



El rol de la UNAM frente al país en el equilibrio de la balanza entre los peligros y beneficios de la inteligencia artificial

Carlos A. Coello Coello

Instituto Politécnico Nacional, Centro de Investigación y de Estudios Avanzados, Departamento de Computación, Ciudad de México, México.
ORCID: [0000-0002-8435-680X](https://orcid.org/0000-0002-8435-680X)

Guy Paic

Universidad Nacional Autónoma de México, Instituto de Ciencias Nucleares, Ciudad de México, México.
ORCID: [0000-0003-2513-2459](https://orcid.org/0000-0003-2513-2459)

Leonid Serkin

Universidad Nacional Autónoma de México, Instituto de Ciencias Nucleares, Ciudad de México, México.
ORCID: [0000-0003-4749-5250](https://orcid.org/0000-0003-4749-5250)

Recepción: 20 de marzo de 2024.

Aceptación: 13 de mayo de 2024.

Junio 2024 • número de revista 10 • <https://doi.org/10.22201/dgtic.26832968e.2024.10.10>

Este es un artículo de acceso abierto bajo la licencia Creative Commons Atribución-No Comercial 4.0 Internacional (CC BY-NC 4.0).

2683-2968/© 2024 UNAM. TIES, Revista de Tecnología e Innovación en Educación Superior es editada por la Universidad Nacional Autónoma de México a través de la Dirección General de Cómputo y de Tecnologías de Información y Comunicación. ISSN: 2683-2968. Reserva de Derechos de Autor: 04-2019-011816190900-203

El rol de la UNAM frente al país en el equilibrio de la balanza entre los peligros y beneficios de la inteligencia artificial

Resumen

En 2019, GPT-2 no podía contar de manera confiable hasta diez. Sólo cinco años después, los sistemas de inteligencia artificial (en adelante, IA) generativa pueden escribir software, asesorar sobre temas científicos y combinar el procesamiento de lenguaje e imagen para generar escenas fotorrealistas, acercándose más y más a la inteligencia de un humano. Tampoco cabe duda de que la IA encarna una dualidad de riesgos y beneficios para la sociedad. La pregunta importante, entonces, es qué significa en términos de consecuencias y qué se podría hacer a nivel de la Universidad Nacional para aprovechar los beneficios y tratar de gestionar los riesgos. El artículo tiene por objeto analizar el papel crucial que tiene la UNAM en el equilibrio de esta balanza, promoviendo un desarrollo y aplicación de la IA que sea ético, equitativo y lleno de beneficios para México y el mundo.

Palabras clave: peligros, beneficios, inteligencia artificial, UNAM, México.

The role of UNAM in facing the nation's challenges of balancing the risks and benefits of artificial intelligence

Abstract

In 2019, GPT-2 was unable to accurately count up to ten. Just five years later, generative artificial intelligence (hereinafter, AI) systems can write software, advise on scientific topics, and combine language and image processing to generate photorealistic scenes, coming closer and closer to human-level intelligence. There is also no doubt that AI embodies a duality of risks and benefits for society. The important question, then, is what this means in terms of consequences and what could be done at the National University level to harness the benefits and try to manage the risks. This article aims to analyze the crucial role that the National Autonomous University of Mexico (UNAM) plays in balancing this scale, promoting ethical, equitable, and beneficial development and application of AI for Mexico and the world.

Keywords: risks, benefits, artificial intelligence, UNAM, Mexico.

Los cuatro pilares de la inteligencia artificial

La inteligencia artificial (IA) puede ser la tecnología que moldee este siglo, con el potencial de rediseñar casi todos los aspectos de la vida como la conocemos. Usamos la IA en nuestra vida diaria, desde pedirle canciones a Alexa [1], hasta recibir recomendaciones de películas de Netflix [2] o utilizar reconocimiento facial para desbloquear nuestros teléfonos inteligentes [3].

La capacidad de la IA para analizar grandes volúmenes de datos es fundamental para abordar problemas complejos relacionados con la educación, la investigación, las ciencias y las humanidades. Su integración en diversos campos del conocimiento y la industria promete avances sin precedentes en eficiencia, capacidad de innovación y solución de problemas complejos, vean por ejemplo [4]; [5]; [6] y sus referencias.

La IA se sostiene sobre cuatro pilares fundamentales, indispensables para su avance y efectiva integración en la sociedad. El primer pilar es la formación y disponibilidad de personal especializado en IA en cada uno de sus campos de aplicación. Un claro ejemplo lo encontramos en el informe del Departamento de Educación de los Estados Unidos, publicado en 2023, que detalla el impacto de la IA en la educación [7]. Asimismo, el plan de la Universidad de Florida para graduar a miles de estudiantes con capacitación en IA, preparados para integrar estas tecnologías en el mercado laboral, es otra muestra significativa [8].

El segundo pilar, los datos, constituye una parte crucial, ya que la IA se nutre del análisis extenso de estos para aprender y tomar decisiones. La manera en que se recolectan los datos es fundamental, subrayando la importancia de la soberanía de datos por país. Este aspecto es particularmente relevante en sectores como la medicina [9] o la climatología [10], donde las patologías o la precipitación pueden variar notablemente de una región a otra. Además, asegurar el almacenamiento de datos implica implementar medidas robustas no solo de ciberseguridad, sino también de privacidad y acceso [11].

El tercer pilar se centra en el acceso y desarrollo de software abierto, esencial para el funcionamiento y desarrollo de la IA, lo que permite la implementación de algoritmos y sistemas de IA en un amplio espectro de aplicaciones y servicios [12]; [13]; [14].

El cuarto y último pilar esencial es la disponibilidad de recursos de computación de alto rendimiento y vanguardia [15]. El acceso a este tipo de infraestructura avanzada es un requisito indispensable para que cualquier nación pueda considerarse parte integral de la comunidad global de IA. Su importancia radica en la capacidad de realizar análisis de datos a gran escala y entrenar modelos de IA complejos en áreas como la salud, la educación, la industria y la investigación científica.

¿Cómo lograr que la IA no rompa la primera ley de Asimov?

A lo largo del último año (2023), la conversación en torno a la IA ha evolucionado de manera significativa, abarcando tanto sus promesas como sus desafíos [16]; [17]. Han surgido preguntas sobre la ética y las implicaciones sociales de la IA, en particular, ¿quién la desarrolla, para quién y con qué propósitos? ¿Qué posibles consecuencias puede traer consigo? ¿Cómo evitar que se utilice de manera perjudicial, respetando la primera ley de Asimov [18]: “Una máquina no hará daño a un ser humano, ni por inacción permitirá que un ser humano sufra daño”?

Uno de los conceptos fundamentales en la seguridad de la IA es el problema de la alineación [19]. Es importante destacar que, en la actualidad, la ciencia no sabe cómo diseñar un sistema de IA que produzca soluciones o respuestas completamente predecibles. Existe una discrepancia entre lo que hacen esos sistemas de IA y nuestras intenciones, e incluso no está claro si esto es algo que se pueda resolver.

Hemos observado problemas de alineación en las respuestas de los sistemas de IA actuales, como discriminación, sesgo y falta de equidad [4], que se originan debido a la implementación de la IA y la manera en que las decisiones no siempre están alineadas con los valores humanos. Estos problemas no surgen necesariamente debido a malas intenciones por parte de los desarrolladores de IA, sino más bien porque aún no han encontrado una forma de diseñar sus sistemas para que se comporten de acuerdo con nuestras normas [20]. A medida que los sistemas de IA se vuelven más poderosos, especialmente si caen en las manos equivocadas y no los configuramos adecuadamente para evitar amenazas, la IA podría ser un riesgo catastrófico a nivel global [21]; [22].

El otro problema es que, incluso si resolvemos la cuestión de la seguridad y la alineación, ¿quién tiene el control de estos sistemas de IA? Incluso si se comportan según lo previsto por quienes los operan, aquellos que los manejan pueden hacer un uso inadecuado de estos sistemas en muchas áreas, ya sea en educación, investigación, democracia o seguridad nacional [23].

Enfoque multilateral sobre las estrategias de mitigación

Estas preocupaciones evidenciaron la necesidad de una solución colaborativa y multilateral para enfrentar los desafíos globales relacionados con esta nueva tecnología. La redacción de los “Principios de Asilomar para la Inteligencia Artificial” en 2017 [24], la “Declaración de Montreal para el Desarrollo Responsable de la Inteligencia Artificial” en 2018 [25], los “Principios de la Organización para la Cooperación y el Desarrollo Económicos (OCDE) sobre Inteligencia Artificial” en 2019 [26], la recomendación de la Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura (UNESCO) sobre la “Ética de la Inteligencia Artificial” en 2021 [27], las políticas sobre el uso de la IA en favor de la infancia del Fondo de las Naciones Unidas

para la Infancia [28], y los estudios de la OCDE sobre el uso estratégico y responsable de la IA en el sector público de América Latina y el Caribe en 2022 [29], representan solo los primeros pasos en este proceso.

En particular, diversos países de América Latina han avanzado significativamente en el desarrollo de estrategias nacionales de IA, con el objetivo de impulsar su desarrollo socioeconómico, fortalecer su soberanía tecnológica y garantizar la protección de los derechos de sus ciudadanos. Por ejemplo, la política argentina considera posibles aplicaciones de la IA para mejorar las condiciones de las personas con discapacidad [30]. La estrategia de Brasil prioriza la formación de niños en tecnologías de IA, considerando sus necesidades específicas y protegiendo sus derechos [31]. Chile ha establecido como uno de sus objetivos principales la adopción de la IA para mitigar los efectos del cambio climático [32]. En el caso de Perú, su estrategia incluye entre sus objetivos estratégicos la reducción de la brecha de género en la formación en IA, así como la inclusión de minorías y lenguas nativas [33], entre otros aspectos relevantes.

Tras la presentación pública de GPT-3.5 y GPT-4 [34], muchos expertos en el campo de la IA han abogado por una pausa temporal en el desarrollo de sistemas de IA aún más avanzados [35]. Esta discusión se enmarca dentro de la evolución de modelos de lenguaje [4] como GPT-2, que en 2019 presentaba avances significativos en capacidades de procesamiento de lenguaje natural, pero mostraba limitaciones significativas en tareas simples como aritmética [36]. Esto llevó a la Organización de las Naciones Unidas (ONU) a establecer a finales de 2023 un Consejo Consultivo, con el propósito de analizar y promover recomendaciones para la gobernanza internacional de la IA [37].

Como resultado, las tres principales recomendaciones para manejar el crecimiento del poder de los sistemas de IA en los próximos 2 a 5 años son: (1) regulación y tratados internacionales, (2) investigación en seguridad de la IA y (3) desarrollo de contramedidas. La primera recomendación propone que los desarrolladores de IA tengan la responsabilidad de demostrar a la comunidad y a la sociedad que sus sistemas son lo suficientemente seguros, en lugar de que el gobierno y/o los usuarios finales tengamos que encontrar qué está mal con ellos. La segunda recomendación enfatiza la necesidad de una comprensión de los riesgos y las estrategias de mitigación. Por ejemplo, ¿cuál es el umbral de lo que es seguro y qué no lo es, y cómo lo medimos? Esto es algo para lo que aún no tenemos respuestas. Y la tercera: preparar un plan B en caso de que los sistemas de IA tengan objetivos peligrosos y sean lo suficientemente poderosos como para amenazar nuestras vidas y/o seguridad nacional.

Hay un concepto importante aquí en lo que respecta a los riesgos, y es que se necesita evitar el punto único de fallo [38]. Esto significa evitar la concentración de poder en un solo sistema de IA, que podría caer en manos equivocadas o salir fuera de control debido a un error humano [39].

Índice Latinoamericano y las iniciativas de IA en México

En el contexto de nuestro país, es importante destacar que México fue uno de los primeros 10 países del mundo en presentar en 2018 una “Estrategia Nacional de Inteligencia Artificial” [40]. Esta estrategia tenía como objetivo establecer un marco para promover el desarrollo y la adopción de la IA en México, reconociendo su potencial transformador en diversos sectores, desde la salud hasta la educación y la economía.

Asimismo, en 2018 México participó en el desarrollo de normas mundiales por la Organización Internacional de Normalización (ISO) y fue un país observador en el establecimiento de las primeras normas ISO/IEC JTC 1/SC 42 para la IA [41]. Más aún, en 2020 México fungió como miembro fundador de la Alianza Global sobre la Inteligencia Artificial [42], una iniciativa de cooperación global sobre la innovación y gobernanza de la IA que tiene participación continua en foros internacionales como el G20, G7, OCDE, UNESCO y otros organismos de la ONU [43].

Paradójicamente, a pesar de que la estrategia nacional fue firmada hace seis años, según el Índice Latinoamericano de IA (ILIA) de 2023, México ocupa el quinto lugar entre 12 países de la región en cuestiones de infraestructura, capital humano, desarrollo, adopción y regulación [44]. Las conclusiones del índice ILIA, preparado por el Centro Nacional de IA de Chile, con el apoyo de la Comisión Económica para América Latina y el Caribe, la Organización de los Estados Americanos y la UNESCO señalan [45]: “en cuanto al capital humano avanzado, si bien tiene un sistema de formación maduro y robusto, (México) se encuentra por debajo del promedio regional en todos los indicadores al normalizar por población”. El informe [45] también menciona que México debe “implementar una estrategia sólida a nivel nacional y fortalecer la gobernanza, participación ciudadana y la regulación en el campo de la IA”.

Tres años después de adherirse como estado miembro al acuerdo de la UNESCO sobre la ética de la IA [27], nuestro país aún no ha implementado los mecanismos regulatorios necesarios para cumplir con los compromisos relacionados con la IA. Hasta la fecha no se han establecido normas oficiales mexicanas (NOMs) específicas para la IA, según el Instituto Federal de Telecomunicaciones [46].

México está en una posición rezagada en comparación con otros países del continente americano y del mundo en cuanto a la regulación y la gestión de los impactos de la IA en la vida cotidiana. Las iniciativas de la Ley Federal de Ciberseguridad [47] y la Ley de Regulación Ética de la Inteligencia Artificial y la Robótica [48] fueron presentadas ante el Senado de la República apenas en 2023, y se espera que los dictámenes correspondientes se emitan hasta mediados del año 2024.

El rol de la Universidad Nacional al frente del desafío

Esta delicada situación resalta una oportunidad para que instituciones como la Universidad Nacional Autónoma de México (UNAM) encabecen el camino en la promoción de la IA a nivel nacional [49]. Como líder en educación e investigación dentro del país, la UNAM no debe limitarse a observar pasivamente los desarrollos tecnológicos. Es imperativo que emplee todas sus capacidades para establecerse como un referente en la incorporación de la IA en procesos educativos y en la aplicación de métodos de IA en la investigación. Esto implica no solo formar a las futuras generaciones de estudiantes capaces de integrar estas nuevas tecnologías en el mercado laboral mexicano, sino también investigar las implicaciones sociales y éticas, así como liderar el desarrollo de políticas públicas que promuevan un manejo responsable de la IA en el país.

Es fundamental que la UNAM difunda en todas sus facultades y escuelas las recomendaciones del uso de la IA en la educación superior [50]; [51], y especialmente aquellas surgidas de la “I Jornada de Inteligencia Artificial Generativa en Educación de la UNAM” [52]. Al promover estas recomendaciones entre estudiantes y profesores de la UNAM, se busca sensibilizar a la comunidad universitaria sobre los principios éticos que deben guiar el desarrollo y la implementación de tecnologías de IA generativa en la educación [53].

La disponibilidad de tecnología de computación de punta es crucial para apoyar de manera efectiva el desarrollo y la ejecución de sistemas de IA complejos [54], así como la contribución a la actualización y modernización de los criterios de ciberseguridad [55]; [56].

Así mismo, es crucial que la UNAM establezca y mantenga alianzas interinstitucionales para promover el desarrollo y uso de la IA en la investigación, así como fomentar la vinculación y la transferencia de conocimientos apoyando el desarrollo de capacidades tecnológicas [57]; [58].

Por otro lado, vale la pena reevaluar el código de ética [59] desde una perspectiva crítica sobre los desafíos planteados por la integración de la IA [46], así como de las implicaciones bioéticas estudiadas por la Comisión Nacional de Bioética de la Secretaría de Salud [60].

Es imperativo también dar continuidad a discusiones entre expertos nacionales y extranjeros sobre los aspectos científicos, aplicaciones, cuestiones éticas y legales de la IA, tales como los simposios internacionales “Inteligencia artificial para la ciencia, la industria y la sociedad”, organizados por la UNAM, el Centro Europeo de Investigaciones Nucleares y la OCDE [61].

La organización de una mesa redonda sobre el uso de la IA, con la participación de directores de facultades, escuelas e institutos afiliados a la UNAM, puede servir como un punto de encuentro clave para generar estrategias y recomendaciones. La finalidad de la mesa redonda sería la formación de un cuerpo multidis-

ciplinario compuesto por expertos de diversas áreas académicas y disciplinas relevantes para la IA, como informática, matemáticas, ingeniería, ética, derecho, sociología, entre otras, dedicado a plantear respuestas a los desafíos tales como: ¿cuáles son las estrategias específicas para integrar con éxito la IA en los programas educativos de la UNAM? ¿O cómo acelerar el desarrollo científico gracias al uso de la IA en el análisis de datos?

Y finalmente, una iniciativa para establecer un instituto multidisciplinario de inteligencia artificial fortalecería la posición de la UNAM como un centro de colaboración internacional en investigación y desarrollo de IA. La creación del instituto con un enfoque similar al del Instituto A3D3 en EE.UU. [62], el Centro Nacional de IA de Chile [45] o el Instituto de Aprendizaje de Algoritmos de Montreal [63], orientado al desarrollo de talentos y la articulación público-privada, posicionaría a la UNAM a la vanguardia de la investigación científica y tecnológica sobre IA no solo a nivel nacional, sino también a nivel mundial.

Conclusión

El uso de la inteligencia artificial ha emergido como una fuerza transformadora con un potencial significativo en casi todos los aspectos de nuestra vida cotidiana, pero indudablemente presenta una dualidad de riesgos y beneficios para la sociedad. A pesar de la temprana participación de México en el desarrollo de normas e iniciativas globales sobre la IA, nuestro país sigue rezagado en infraestructura, capital humano, desarrollo, adopción y regulación de la IA.

Reconociendo la necesidad de una regulación ética y responsable, así como la promoción de la investigación y la formación en IA, la UNAM debe liderar el camino hacia un uso inclusivo y equitativo de esta tecnología y desempeñar un papel crucial en el equilibrio entre los riesgos y beneficios de la IA en México. Al promover las recomendaciones para el uso de la IA en la docencia y la investigación, estableciendo y manteniendo alianzas y colaboraciones interdisciplinarias e internacionales, así como formando un cuerpo e instituto multidisciplinario dedicado a la IA, la UNAM puede posicionarse como un centro de excelencia en la investigación y desarrollo de IA, contribuyendo al avance de la sociedad mexicana.

Referencias

- [1] Amazon Alexa, "What Is Conversational AI?," 2024. [En línea]. Disponible: <https://developer.amazon.com/en-US/alexa/alexa-skills-kit/conversational-ai>.
- [2] Netflix Research, "Machine Learning," 2024. [En línea]. Disponible: <https://research.netflix.com/research-area/machine-learning>.

- [3] Samsung SDS. "Vision API," 2024. [En línea]. Disponible: <https://www.samsungsds.com/en/ai-vision-api/vision-api.html>.
- [4] T. B. Brown, B. Mann, N. Ryder, M. Subbiah, J. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell, S. Agarwal, A. Herbert-Voss, G. Krueger, T. Henighan, R. Child, A. Ramesh, D. M. Ziegler, J. Wu, C. Winter, et al., "Language Models are Few-Shot Learners," 2020. [En línea]. Disponible: <https://doi.org/10.48550/arXiv.2005.14165>.
- [5] J. Wei, Y. Tay, R. Bommasani, C. Raffel, B. Zoph, S. Borgeaud, D. Yogatama, M. Bosma, D. Zhou, D. Metzler, E. H. Chi, T. Hashimoto, O. Vinyals, P. Liang, J. Dean, y W. Fedus, "Emergent abilities of large language models," 2022. [En línea]. Disponible: <https://doi.org/10.48550/arXiv.2206.07682>.
- [6] L. Wang, C. Ma, X. Feng, Z. Zhang, H. Yang, J. Zhang, Z. Chen, J. Tang, X. Chen, Y. Lin, W. X. Zhao, Z. Wei, y J. R. Wen, "A Survey on Large Language Model based Autonomous Agents," 2023. [En línea]. Disponible: <https://doi.org/10.48550/arXiv.2308.11432>.
- [7] Gobierno de los EE. UU., "Artificial Intelligence and Future of Teaching and Learning: Insights and Recommendations," U.S. Department of Education, Office of Educational Technology, 2023. [En línea]. Disponible: <https://www2.ed.gov/documents/ai-report/ai-report.pdf>.
- [8] University of Florida, "Artificial Intelligence," 2024. [En línea]. Disponible: <https://ai.ufl.edu>.
- [9] The New England Journal of Medicine, "Review article series: AI in Medicine," 2024. [En línea]. Disponible: <https://www.nejm.org/ai-in-medicine>.
- [10] R. Lam, A. Sanchez-Gonzalez, M. Willson, P. Wirnsberger, M. Fortunato, F. Alet, S. Ravuri, T. Ewalds, Z. Eaton-Rosen, W. Hu, A. Merose, S. Hoyer, G. Holland, O. Vinyals, J. Stott, A. Pritzel, S. Mohamed, y P. Battaglia, "Learning skillful medium-range global weather forecasting," *Science*, vol. 382, no. 6677, pp. 1416-1421, 2023. [En línea]. Disponible: <https://doi.org/10.1126/science.adi2336>.
- [11] Unión Europea, "Reglamento general de protección de datos," Diario Oficial de la Unión Europea, 2016. [En línea]. Disponible: <https://eur-lex.europa.eu/legal-content/ES/TXT/PDF/?uri=CELEX:32016R0679>.
- [12] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, R. Jozefowicz, Y. Jia, L. Kaiser, M. Kudlur, ... X. Zheng, "TensorFlow: Large-scale machine learning on heterogeneous systems," 2015. [En línea]. Disponible: <https://www.tensorflow.org>.
- [13] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Köpf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, y S. Chintala, "PyTorch: An Imperative Style, High-Performance Deep Learning Library," 2019. [En línea]. Disponible: <https://doi.org/10.48550/arXiv.1912.01703>.

- [14] M. Chen, J. Tworek, H. Jun, Q. Yuan, H. Ponde de Oliveira Pinto, J. Kaplan, H. Edwards, Y. Burda, N. Joseph, G. Brockman, A. Ray, R. Puri, G. Krueger, M. Petrov, H. Khlaaf, G. Sastry, P. Mishkin, B. Chan, ..., W. Zaremba, "Evaluating Large Language Models Trained on Code," 2021. [En línea]. Disponible: <https://doi.org/10.48550/arXiv.2107.03374>.
- [15] Nvidia, "High-Performance Computing and Artificial Intelligence," Nvidia Corporation, 2024. [En línea]. Disponible: <https://www.nvidia.com/en-us/high-performance-computing/hpc-and-ai/>.
- [16] Y. Bai, S. Kadavath, S. Kundu, A. Askell, J. Kernion, A. Jones, A. Chen, A. Goldie, A. Mirhoseini, C. McKinnon, C. Chen, C. Olsson, C. Olah, D. Hernandez, D. Drain, D. Ganguli, D. Li, E. Tran-Johnson, E. Perez, ... J. Kaplan, "Constitutional AI: Harmlessness from AI Feedback," 2022. [En línea]. Disponible: <https://doi.org/10.48550/arXiv.2307.16338>.
- [17] Y. Bengio, G. Hinton, A. Yao, D. Song, P. Abbeel, Y. N. Harari, Y.-Q. Zhang, L. Xue, S. Shalev-Shwartz, G. Hadfield, J. Clune, T. Maharaj, F. Hutter, A. G. Baydin, S. McIlraith, Q. Gao, A. Acharya, D. Krueger, A. Dragan, ..., S. Mindermann, "Managing AI Risks in an Era of Rapid Progress," 2023. [En línea]. Disponible: <https://doi.org/10.48550/arXiv.2310.17688>.
- [18] I. Asimov, "Strange playfellow," *Super Science Stories*, vol. 1, no. 4, pp. 67-77, 1940.
- [19] R. Ngo, L. Chan, y S. Mindermann, "The alignment problem from a deep learning perspective," 2023. [En línea]. Disponible: <https://doi.org/10.48550/arXiv.2209.00626>.
- [20] P. P. Ray, "ChatGPT: A comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope," *Internet of Things and Cyber-Physical Systems*, vol. 3, pp. 121-154, 2023. [En línea]. Disponible: <https://doi.org/10.1016/j.iotcps.2023.04.003>.
- [21] A. Turchin y D. Denkenberger, "Classification of Global Catastrophic Risks Connected with Artificial Intelligence," *AI and Society*, vol. 35, no. 1, pp. 147-163, 2020. [En línea]. Disponible: <https://doi.org/10.1007/s00146-018-0845-5>.
- [22] J. Achiam, S. Adler, S. Agarwal, L. Ahmad, I. Akkaya, F. L. Aleman, D. Almeida, J. Altenschmidt, S. Altman, S. Anadkat, R. Avila, I. Babuschkin, S. Balaji, V. Balcom, P. Baltescu, H. Bao, M. Bavarian, J. Belgum, I. Bello, J. Berdine, ... B. Zoph, "GPT-4 Technical Report," 2023. [En línea]. Disponible: <https://doi.org/10.48550/arXiv.2303.08774>.
- [23] Y. Bengio, "AI and Catastrophic Risk," *Journal of Democracy*, vol. 34, no. 4, pp. 111-121, 2023. [En línea]. Disponible: <https://doi.org/10.1353/jod.2023.a907692>.
- [24] Future of Life Institute, "The Asilomar AI Principles," 2017. [En línea]. Disponible: <https://futureoflife.org/open-letter/ai-principles>.

- [25] Université de Montréal, “Declaración de Montreal para un Desarrollo Responsable de la Inteligencia Artificial,” 2018. [En línea]. Disponible: https://declarationmontreal-iaresponsable.com/wp-content/uploads/2023/01/ES-UdeM_Decl-IA-Resp_LA-Declaration_v4.pdf.
- [26] OCDE, “Recommendation of the Council on Artificial Intelligence,” Organisation for Economic Co-operation and Development (OCDE), 2019. [En línea]. Disponible: [https://one.oecd.org/document/C/MIN\(2019\)3/FINAL/en/pdf](https://one.oecd.org/document/C/MIN(2019)3/FINAL/en/pdf).
- [27] UNESCO, “Recomendación sobre la ética de la inteligencia artificial,” Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura (UNESCO), 2021. [En línea]. Disponible: https://unesdoc.unesco.org/ark:/48223/pf0000381137_spa.
- [28] UNICEF, “Orientación de políticas sobre el uso de la inteligencia artificial en favor de la infancia,” Fondo de las Naciones Unidas para la Infancia (UNICEF), 2021. [En línea]. Disponible: https://www.unicef.org/globalinsight/media/2636/file/UNICEF-Global-Insight-policy-guidance-AI-children-2.0-2021_ES.pdf.
- [29] OCDE, “Estudios de la Organización para la Cooperación y el Desarrollo Económicos (OCDE) sobre el uso estratégico y responsable de la inteligencia artificial en el sector público de América Latina y el Caribe,” 2022. [En línea]. Disponible: <https://oecd-opsi.org/wp-content/uploads/2022/09/IA-ALC-Report.pdf>.
- [30] Gobierno de Argentina, “Plan Nacional de Inteligencia Artificial,” Presidencia de la Nación, Secretaría de Ciencia, Tecnología e Innovación Productiva, 2019. [En línea]. Disponible: <https://oecd-opsi.org/wp-content/uploads/2021/02/Argentina-National-AI-Strategy.pdf>.
- [31] Gobierno de Brasil, “Estrategia Brasileña de Inteligencia Artificial,” Ministerio de Ciencia, Tecnología e Innovación, 2021. [En línea]. Disponible: <https://www.gov.br/mcti/pt-br/acompanhe-o-mcti/transformacaodigital/inteligencia-artificial>.
- [32] Gobierno de Chile, “Política Nacional de Inteligencia Artificial,” Ministerio de Ciencia, Tecnología, Conocimiento e Innovación, 2021. [En línea]. Disponible: https://minciencia.gob.cl/uploads/filer_public/bc/38/bc389daf-4514-4306-867c-760ae7686e2c/documento_politica_ia_digital_.pdf.
- [33] Estado Peruano, “Estrategia Nacional de Inteligencia Artificial para el periodo 2021-2026,” Secretaría de Gobierno y Transformación Digital, 2021. [En línea]. Disponible: <https://cdn.www.gob.pe/uploads/document/file/1899077/Estrategia%20Nacional%20de%20Inteligencia%20Artificial.pdf>.
- [34] OpenAI, “ChatGPT. Large language model,” 2023. [En línea]. Disponible: <https://chat.openai.com/chat>.
- [35] Future of Life Institute, “Open Letter: Pause Giant AI Experiments,” 2023. [En línea]. Disponible: <https://futureoflife.org/open-letter/pause-giant-ai-experiments>.
- [36] M. Hanna, O. Liu, y A. Variengien, “How does GPT-2 compute greater-than?: Interpreting mathematical abilities in a pre-trained language model,” 2023. [En línea]. Disponible: <https://doi.org/10.48550/arXiv.2305.00586>.

- [37] ONU, "Creación del Consejo Consultivo que se enfocará en analizar y promover recomendaciones para la gobernanza internacional de la inteligencia artificial," Organización de las Naciones Unidas (ONU), 2023. [En línea]. Disponible: <https://news.un.org/es/story/2023/10/1525252>.
- [38] G. Sastry, L. Heim, H. Belfield, M. Anderljung, M. Brundage, J. Hazell, C. O'Keefe, G. K. Hadfield, R. Ngo, K. Pilz, G. Gor, E. Bluemke, S. Shoker, J. Egan, R. F. Trager, S. Avin, A. Weller, Y. Bengio, y D. Coyle, "Computing Power and the Governance of Artificial Intelligence," 2024. [En línea]. Disponible: <https://doi.org/10.48550/arXiv.2402.08797>.
- [39] J. Cameron y W. Wisner, "Terminator 2: Judgment Day," TriStar Pictures, 1991. [En línea]. Disponible: <https://www.sonypictures.com/>.
- [40] Gobierno de México, "Estrategia de Inteligencia Artificial de México 2018," Presidencia de la República, 2018. [En línea]. Disponible: <https://www.gob.mx/epn/articulos/estrategia-de-inteligencia-artificial-mx-2018>.
- [41] ISO, "ISO/IEC JTC 1/SC 42 Artificial Intelligence," International Organization for Standardization (ISO), 2018. [En línea]. Disponible: <https://www.iso.org/es/contents/data/committee/67/94/6794475.html?view=participation>.
- [42] Gobierno de México, "Declaración Conjunta de los miembros fundadores de la Alianza Global sobre la Inteligencia Artificial," Secretaría de Relaciones Exteriores, 2020. [En línea]. Disponible: <https://www.gob.mx/sre/prensa/declaracion-conjunta-de-los-miembros-fundadores-de-la-alianza-global-sobre-la-inteligencia-artificial>.
- [43] GPAI, "2023 Ministerial Declaration," Global Partnership on Artificial Intelligence (GPAI), 2023. [En línea]. Disponible: <https://gpai.ai/2023-GPAI-Ministerial-Declaration.pdf>.
- [44] CEPAL, "IA para el desarrollo sostenible de América Latina: Lanzamiento del primer Índice Latinoamericano de Inteligencia Artificial," Comisión Económica para América Latina y el Caribe (CEPAL), 2023. [En línea]. Disponible: <https://www.cepal.org/es/eventos/ia-desarrollo-sostenible-america-latina-lanzamiento-primero-indice-latinoamericano>.
- [45] CENIA, "Índice Latinoamericano de IA elaborado por el Centro Nacional de Inteligencia Artificial de Chile (CENIA). Ficha de México," 2023. [En línea]. Disponible: <https://indicelatam.cl/wp-content/uploads/2023/08/CAP-G-MEXICO.pdf>.
- [46] IFT, "Recomendaciones que emite el Consejo Consultivo del IFT para implementar en su ámbito de competencia la recomendación sobre la ética de la inteligencia artificial de la UNESCO," Instituto Federal de Telecomunicaciones (IFT), 2023. [En línea]. Disponible: https://www.ift.org.mx/sites/default/files/ag_ii-5_recomendacion_inteligencia_artificial_cci_acc.pdf.

- [47] Gobierno de México, "Iniciativa de Ley Federal de Ciberseguridad," Cámara de Diputados LXV Legislatura, Gaceta Parlamentaria, 2023. [En línea]. Disponible: https://www.diputados.gob.mx/LeyesBiblio/iniclave/65/CD-LXV-II-2P-292/02_iniciativa_292_25abr23.pdf.
- [48] Gobierno de México, "Iniciativa de Ley de Regulación Ética de la Inteligencia Artificial y la Robótica," Cámara de Diputados LXV Legislatura, Gaceta Parlamentaria, 2023b. [En línea]. Disponible: https://www.senado.gob.mx/65/gaceta_comision_permanente/documento/135000.
- [49] L. Ho, J. Barnhart, R. Trager, Y. Bengio, M. Brundage, A. Carnegie, R. Chowdhury, A. Dafoe, G. Hadfield, M. Levi, y D. Snidal, "International Institutions for Advanced AI," 2023. [En línea]. Disponible: <https://doi.org/10.48550/arXiv.2307.04699>.
- [50] UNESCO, "ChatGPT e Inteligencia Artificial en la educación superior: Guía de inicio rápido," Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura (UNESCO), 2023. [En línea]. Disponible: https://unesdoc.unesco.org/ark:/48223/pf0000385146_spa.locale=es.
- [51] Universidad Carlos III de Madrid, "Recomendaciones para la docencia con inteligencias artificiales generativas," Vicerrectorado de Infraestructuras, 2023. [En línea]. Disponible: <https://e-archivo.uc3m.es/entities/publication/e560161f-44a3-43f5-9a4e-5175a052c2ec>.
- [52] UNAM, "Recomendaciones para el uso de la inteligencia artificial generativa en la docencia," Universidad Nacional Autónoma de México, 2023b. [En línea]. Disponible: https://iagenedu.unam.mx/docs/recomendaciones_uso_iagen_docencia_unam_2023.pdf.
- [53] Z. Swiecki, H. Khosravi, G. Chen, R. Martinez-Maldonado, J. M. Lodge, S. Milligan, N. Selwyn, y D. Gašević, "Assessment in the age of artificial intelligence," *Computers and Education: Artificial Intelligence*, vol. 3, p. 100075, 2022. [En línea]. Disponible: <https://doi.org/10.1016/j.caeai.2022.100075>.
- [54] UNAM, "Ofrece el LAMOD nuevas posibilidades para la investigación," Gaceta de la Universidad Nacional Autónoma de México, 2023a. [En línea]. Disponible: <https://www.gaceta.unam.mx/la-unam-con-gran-potencial-en-proyectos-de-frontera>.
- [55] UNAM, "Lineamientos generales y políticas sobre almacenamiento e información compartida entre los sistemas existentes," Universidad Nacional Autónoma de México, 2021b. [En línea]. Disponible: https://www.red-tic.unam.mx/recursos/2021/2021_Lineamiento_RedResponsablesTIC_01.pdf.
- [56] Gobierno de Reino Unido, "Cumbre de Seguridad de la IA 2023: Declaración Bletchley," 2023. [En línea]. Disponible: <https://www.gov.uk/government/publications/ai-safety-summit-2023-the-bletchley-declaration>.
- [57] UNAM, "Alianza para promover el Desarrollo de Capacidades Digitales en México," Universidad Nacional Autónoma de México, 2021a. [En línea]. Disponible: <https://alianza.unam.mx/>.

- [58] M. J. Santos Corral, R. de Gortari Rabiela, y M. Lopátegui, "Construir vinculación desde la Inteligencia Artificial: Análisis de una alianza interinstitucional," *Entreciencias: diálogos en la sociedad del conocimiento*, vol. 11, no. 25, p. e2584175, 2023. [En línea]. Disponible: <https://doi.org/10.22201/enesl.20078064e.2023.25.84175>.
- [59] UNAM, "Acuerdo por el que se establecen los Lineamientos para la Integración, Conformación y Registro de los Comités de Ética en la Universidad Nacional Autónoma de México," *Gaceta UNAM*, 2019a. [En línea]. Disponible: <https://www.gaceta.unam.mx/wp-content/uploads/2019/10/190829-Convocatorias.pdf>.
- [60] Gobierno de México, Secretaría de Salud, Comisión Nacional de Bioética, "Bioética e inteligencia artificial," 2023c. [En línea]. Disponible: https://www.gob.mx/cms/uploads/attachment/file/846953/gaceta48_final_completa__3_.pdf.
- [61] UNAM, "Expertos debaten en la UNAM retos y alcances de la inteligencia artificial," *Boletín UNAM-DGCS-746*, Universidad Nacional Autónoma de México, 2019b. [En línea]. Disponible: https://www.dgcs.unam.mx/boletin/bdboletin/2019_746.html.
- [62] A3D3 Institute, "About A3D3," Accelerated AI Algorithms for Data-Driven Discovery (A3D3) Institute, 2023. [En línea]. Disponible: <https://a3d3.ai/about/>.
- [63] MILA, "Quebec AI Institute (Montreal Institute for Learning Algorithms)," 2024. [En línea]. Disponible: <https://mila.quebec/en/mila/>.