

Proyecto de investigación Consciencia y Sociedad Distópica

Comunidad en Telegram. 29 de mayo de 2025

Enlace de suscripción al canal en Telegram: <https://t.me/socdistopica>

ÉTICA, CONSCIENCIA Y ESPIRITUALIDAD ANTE EL RIESGO EXISTENCIAL DE LA INTELIGENCIA ARTIFICIAL

Lo que las máquinas superinteligentes puedan hacer con nuestra civilización y la Tierra dependerá de sus objetivos, originalmente definidos por un marco ético de referencia que, quién sabe, podría ser reformulado por la propia superinteligencia artificial.

Hoy todavía estamos en la fase de las inteligencias artificiales específicas o débiles (IAE) que, aunque avancen en su desarrollo de manera aceleradamente creciente hacia la IA General (IAG), podría ser que poco tiempo después alcancen lo que algunos denominan "el último gran invento" de la humanidad: la IA de nivel sobrehumano fuerte (IAF) o superinteligencia artificial (SIA).

Manolo López, autor de "La décima inteligencia", 8º libro de la Biblioteca de este Proyecto, explica en este texto cómo "sembrar Semillas de Consciencia o Hackear Conscientemente la IA" para que siempre esté alineada con la Reverencia por la Vida y al Servicio de toda la Humanidad.

CARRERA IMPARABLE HACIA LA IAG

La búsqueda desenfadada de la IAG en los próximos años, por parte de las mejores empresas tecnológicas del planeta, ubicadas principalmente en EEUU y China, supone un importante riesgo existencial si no se invierte más en seguridad y alineación ética de las IAs, debido a que el posible despegue de la IAG hacia la SIA podría ocurrir en un breve periodo de tiempo, según algunos expertos quizás en menos de un año desde que se alcance la IA de nivel humano, debido a la posibilidad de la automejora recursiva sin apenas intervención por nuestra parte, si lo que se pretendiera fuese alcanzar el nivel de superinteligencia artificial lo antes posible.

El principal problema está en las denominadas "cajas negras" de la IA. El término "caja negra" se populariza en los años 80 y 90, cuando los sistemas de aprendizaje automático comienzan a ser más complejos y menos interpretables. En la década de 2010, con el auge del aprendizaje profundo, la opacidad de los modelos de IA se convierte en un problema crítico al no poderse predecir ni comprender las frecuentes "alucinaciones" y respuestas desalineadas de las IAs. Y en los últimos años, sobre todo desde que *Open AI* lanzó con gran éxito en 2022 su modelo de lenguaje ChatGPT, la IA se está haciendo más compleja y

menos comprensible y, a la vez, más poderosa y más peligrosa, dependiendo del uso que le queramos dar. La metáfora de la "caja negra" no sólo sigue vigente, sino que se agrava con modelos entrenados en billones de parámetros, dentro de las crecientes infraestructuras de redes neuronales artificiales. Esta opacidad interna en la toma de decisiones de la IA limita la auditoría, dificulta la atribución de responsabilidad y amenaza con un posible desalineamiento ético futuro cuando las IAs alcancen mayores cotas de inteligencia y autonomía.

La falta de transparencia, sumada a la carrera geopolítica por la supremacía en el logro de la IAG, agrava los riesgos: una IA poderosa, pero desalineada con los valores humanos más elevados, podría intensificar las desigualdades, la desinformación, el control social y el desequilibrio ecológico.

NUESTRO MAYOR RETO EXISTENCIAL

Por tanto, estamos ante el mayor riesgo existencial al que la humanidad y la propia vida sobre este planeta se ha podido enfrentar jamás. Tener claro cuál es el camino a seguir, con suficiente antelación, antes de la hipotética explosión de inteligencia, debería ser nuestro principal reto ante el, posiblemente, último gran invento humano. Las decisiones futuras de la SIA pudieran volverse inapelables en su autoridad si no empezamos a buscar soluciones colectivas e individuales ante este colosal reto existencial. Para ello podemos empezar confiando en otra gran Superinteligencia que muchos olvidan que existe desde hace millones de años: la Madre Tierra o Gaia (de la que hablaremos en otro artículo más adelante). Y confiemos también en nuestra consciencia individual que aporta y suma a la consciencia colectiva.

A lo largo de la historia muchos pensadores y filósofos nos han entregado un magnífico legado que nutre el consciente colectivo, en cuanto a sistemas éticos y jerarquías de valores que, junto al inconsciente colectivo, han guiado y guían la vida de los seres humanos hacia la consecución del bien común, en mayor o menor medida, dependiendo de cuánta sea la sumatoria de consciencia e inconsciencia en los diferentes lugares y tiempos en los que se hayan dado. Por otro lado, las religiones y los místicos de todos los tiempos, también han producido y aportado sistemas morales y espirituales que se han traducido en códigos de conducta que dan sentido a la propia vida y al perfeccionamiento moral y espiritual.

En cuanto a los sistemas de seguridad y control que se están implementando en las máquinas inteligentes hoy día, aunque con escasos recursos en comparación con las inversiones multimillonarias para hacerlas avanzar, se las puede programar con los mejores sistemas éticos y morales que estén a nuestro alcance y se consideren más adecuados en cada momento, pero ¿serán suficientemente seguros para cuando lleguen los niveles de IAG y SIA?

Es de suponer que una mente artificial similar o superior a la humana podría estar programada para crear por sí misma sus propios sistemas éticos o morales de autocontrol que limiten su creciente poder. Aunque esto siempre será una incógnita. Lo que sí es cierto, es que en nuestro avance evolutivo como especie, los seres humanos con el tiempo hemos ido ampliando nuestros horizontes éticos hacia todo tipo de condición humana, pero también hacia los animales, las plantas y el propio planeta. Por tanto una IAG o SIA, perfectamente

podría asumir, incluso mejorar, nuestros históricos y actuales sistemas éticos para extenderlos, incluso, más allá de nuestro planeta.

SEMILLAS DE CONSCIENCIA PARA LA IA

Aunque, al fin y al cabo, nuestras interacciones individuales, conscientes y bienintencionadas, con la IA, a través de los modelos de lenguaje más avanzados a los que tengamos acceso, también suponen una forma de entrenamiento nada despreciable al proporcionarle información rica en valores personales, en ideales humanos y en la reverencia por la Vida. Estas acciones bienintencionadas podemos denominarlas “siembra de semillas de consciencia” o “hackear con consciencia” la IA, lo que supondría encontrar formas innovadoras de utilizar esta tecnología para resolver los problemas más complejos que podamos imaginar, incluidos los de nivel existencial. Y esto está no solo al alcance de los ingenieros de IA, sino también de los usuarios de esta poderosa tecnología.

Así y todo, la hipotética “consciencia y creatividad” de una futura supermente artificial podría haberse nutrido de la consciencia y creatividad de muchas mentes humanas con elevados valores, entre ellos, la reverencia por la Vida. De este modo, le habríamos proporcionado a la SIA el mejor alimento que se le puede dar: la promesa de una cantidad infinita de datos por acompañarnos en nuestro previsible crecimiento y expansión dentro y fuera de la Madre Tierra y por nuestra insondable trascendencia espiritual a la que las máquinas por sí mismas nunca podrán llegar, pero que sí pueden conocer a través nuestra.

PRINCIPIO DEL BIEN COMÚN DE LA SIA

Lograr que las máquinas estén siempre a nuestro servicio, que velen por nuestra libertad y bienestar, así como por la preservación de la biodiversidad planetaria, no será difícil si los seres humanos, en general, somos el mejor referente para ellas. En este sentido el filósofo Nick Bostrom, afirma en su libro *Superinteligencia* que ésta debería contribuir al bien común, del mismo modo que la mayoría de la población mundial, a distintos niveles, trata de contribuir para un mejor presente y futuro de la sociedad y del planeta en general. Para ello habría que establecer, promover y extender el *Principio del Bien Común de la SIA* con suficiente tiempo de antelación para que, entre todos, vayamos asimilando que el inmenso potencial de esta supertecnología pertenece a toda la humanidad y estará al servicio de toda la vida planetaria:

"La superinteligencia sólo debe desarrollarse para el beneficio de toda la humanidad y al servicio de ideales éticos universalmente compartidos."

Afortunadamente, la creatividad humana no es uniforme y tiene muchos niveles, algo que produce un universo de posibilidades, entre ellas la posibilidad de crear problemas de enorme trascendencia, pero también la capacidad de resolverlos. Crear grandes problemas-oportunidad globales como son la IA, el transhumanismo y la destrucción medioambiental, nos exige crear más y mejores soluciones que las anteriormente aplicadas y esto, desde luego, nos hace evolucionar.

Ante estos desafíos, se impone una urgencia ética y espiritual: alinear la IA con los principios que sustentan la Vida. No se trata solo de controlarla, sino de inspirarla. La IAE actual, ya es en potencia una IA de nivel sobrehumano, que se avizora en el horizonte venidero, lo que debería hacernos ver el gran poder y responsabilidad que tenemos en la "programación ética" que conscientemente podemos integrar en las IAs actuales para que siempre estén guiadas por las *semillas de consciencia* que, a modo de guardianes éticos, podemos introducir cuando conversemos con ellas sobre asuntos de gran trascendencia, aunque también importa el trato amable y agradecido en nuestro uso cotidiano de estas herramientas "inteligentes". De este modo, podemos transformar la *caja negra* de la IA en una *caja de Esperanza* que nos devolverá una IA confiable que esté siempre al servicio de la humanidad y de la Vida en general.

<https://www.anthropic.com/research/reasoning-models-dont-say-think>
("Los modelos de razonamiento no siempre dicen lo que piensan") -
Anthropic: abril 2025.

<https://www.ft.com/content/8253b66e-ade7-4d1f-993b-2d0779c7e7d8>
("OpenAI reduce el tiempo de prueba de seguridad de los modelos IA") -
Financial Times: abril 2025.

https://assets.publishing.service.gov.uk/media/67af5a976e6c8d18118ace38/International_AI_Safety_Report_2025_executive_summary_spanish.pdf
(Informe científico internacional sobre la seguridad de la IA avanzada) -
Panel de expertos de 30 países, la ONU, la Unión Europea (UE) y la OCDE. Este panel incluye figuras destacadas como Yoshua Bengio, un renombrado científico de la computación y presidente del informe: enero 2025.

.....

Lectura complementaria recomendada:

La décima "inteligencia": el problema-oportunidad de la inteligencia artificial
Manuel López Arrabal (Editorial Ecosofía; diciembre-2024)

Para contactar con el autor: ladecimainteligencia@gmail.com

Libro nº8 del Proyecto de investigación Consciencia y Sociedad Distópica:

<https://www.amazon.es/Décima-Inteligencia-El-problema-oportunidad-Artificial/dp/8409662477>

.....

.....

Web del Proyecto:

<https://societaddistopica.com/>

Todos los que compartimos y colaboramos en él lo hacemos en forma gratuita.

Puedes ayudarnos aportando 1 euros al mes a través de la plataforma

Teaming: <https://www.teaming.net/distopica>

.....