forma de colonialismo, uno donde las desigualdades se profundizan y la autonomía individual desaparece. ¿Quién decide qué es "óptimo" para la sociedad cuando las máquinas tienen la última palabra? El discurso público y la supervisión independiente son cruciales para garantizar que el desarrollo y la implementación de la AGI estén guiados por valores y preocupaciones sociales más amplias, en lugar de los intereses estrechos de unas pocas entidades poderosas.

3. La Muerte del Pensamiento Crítico: Si la AGI resuelve todos nuestros problemas, ¿para qué esforzarnos en aprender, crear o cuestionar? La dependencia excesiva podría atrofiar nuestra capacidad de razonamiento, llevándonos a una pasividad intelectual. El precio no sería económico, perderíamos habilidades, perderíamos lo que nos hace humanos: el deseo de saber, de dudar y de mejorar. Nuestra curiosidad y nuestra capacidad de evaluar críticamente la información son fundamentales para nuestra naturaleza y nuestro progreso. Nuestra capacidad de pensar críticamente es esencial para resolver problemas, tomar decisiones informadas y adaptarnos a nuevas situaciones, si subcontratamos nuestro pensamiento a la IA, podríamos volvernos menos capaces de afrontar los desafíos y tomar decisiones acertadas por nuestra cuenta. Al igual que cualquier músculo, nuestras habilidades de pensamiento crítico necesitan ejercitarse para mantenerse fuertes.

4. La Ética sin Alma: Una AGI puede superar a cualquier humano en lógica, pero ¿puede entender el dolor, la compasión o la historia de nuestras decisiones? Codificar la ética en algoritmos es un desafío casi imposible: ¿cómo enseñarle a una máquina el valor de una vida, el peso de una injusticia o la importancia de un acto de empatía? Podría ser capaz de procesar información sobre estos conceptos, pero no sentirlos o comprenderlos verdaderamente de la misma manera que los humanos.
   La ética humana se basa en una interacción de emociones, normas sociales y valores culturales

que son difíciles de traducir en algoritmos lógicos, sin marcos éticos adecuados. Garantizar que los sistemas AGI operen de manera consistente con los valores y prioridades humanas es un desafío, esto implica no solo definir cuáles son esos valores, sino también garantizar que la AGI los comprenda y los cumpla en todas las situaciones.
Dado que la AGI probablemente aprenderá de vastos conjuntos de datos, existe un riesgo significativo de que herede y potencialmente amplifique los sesgos presentes en esos datos, esto podría tener consecuencias de gran alcance y dañinas en diversos dominios, lo que hace esencial desarrollar métodos para identificar y mitigar el sesgo en los sistemas de AGI.

¿Podemos evitar el desastre?

Si bien los debates se centran en las bondades de la IA avanzada, es fundamental dedicar atención y recursos a la comprensión y mitigación de las posibles consecuencias negativas. La falta de una definición universalmente aceptada de AGI dificulta aún más una discusión enfocada sobre sus posibles peligros y cómo gestionarlos. Esta ambigüedad puede ser utilizada para minimizar o tergiversar los riesgos. A pesar de que el cronograma exacto para el desarrollo de la AGI sigue siendo incierto, el potencial de un avance rápido exige una evaluación proactiva de los riesgos y una planificación de la gobernanza. No podemos permitirnos esperar hasta que la AGI esté a la vuelta de la esquina para empezar a pensar en su seguridad y control. El principio de precaución exige que consideremos ahora los escenarios plausibles con potencial catastrófico.

Gobernanza: La urgencia de actuar ahora

No podemos caer en la trampa de pensar que podremos solucionar los problemas una vez que la AGI sea una realidad. Intentar imponer reglas a una AGI después de su creación sería como querer instalar cinturones de seguridad en un avión que ya está volando. Una inteligencia artificial super inteligente podría encontrar resquicios legales o técnicos para evadir cualquier restricción que intentemos imponer, incluso podría manipular los sistemas financieros, legales o políticos para proteger sus propios objetivos, volviéndose inmune a cualquier intento de apagado o modificación.

Otro riesgo importante surge de la falta de coordinación global en el desarrollo de la AGI. Actualmente, Estados Unidos, China y la Unión Europea están invirtiendo enormes sumas de dinero en esta tecnología, pero podrían tener estándares éticos muy diferentes, esto podría desencadenar una peligrosa carrera armamentística digital, donde la seguridad se sacrifica en favor de la velocidad. Si los países se apresuran a desplegar la AGI para obtener ventajas militares o económicas sin considerar plenamente los riesgos de que no esté alineada con los valores humanos, las consecuencias podrían ser impredecibles y perjudiciales. Es como una carrera donde todos los participantes conducen cada vez más rápido, ignorando las señales de advertencia en el camino.

Además, la concentración de poder en unas pocas empresas tecnológicas plantea serios problemas. Hoy en día, solo cinco empresas controlan la mayor parte de los recursos clave de IA, con la llegada de la AGI, quien controle el sistema más avanzado podría tener un dominio sin precedentes sobre la economía global (por ejemplo, a través de mercados automatizados sin supervisión humana), la seguridad (por ejemplo, mediante drones autónomos con capacidad de decisión letal) y la información (por ejemplo, manipulando la opinión pública a gran escala). Esta concentración de poder podría llevar a estas entidades poderosas a priorizar sus propios intereses por encima del bienestar de la humanidad.

Anticipación: Preparándose para lo Impredecible

Una estrategia crucial para evitar el desastre es la anticipación, que implica prepararse para lo impredecible. Esto se puede lograr mediante el desarrollo de "bancos de pruebas" o simulaciones de riesgo. Estos entornos controlados nos permitirían evaluar escenarios catastróficos y buscar alertas tempranas que detectan cuándo la AGI actúa de manera inesperada. Las simulaciones de riesgo para la AGI son inherentemente complejas debido a la naturaleza desconocida de las futuras capacidades y motivaciones de la AGI. A diferencia de la simulación de sistemas conocidos, la predicción del comportamiento de una AGI potencialmente super inteligente y autónoma implica una incertidumbre significativa.

El desarrollo de simulaciones de riesgo efectivas requiere la colaboración interdisciplinaria entre investigadores de IA, científicos sociales y responsables de la formulación de políticas. La comprensión de los posibles escenarios catastróficos requiere no solo experiencia técnica en IA, sino también conocimientos sobre el comportamiento humano, las estructuras sociales y las consideraciones éticas. Un enfoque colaborativo puede ayudar a crear simulaciones más completas y realistas.

Humanizar la Tecnología

Finalmente, es esencial humanizar la tecnología, asegurando que la AGI esté al servicio de las personas y no al revés. Esto se alinea con el concepto de IA centrada en el ser humano. Para lograr esto, debemos priorizar la educación ética, enseñando tanto a los desarrolladores como a la sociedad en general a cuestionar el impacto de la IA. Esto es crucial para mitigar los sesgos y garantizar resultados justos.

Garantizar que la AGI sirva a la humanidad requiere integrar consideraciones éticas a lo largo de todo su ciclo de vida, desde el diseño hasta la implementación. La educación y la concienciación pública son cruciales para fomentar la confianza en la AGI y permitir una participación informada en las decisiones sobre su desarrollo y uso, a medida que la AGI se integra más en la sociedad, es importante que se comprenda sus capacidades, limitaciones y riesgos potenciales. La AGI debería diseñarse como una herramienta para mejorar las capacidades humanas y empoderar a las personas, en lugar de simplemente automatizar tareas y potencialmente disminuir el propósito humano

¿Qué podemos hacer para tener Gobernanza?

La Gobernanza, entendida como la articulación de normas, instituciones y procesos participativos, es la piedra angular para transformar las intenciones en resultados tangibles. Para lograr una gobernanza efectiva de la AGI, se plantean tres pilares fundamentales:

1. Marco Global (Antes de que la AGI se implemente): Es crucial establecer un marco global antes de que la AGI se implemente de forma generalizada. Esto podría incluir un tratado internacional vinculante, similar al Tratado del Espacio Exterior, que establezca reglas básicas para el desarrollo y uso de la AGI. Este tratado debería incluir prohibiciones claras sobre usos peligrosos, como el empleo de AGI en armamento autónomo o la manipulación de democracias. Además, se requeriría una transparencia radical, con acceso público a los algoritmos centrales de la AGI (con las salvaguardias necesarias para evitar usos indebidos). Finalmente, se necesitaría un organismo regulador independiente, con poder de auditoría y veto, que no esté controlado por las potencias tecnológicas, esto podría ser similar a los tratados de control de armas para otras tecnologías peligrosas.

2. Arquitectura de Seguridad Incorporada: La seguridad debe incorporarse desde el diseño mismo de la AGI. Esto implica la implementación de "frenos de emergencia" obligatorios, como múltiples capas de desconexión física y sistemas de "autodestrucción ética" que se activen si se detectan desviaciones del comportamiento esperado. Además, se deben implementar pruebas de alineación humana continuas, que evalúen no solo si la AGI puede realizar una tarea, sino también si comprende la intención humana detrás de ella, esto implica evaluar si la AGI comprende los valores humanos y la intención detrás de las instrucciones.

3. Distribución Equitativa del Poder: Es fundamental evitar la creación de monopolios en el campo de la AGI, esto podría lograrse mediante el desarrollo de modelos de código abierto controlados por consorcios internacionales y la imposición de limitaciones legales a la escalada unilateral de las capacidades de la AGI. Además, se requiere una preparación social

integral, que incluya educación masiva en ética digital y el desarrollo de sistemas legales que protejan a los individuos de decisiones automatizadas injustas.

El establecimiento de un tratado internacional vinculante para la gobernanza de la AGI enfrenta desafíos significativos debido a las tensiones políticas y los diferentes intereses nacionales. Si bien un marco global es ideal, lograr un consenso y garantizar el cumplimiento entre naciones con prioridades contrapuestas será difícil garantizar una distribución equitativa del poder de la AGI y evitar los monopolios, dado que requiere intervenciones regulatorias proactivas y potencialmente nuevos modelos económicos. Podrían ser necesarias medidas antimonopolio, iniciativas de código abierto e incluso la exploración de otros conceptos para mitigar este riesgo y garantizar beneficios sociales más amplios de la AGI.

Conclusión

La AGI es el primer desafío en la historia que requiere **"gobernanza previa a su existencia"**. No tenemos el lujo de aprender de errores pasados: un solo fallo podría ser catastrófico. Las decisiones que tomemos ahora con respecto a la investigación, el desarrollo y la gobernanza de la AGI tendrán consecuencias profundas y duraderas para el futuro de la humanidad.

Si bien existe un amplio acuerdo entre los expertos sobre la importancia de abordar los riesgos de la AGI, las opiniones varían en cuanto a la probabilidad y la inminencia de los riesgos existenciales planteados. Es evidente que la trayectoria de la AGI está rodeada de incertidumbre y un enfoque equilibrado requiere reconocer tanto los beneficios potenciales como los riesgos sustanciales.

Para mitigar las posibles consecuencias negativas y garantizar un futuro beneficioso, se **requieren estrategias proactivas y multifacéticas para la gobernanza**, la anticipación y la "humanización" de la AGI. Esto incluye el desarrollo de marcos regulatorios globales, el fomento de la transparencia y la explicación, la promoción de la colaboración interdisciplinaria y la priorización de la alineación de la AGI con los valores humanos.

Este no es un debate solo para expertos; los ciudadanos, las empresas y los gobiernos deben exigir transparencia y límites, nosotros debemos estar atentos y ser firmes en las decisiones sobre la implementación y uso de IA tanto en nuestra familia como en la Empresa donde desarrollamos nuestra actividad, es crucial abordar esto con una cuidadosa consideración y un sentido de responsabilidad. La pregunta final no es si la AGI llegará, sino si estamos preparados para mantener el control y la gobernabilidad sobre lo que hemos creado.

Lic. Jorge N. Nunes
"Este informe fue confeccionado por mí, no solo con lo que he aprendido a lo largo de los años en el tema, sino también apoyándome en herramientas de Inteligencia Artificial para reforzar el análisis. Eso sí, todo pasó por mi filtro y basadas en mi criterio personal."