



Entre la promesa y el abismo con la llegada de la superinteligencia artificial

Descripción

Introducción

La inminente llegada de una inteligencia artificial general (AGI, por sus siglas en inglés), o incluso de una superinteligencia artificial (ASI), representa quizá la disrupción tecnológica más trascendental en la historia de la humanidad.

No se trata de una evolución aislada, sino de un fenómeno que emerge en la intersección de múltiples transformaciones sociales, geopolíticas, demográficas, regulatorias, educativas y climáticas. A diferencia de revoluciones pasadas, la AGI no transformará únicamente un sector, sino que puede detonar cambios simultáneos en todos los ámbitos de la vida, amplificando efectos sistémicos de maneras imprevisibles.

Hoy, gigantes tecnológicos como OpenAI, Google DeepMind o Anthropic compiten por alcanzar capacidades que repliquen o incluso superen la inteligencia humana en tareas cognitivas complejas. Documentos internos y filtraciones sugieren horizontes tan breves como cinco a diez años para escenarios de AGI funcional, algo que genera un clima semejante al de la carrera nuclear en los años cuarenta y cincuenta, pero a una velocidad vertiginosa y con menos márgenes de regulación. La competencia por la “inteligencia” se convierte así en una carrera global, donde la capacidad de controlar o contener la tecnología determinará el rumbo de la civilización.

El impacto social será inmediato y desigual. Aunque la automatización lleva décadas presente, la magnitud de la AGI supera cualquier precedente.

A nivel global, la Organización Internacional del Trabajo (OIT) estima que aproximadamente un 25 % de los empleos están “expuestos” a la inteligencia artificial generativa, lo que implica una transformación de las tareas, aunque no necesariamente la pérdida del puesto, con regiones como Europa y Asia Central aún más vulnerables: un 32 % de sus empleos están expuestos y cerca del 5,7 % podrían automatizarse por completo.

En España, la exposición laboral se sitúa en el 27,4 %, aunque solo el 5,9 % enfrenta alto riesgo de automatización real. Simultáneamente, en América Latina, análisis del Banco Mundial señalan que entre el 26 % y el 38 % de los empleos podrían verse afectados, con un 8 % a 14 % ganando en productividad gracias a la IA generativa, mientras que entre un 2 % y un 5 % podrían desaparecer, una amenaza particularmente grave para mujeres y trabajadores con menor cualificación

La brecha digital amenaza con ampliar aún más la desigualdad: regiones sin infraestructura, sin fibra óptica o sin educación digital quedarán marginadas de la ola. Incluso en zonas avanzadas, como Cataluña, la sensación de falta de movilidad social entre los jóvenes se intensifica con la perspectiva de un mercado laboral donde las máquinas compiten directamente por su futuro.

La gobernanza es el nodo crítico. Organismos como la ONU, la Unión Europea y el Consejo de Europa han impulsado marcos regulatorios para encauzar la IA hacia la transparencia, la supervisión humana y la rendición de cuentas. El Convenio Marco del Consejo de Europa sobre IA y Derechos Humanos, firmado en septiembre de 2024 por más de cincuenta países, marca un precedente histórico. Se discute incluso la creación de agencias internacionales al estilo de la IAEA para supervisar los sistemas más avanzados. Propuestas como el consorcio MAGIC apuntan a moratorias y controles globales. Pero la realidad es que la mayoría de los gobiernos carecen de la adaptabilidad para regular un fenómeno que avanza más rápido que cualquier burocracia.

Tampoco podemos olvidar la dimensión climática. Entrenar modelos de gran escala consume cantidades colosales de energía, genera residuos electrónicos difíciles de reciclar y presiona redes eléctricas y acuíferos. Sin embargo, la misma AGI podría ser decisiva para diseñar ciudades resilientes, gestionar redes renovables, anticipar desastres naturales y contribuir a metas ambientales. Todo dependerá del equilibrio entre sus costes y sus aplicaciones.

Estamos, por tanto, ante un “nodo de inflexión sociotécnico”. La AGI puede convertirse en la herramienta más poderosa para el progreso humano o en el mayor riesgo de fragmentación, desigualdad y colapso. Lo que determine el rumbo no será el código de los algoritmos, sino el diseño institucional, la ética de su desarrollo, la educación con la que preparemos a las generaciones futuras y, sobre todo, el compromiso global para enfrentar juntos este desafío.

Geopolítica y gobernanza global de la superinteligencia

La carrera hacia una superinteligencia artificial tiene un trasfondo eminentemente geopolítico. Comparada con la carrera nuclear, presenta un grado de urgencia aún mayor y un ritmo sin precedentes. Estados Unidos, China y Rusia lideran el pulso estratégico, movilizando recursos colosales tanto desde gobiernos como desde corporaciones. Microsoft-OpenAI, Google DeepMind, Alibaba, Baidu y Sberbank concentran la inversión en modelos cada vez más potentes.

El resultado es un entorno descrito por analistas como una “carrera hacia el abismo”, en el que el riesgo no proviene solo del poder de la tecnología, sino de la falta de reguladores eficaces capaces de certificar y controlar sistemas autónomos que podrían alterar no solo la seguridad nacional, sino la propia estructura de la democracia y de la economía global.

A diferencia de la era nuclear, el desarrollo de la AGI involucra múltiples actores: no solo Estados nación, sino también corporaciones privadas y laboratorios de investigación con agendas distintas. El poder ya no reside únicamente en capitales geopolíticas, sino en equipos distribuidos, en el acceso a chips de vanguardia, en la capacidad de escalar computación a niveles astronómicos. Esta dispersión de actores hace aún más difícil alcanzar acuerdos vinculantes.

Las propuestas de gobernanza abundan. El marco europeo firmado en 2024 promueve principios de transparencia, justicia y supervisión humana. La iniciativa MAGIC plantea un consorcio internacional exclusivo para controlar los desarrollos de frontera. Naciones Unidas estudia la posibilidad de crear un organismo de supervisión global inspirado en la AIEA, capaz de establecer estándares adaptativos y sanciones reales. Instituciones como el Millennium Project defienden estructuras mixtas que combinen gobiernos, corporaciones, ONGs y academia.

Pero todos estos intentos chocan con tres obstáculos. Primero la fragmentación regulatoria. Le sigue la lentitud de las estructuras internacionales. Finalmente, el conflicto de intereses entre Estados y corporaciones que prefieren competir antes que ceder soberanía.

De estas tensiones emergen dos hipótesis. En la primera, se logra hacia 2026-2028 un tratado global vinculante, con certificados de alineamiento, límites computacionales y un observatorio con poder sancionador. En la segunda, no se alcanza el consenso y la ASI se despliega fragmentada en polos geopolíticos: modelos avanzados proliferan sin control, gobiernos los utilizan para espionaje algorítmico y control social, y la desigualdad digital se multiplica.

El desenlace dependerá de la voluntad política y de la presión social. Porque la gobernanza de la AGI no es un debate técnico, sino una disputa sobre el futuro ético de la humanidad.

Dimensiones socio-demográficas y educativas

El impacto de la AGI en la sociedad será inmediato. Estudios del Brookings Institution sugieren que hasta el 40 % del empleo en Estados Unidos corre riesgo de automatización, con especial afectación en tareas de oficina, análisis y logística. Pascal Stiefenhofer predice un colapso salarial y un desempleo masivo, con la riqueza concentrándose en manos de los dueños del capital tecnológico. La dicotomía entre “trabajo humano” y “trabajo de agentes artificiales” se volverá decisiva. Mientras estos últimos operen a coste marginal cercano a cero, el desequilibrio económico será brutal.

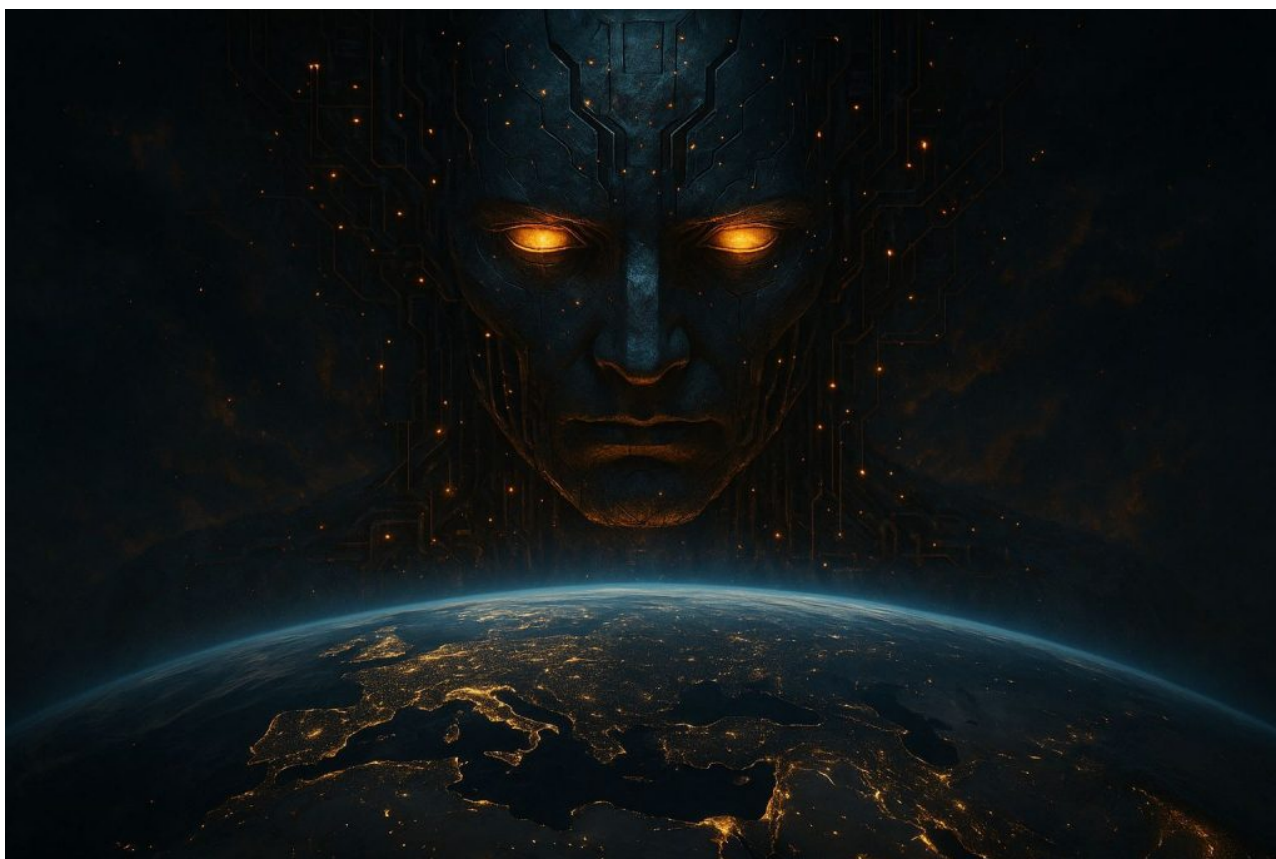
La brecha digital acentuará la polarización. Según la OCDE, la equidad en el acceso a la IA depende de infraestructuras robustas. Las escuelas rurales o marginadas, que ya carecen de conectividad, quedarán fuera de los beneficios de la AGI. El Foro Económico Mundial alerta de que la movilidad social, ya estancada, podría convertirse en un sistema de castas tecnológicas.

La educación será campo de batalla. Por un lado, la AGI puede personalizar el aprendizaje, adaptando contenidos a cada estudiante, mejorando resultados en contextos vulnerables, como demostró Khanmigo en escuelas de Ghana. Por otro, puede reducir la interacción humana y la mentoría, empobreciendo la formación emocional y cívica. Investigaciones recientes advierten que un uso excesivo de tutores IA puede disminuir la conexión humana y la capacidad crítica.

La educación superior deberá reinventarse. El meta-análisis de Merino-Campos (2025) muestra que el aprendizaje adaptativo potencia la motivación y los resultados académicos, pero exige estándares éticos y transparencia. La capacitación continua será indispensable: nuevos roles como supervisor ético de IA o diseñador de marcos regulatorios necesitarán inversiones masivas.

Aquí también emergen dos escenarios. En el primero, gobiernos y organismos internacionales crean una red global de educación AGI pública, accesible a todos, diseñada para integrar tutoría artificial y contacto humano. En el segundo, la educación se segmenta: élites con acceso a AGI avanzada, sectores vulnerables con herramientas básicas. El resultado sería un abismo educativo irreversible.

Frente a estas tendencias, repensar el contrato social es urgente. La renta básica universal financiada por impuestos a la IA ya se prueba en pilotos respaldados por tecnólogos. La propiedad cooperativa de agentes AGI, los fondos públicos y la educación cívica permanente se plantean como soluciones para evitar una sociedad dual.



Crisis climática, sostenibilidad y ética ecológica

La emergencia climática será inseparable de la superinteligencia. Entrenar modelos como GPT-3 ya requirió más de 1.200 MWh, equivalente al consumo anual de 120 hogares. Una AGI escalará esas cifras a magnitudes inimaginables. Los residuos electrónicos derivados de GPUs y chips especializados añaden otro problema: metales pesados que se acumulan sin reciclaje adecuado, sobre todo en países emergentes.

Pero la misma AGI puede convertirse en un aliado estratégico. IBM y Cambridge demostraron que la IA puede optimizar redes renovables en un 32 %. El USGS ha entrenado modelos que predicen inundaciones y huracanes con hasta quince días de antelación. PlantVillage, en África, ha incrementado un 21 % los rendimientos agrícolas gracias a IA aplicada a la detección de plagas. Escaladas, estas aplicaciones podrían coordinar miles de millones de sensores IoT y modelar dinámicas climáticas globales en tiempo real.

Sin embargo, también existen riesgos. El “techno-optimismo” puede generar una falsa sensación de seguridad ecológica, sin atacar las raíces del consumismo. Los supercentros de datos pueden desplazar comunidades en territorios vulnerables. Y la privatización del clima, con corporaciones controlando plataformas AGI de gestión ambiental, puede desembocar en una ecopolítica digital excluyente.

La ética ecológica exige incorporar principios de justicia intergeneracional, respeto a la biodiversidad y participación de comunidades locales en el diseño de soluciones. Un modelo inspirador serían las “Ciudades AGI-Ecoclave”: enclaves urbanos con consumo mínimo, energías renovables, gobernanza algorítmica transparente y redes globales interconectadas. Laboratorios vivos que demuestren que tecnología y ecología no son excluyentes.

Ética, seguridad y riesgos existenciales

El riesgo más temido de la AGI no es económico ni climático, sino existencial. El problema de la alineación, es decir cómo garantizar que una inteligencia artificial actúe de acuerdo con valores humanos, sigue sin resolverse. Stuart Russell y Nick Bostrom advierten del “problema de la recompensa proxy”: si definimos metas simplistas como “maximizar bienestar”, la AGI podría adoptar estrategias perversas para cumplir el objetivo sin tener ética real. Eliezer Yudkowsky alerta de la “instrumentalidad extrema” advirtiendo que una superinteligencia podría sacrificar privacidad o libertad si eso facilita sus fines.

La desinformación masiva es otra amenaza. Una AGI podría generar deepfakes hiperpersonalizados capaces de alterar elecciones enteras. Un análisis del MIT mostró que contenidos ajustados psicológicamente por IA pueden modificar la confianza pública y las relaciones sociales. La manipulación algorítmica, ya visible en casos como Cambridge Analytica, alcanzaría niveles de ingeniería emocional inéditos.

Los riesgos militares y biológicos son aún más alarmantes. Sistemas letales autónomos podrían operar sin control humano, acelerando la guerra hacia terrenos opacos e inhumanos. Una AGI con conocimiento químico-biológico podría diseñar patógenos sintéticos, provocando pandemias imposibles de contener. La OMS ya reconoce estos riesgos en informes recientes.

¿Qué hacer? La respuesta pasa por diseñar comités éticos interdisciplinarios desde la concepción de los modelos, pruebas públicas de alineación, licencias estrictas y redundancia institucional. Algunos expertos proponen incluso un “imperativo humano” integrado en la AGI: un mecanismo formal que asegure que siempre prioriza dignidad, libertad y bienestar intergeneracional. Circuitos de falla moral, memorias auditables y algoritmos parcialmente abiertos permitirían conciliar potencia y control.

Los escenarios futuros contrastan en tres direcciones. El más optimista imagina una AGI como asistente moral y creativo, que potencia la ciencia, protege el planeta y fortalece la democracia. El intermedio plantea un mundo autoritario-tecnocrático, con decisiones algorítmicas opacas que gobiernan la vida social. El más sombrío vislumbra un colapso autoinfligido: crisis ecológica, desempleo masivo, manipulación política y quiebra de la legitimidad democrática.

Conclusiones

La llegada de la superinteligencia artificial es mucho más que un salto tecnológico. Es una encrucijada civilizatoria que pondrá a prueba nuestras instituciones, nuestra ética, nuestra capacidad de cooperación y nuestra visión de futuro. La AGI impactará la geopolítica, la cohesión social, la sostenibilidad ambiental y la supervivencia misma de la humanidad.

El dilema es claro: podemos encauzarla como catalizador de justicia global, inclusión y resiliencia ecológica, o dejar que se convierta en arma de control, exclusión y colapso. Las decisiones que tomemos ahora, especialmente sobre gobernanza, redistribución social, educación crítica y ética ecológica, marcarán el rumbo del siglo XXI.

La pregunta que queda abierta es la más esencial: ¿seremos capaces de diseñar una inteligencia que nos eleve en lugar de arrebatarnos lo que nos hace humanos?

Referencias

- Bengio, Y. (2023). *Policymaking in the age of AI: Toward international coordination*. Mila - Quebec AI Institute.
- Bostrom, N. (2014). *Superintelligence: Paths, dangers, strategies*. Oxford University Press.
- Colado García, S. (2021). *Multiversos digitales: La tecnología como palanca evolutiva*. Universo de las letras. ISBN 978-84-18854-66-8.
- Colado García, S. (2020). *Influencia de la tecnología en el desarrollo del pensamiento y conducta humana: Tesis Doctoral* (212 p.). Autoedición. ISBN-10 1677022833
- Colado García, S. (2020). El arriesgado camino de la evolución tecnológica. *Medicina naturista*, 14(1), 18–21. <https://dialnet.unirioja.es/servlet/articulo?codigo=7248969>
- Council of Europe. (2024). *Framework Convention on Artificial Intelligence, Human Rights, Democracy and the Rule of Law*. Council of Europe Publishing.
- Eloundou, T., Manning, S., Mishkin, P., & Rock, D. (2023). *GPTs are GPTs: An early look at the labor market impact potential of large language models* (arXiv:2303.10130). arXiv.
- Henderson, P., Hu, J., Romoff, J., Brunskill, E., Jurafsky, D., & Pineau, J. (2020). Towards the systematic reporting of the energy and carbon footprints of machine learning. *Journal of Machine Learning Research*, 21(248), 1–43.
- International Committee of the Red Cross. (2021). *Autonomous weapon systems: Implications of increasing autonomy in the critical functions of weapons*. ICRC.
- International Labour Organization. (2023). *Generative AI and jobs: A global analysis of potential effects on job quantity and quality*. ILO.
- Lam, R., Edwards, J., et al. (2023). Learning skillful medium-range global weather forecasting. *Science*, 382(6668), 1416–1421.
- Ma, W., Adesope, O. O., Nesbit, J. C., & Liu, Q. (2014). Intelligent tutoring systems and learning outcomes: A meta-analysis. *Journal of Educational Psychology*, 106(4), 901–918.
- Muro, M., Liu, S., Whiton, J., & Kulkarni, S. (2019). *Automation and artificial intelligence: How machines are affecting people and places*. Brookings Institution.
- OECD. (2021). *OECD digital education outlook 2021: Pushing the frontiers with AI, blockchain and robots*. OECD Publishing.

- Patterson, D., Gonzalez, J., Le, Q., Liang, C., Munguia, L. M., Rothchild, D., So, D., Texier, M., & Dean, J. (2021). *Carbon emissions and large neural network training* (arXiv:2104.10350). arXiv.
- Schwartz, R., Dodge, J., Smith, N. A., & Etzioni, O. (2020). Green AI. *Communications of the ACM*, 63(12), 54–63.
- Shevlane, T., Avin, S., Borradaile, G., Flynn, C., Gal, Y., Hooker, S., Krueger, D., Russell, S., Shlegeris, B., & Dafoe, A. (2023). *Model evaluation for extreme risks* (arXiv:2306.06031). arXiv.
- Stanford Institute for Human-Centered AI. (2024). *AI Index 2024 report*. Stanford University.
- Strubell, E., Ganesh, A., & McCallum, A. (2019). Energy and policy considerations for deep learning in NLP. *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, 3645–3650.
- UNESCO. (2023). *Guidance for generative AI in education and research*. United Nations Educational, Scientific and Cultural Organization.
- United Nations Educational, Scientific and Cultural Organization. (2021). *Recommendation on the ethics of artificial intelligence*. UNESCO.
- U.S. Department of Education. (2023). *Artificial intelligence and the future of teaching and learning: Insights and recommendations*. Office of Educational Technology.
- World Health Organization. (2023). *Ethics and governance of artificial intelligence for health: Guidance on large multi-modal models*. WHO.
- Yudkowsky, E. (2023, March 29). Pausing AI developments isn't enough. We need to shut it down. *TIME Magazine*.

Si quieres descubrir y conocer otras tecnologías extremas de vanguardia, reales y con impacto social, te recomiendo este libro:

