

V REVISTA DE EL COLEGIO DE SAN LUIS

ISSN: 2007-8846 • <http://revista.colsan.edu.mx>
Enero a diciembre de 2025 • Año 15 • Número 26 • pp. 1-26
e1712 • <https://doi.org/10.21696/rcsl152620251712>
Sección Artículos

Antroposíntesis social: interacciones y *nomos* Humanos∞Inteligencia Artificial (OpenAI/ChatGPT/Microsoft)

*Social Anthroposynthesis: Interactions and Nomos Humans∞Artificial Intelligence
(OpenAI/ChatGPT/Microsoft)*

Liliana Martínez Pérez

Facultad Latinoamericana de Ciencias Sociales Sede México
marper@flacso.edu.mx  <https://orcid.org/0000-0001-8990-197X>

Santiago Carassale

Facultad Latinoamericana de Ciencias Sociales Sede México
sandres@flacso.edu.mx  <https://orcid.org/0000-0001-9743-5892>

Javier Contreras Alcántara

El Colegio de San Luis, Programa de Estudios Políticos e Internacionales
javier.contreras@colsan.edu.mx  <https://orcid.org/0000-0001-5090-5500>

Recibido: 28 de febrero de 2025. **Aceptado:** 31 de julio de 2025.

Esta obra se encuentra publicada bajo la Licencia Creative Commons 4.0 Internacional
Atribución -Reconocimiento-NoComercial-SinDerivados
<https://creativecommons.org/licenses/by-nc-nd/4.0/>



EL
COLEGIO
DE
SAN LUIS

RESUMEN

Con el objetivo de comprender cómo la inteligencia artificial general dramatiza y configura un “juego profundo” a partir de una “brecha” doble, socioantropológica y sociotecnológica, se analizaron, desde el método pragmático cultural, cuatro momentos cardinales del social/stage drama de la invención~innovación del ChatGPT. Se encontró que la inteligencia artificial general produce una interacción humana con una alteridad digital (“espectro/spectrum”) y una nueva disputa por el lucro, el poder y el futuro humano. Una de las limitaciones del estudio estriba en que las transformaciones sociales develadas demandan conceptualizaciones socioantropológicas y políticas novedosas y críticas. El artículo ofrece una perspectiva teórico-analítica antropológica y dramaturgista. Se concluye que la nueva interacción humano~inteligencia artificial general desencadena un cambio radical de ser humano que re-escala la disputa política y el quehacer científico social.

Palabras clave: tecnologías digitales, otro digital significativo, juego profundo, drama social, ciencias sociales.

ABSTRACT

With the aim of understanding how artificial general intelligence dramatizes and configures a “deep game” based on a double “breach”, socio-anthropological and socio-technological, four cardinal moments of the social/stage drama of ChatGPT’s invention-innovation were analyzed using a cultural pragmatic method. It was found that artificial general intelligence produces human interaction with a digital otherness (“spectro/spectrum”) and a new dispute over profit, power, and the human future. One of the limitations of the study is that the social transformations revealed demand novel and critical socio-anthropological and political conceptualizations. The article offers an anthroposynthetic and dramaturgical theoretical-analytical perspective. It concludes that the new human~artificial general intelligence interaction triggers a radical change in being human that re-scales political disputes and social scientific work.

Keywords: Digital Technologies, Digital Significant Other, Deep Play, Social Drama, Social Sciences.

Introducción¹

Los “avances tecnológicos” digitales actuales son presentados a un ritmo acelerado, anticipatorio y reiterativo, a través de innumerables notas en redes sociodigitales, sitios de internet, productos audiovisuales y escenificaciones públicas: un “carrusel de feria” mediático desde donde se trama una comunicación social en clave de espectáculo pendular que, en un extremo, exalta las expectativas maquinísticas prometeicas y, en el otro, atemoriza con los graves peligros potenciales de su aplicación para la humanidad y sus presentes futuros. Esta exhibición sintética, artificiosa (*open*

1 Este artículo es resultado del trabajo en el marco del proyecto “Antroposíntesis social: problemáticas clave en la relación humano~tecnología en las sociedades contemporáneas”, FLACSO México, 2024-2026.

*play*²), distante de la exposición realista y crítica de las rupturas o revoluciones tecnológicas en las que se basan las invenciones e innovaciones tecnodigitales, encubre o vela, en nuestra opinión, el "juego profundo" (*deep play*³) —más similar a una "ruleta de casino"— de las interacciones y del *nomos*⁴ antroposintéticos que implican el desarrollo de sus procedimientos hiperespecializados y extractivos, sus complejas y costosas infraestructuras y sus densas asimetrías políticas, sociales y científicas.

El objetivo de este artículo se centra en comprender, desde la perspectiva de las ciencias sociales (sociología, antropología y ciencia política), cómo las tecnologías digitales, en particular la inteligencia artificial generativa (IAGn), dramatizan y configuran un "juego profundo", sustentado en una brecha (*breach*) cultural doble: sociotecnológica y socioantropológica.

La primera trastoca el *nomos* moderno —basado en la apropiación y distribución de recursos naturales para la producción y el consumo de bienes y servicios— en un *nomos* fundado en la apropiación masiva y extractiva de datos, la heteromatización y la algoritmización supervisada y autónoma de la producción, el consumo, las reglamentaciones y el orden estabilizado en un sistema social.

La segunda modula y modela artificialmente el "saber hacer" cinegético, abductivo, práctico (Ginzburg, 1999, pp. 138-175; Eco y Sebeok, 1989, pp. 19-30; Chaiklin y Lave, 1996), y la creatividad de los seres humanos, tanto en su heurística: modo, forma, patrones de descubrimiento, como en su contenido: toma de decisiones, criterios de asignación, diseño de patrones, oportunidades de acceso, horizontes de expectativas, espacios de experiencias e interacciones sociales.

2 Utilizamos *open play* para distinguir los procesos aparentes y las apuestas utilitarias en juego en las dinámicas sociales, en contraste con el *deep play* (Geertz, 2003a, pp. 355-356). Asimismo, pretendemos destacar lo manifiesto, lo desplegado por el drama social y desde el *stage drama* a raíz del desarrollo y el uso humano (antroposocial) de las tecnologías de inteligencia artificial, entendido como "la forma dominante de un "espejo" [matricial] estético-cultural en el cual [las formaciones socioeconómicas] alcanzan un cierto grado de auto-reflexividad" (Turner, 1985, p. 291).

3 Por *deep play* entendemos, como lo refiere Geertz (2003a, pp. 355-356) siguiendo a Bentham, un intercambio simbólico en forma de apuesta con probabilidades parejas, de todo o nada, extremo, solo aceptable para quienes están inmersos en la cultura de origen (estatus, poder, orgullo, etcétera), el cual pone en juego el orden social. Un flujo de interacciones implícito, a decir de Turner (1985), que habilita el juego entre el lucro y la regulación en el caso de la relación humanos∞inteligencia artificial.

4 Entendemos por *nomos* la constitución de un espacio abierto, no domesticado ni ordenado, susceptible de ser apropiado, distribuido y producido. En este sentido, recuperamos a Deleuze y Guattari (2015, pp. 384-385, 426), quienes señalan que la raíz griega *nem* de la palabra *nomos* refiere en principio la noción de distribución espacial, pastoril, de los animales (nomadología), así como a Schmitt (2003, pp. 326-328), quien utiliza *nomos* (*nemein*) para aludir al proceso de apropiación para la distribución y la producción de la tierra. Esta doble observación nos permite explorar el "nuevo juego" auspiciado por la brecha tecnológica (*breakthrough*) de la IAGn, la liminalidad (Turner, 1985, p. 295; Deleuze y Guattari, 2015, p. 361), en las interacciones y dramatizaciones sociales y políticas.

Esta doble fractura instaaura un drama social que pugna entre la legitimación (ficcionalización, futurización) y la antropomorfización de las tecnologías artificialmente inteligentes “abierta-manifiesta” (*open*) versus la “virtual-implícita” (*deep*) corporativización, opacidad y “fisión” del ordenamiento, las acciones, las apuestas sociales y políticas, las densas asimetrías y la periferización de lo humano vinculadas al despliegue de estas tecnologías.

La dramatización, futurización y futuridad en torno a la inteligencia artificial (IA) oscilan entre la utopía basada en la potencialidad de la superinteligencia artificial tanto para satisfacer las más altas demandas de democratización, bienestar y gozo de los seres humanos como para resolver los problemas más graves y complejos que aquejan a la especie humana y al planeta; y la distopía del advenimiento de la Singularidad —la aparición de la inteligencia artificial general (IAGI), su total consciencia, autonomía y sujeción (extinción) de los seres humanos (*Matrix*, 1999-2021; *Her*, 2013; *Ex Machina*, 2014)—. Mientras, la acción del *nomos* antropológico se debate en torno al dominio y la apropiación del código y la innovación algorítmica, el lucro corporativo o el acceso abierto y la seguridad o el descontrol.

A continuación, abordaremos sucintamente algunos cambios clave en la relación humanos∞-máquina/IA y las nuevas conceptualizaciones sociológicas requeridas para la comprensión del “juego profundo” antropológico. Luego, a partir del método pragmático cultural (Geertz, 2003b; Turner, 1985; Schechner, 1976),⁵ examinaremos cuatro momentos cardinales asociados a la invención~innovación del Chat Generative Pre-Trained Transformer (en adelante, ChatGPT), avance científico-tecnológico desarrollado por la compañía OpenAI —fundada en 2015 por emprendedores y desarrolladores como Samuel Altman, Gregory Brockman, Ilya Sutskever, Elon Musk, entre otros, y asociada en 2019 a la corporación tecnológica multinacional Microsoft, creada en 1975 por William “Bill” Gates y Paul Allen—, cuyo despliegue permite develar varias de las aristas y tensiones, manifiestas y latentes, de las interacciones y el *nomos* antropológico. Y, al final, desde los resultados de esta investigación sobre el juego antropológico humanos∞IA, sugeriremos algunos puntos clave acerca de las formas de reflexionar de las ciencias sociales.

5 Entendemos por método pragmático cultural una perspectiva conceptual y analítica orientada al examen e interpretación de los flujos posibles entre lo manifiesto, lo latente y lo implícito en los procesos sociales, culturales y políticos, así como sus tensiones y disputas simbólicas y prácticas. Este modelo analítico refiere los acontecimientos, el acontecer, como interacciones situadas —no como códigos simbólicos binarios por develar—, como modulaciones, configuraciones, “juegos” (*play*) en el doble sentido de apuesta y representación.

Relación y “juego profundo” humanos∞máquina/ia

La historia de la relación humanos∞máquina ha estado tramada tanto por la acción como por la recepción y la estética humana. En las sociedades no industriales, el vínculo ha sido básicamente herramental; mientras, en las sociedades industriales preelectrónicas, la relación se ha centrado en la automatización de procesos mecánicos y en la mimesis mutua entre el ser humano y la máquina (Turner, 1985, p. 291; Rabinbach, 2018).

Las innovaciones más significativas en las sociedades industriales fueron la máquina de vapor o motor de combustión externa (1769) e interna (1860, 1876) y el dominio de la electricidad (1870), centradas en la explotación y el uso de un recurso o energía natural y cuyo máximo simbólico sería el “motor humano”; y la máquina o telar de Jacquard (1801), síntesis de un conjunto de experimentaciones (Kang, 2011) y avances tecnológicos previos (1725-1740), cuya innovación radicaba en la capacidad de “codificar” el diseño de patrones en textiles en un bucle cerrado de tarjetas perforadas (primera y segunda revolución industrial).

Durante siglo y medio, la codificación del telar habilitaría, de modo larvado (Cardon *et al.*, 2018), tanto la alegoría de la máquina autómatas calculista/inteligente (Ch. Babbage, 1837, A. Lovelace, 1843) como la producción en serie de las máquinas codificadoras y tabuladoras de H. Hollerith (1889, 1896) por la International Business Machines Corporation (IBM) (Pugh, 1996). Y, desde mediados del siglo XX, se convertiría en pieza clave (*building block*) de la tercera y cuarta revolución científico-tecnológica: la computadora personal, la interconexión, la datificación y la algoritmización o IA simbólica (lógico-deductiva).

La diferencia significativa en la relación humanos∞máquina/IA desplegada por la innovación de la inteligencia artificial conectiva, generativa y de autoaprendizaje (IAGn), desarrollada desde fines del siglo XX, radica en la exteriorización u objetivación de los procesos intelectivos y creativos del ser humano, que no solo lo sustituyen (Thomson Reuters, 2023), sino que parecen periferizarlo, en el sentido de convertirlo en un dispositivo o “periférico” del que la máquina extrae información y valor: en palabras de Ekbja y Nardi, un proceso de heteromatización más que de automatización, es decir, un cambio tecnológico que “más que descartar (*disallow*) la intervención humana en casi todos los puntos del sistema [...], empuja las tareas críticas hacia los usuarios finales como mediadores indispensables” (Ekbja y Nardi, 2014, p. 2). Para develar este “juego profundo” (Geertz, 2003a, pp. 339-372) antroposintético instaurado por esta tecnología IAGn en los humanos y en la sociedad actual, se requiere, asimismo, una conceptualización sociológicamente significativa.

Definimos la IAGn como un “espectroum” de invenciones~innovaciones e interacciones humanos~IA, lo que supone un ensamble sociotecnológico, más de contenido que de forma (Ekbja y Nardi, 2018). Esta concepción nos permite develar la doble operación de la IAGn: por una parte, el despliegue de una gama y gradiente científico tecnológico procedimental artificial (“espectro”); por la otra, la agencia de una máquina humanizada (“*spectrum*”) que habilita un umbral y fase liminoide (*threshold*) (Turner, 1985, p. 295) en el proceso de relación entre humanos y máquinas (*in-line*).

El “espectro” tecnológico de la IAGn supone el diseño deliberado (sintético) de programaciones que hacen posible desde la algoritmización de los procesos digitales extractivos (minería) de datos humanos, la elaboración de datos sintéticos, la identificación de patrones de organización desconocidos de los datos, hasta el aprendizaje de máquina (algorítmico) supervisado y autónomo basado en invenciones matemático-estadísticas (convolución, método Montecarlo, entre otras) y de la ciencia de datos (redes neuronales artificiales), y su implementación en procedimientos prácticos (conducción y operaciones autómatas), discernimientos analíticos y anticipatorios, así como en modelos de lenguaje masivo (MLM) para comunicarse en tiempo real y contextualizado (“natural”) con seres humanos.

Asimismo, esta gama de procedimientos informáticos constituye la tecnología IAGn en un “*spectrum*”, una agencia (Esposito, 2017) antropológica: un ente o entidad artificial significativa (Heimans y Timms, 2024), corpórea (robot) o digital (*chatbot*), que interactúa, sintetiza modos de actuar y, de esa manera, diseña la condición humana misma; es decir, modela y abstrae las singularidades y cualidades de ser humano, así como de su interacción social y natural (evolutiva) (Carassale *et al.*, 2023).

Por otra parte, el desarrollo de esta tecnología implica un *nomos* u ordenamiento socioeconómico similar al de una “máquina de guerra” o al del juego de Go (Deleuze y Guattari, 2015, pp. 360-361). En tanto la materialidad (semiconductores y microprocesadores electrónicos) y el recurso clave (datos humanos) necesarios para su implementación se caracterizan por una funcionalidad de propósito general (*general purpose technology*) —dada su cualidad de carecer de propiedades intrínsecas, como las “piedras” del Go—, la apropiación/distribución/producción (Schmitt, 2003, pp. 324-335) depende de actores y acciones que exploran y aprovechan estas fuentes y bienes de modo subrepticio, y ocupan el espacio de manera radicalmente situada, como los “movimientos” principales desplegados en el Go: bordear, rodear, romper.

Este nuevo *nomos*, además, es subrepticio en su lógica de acumulación de capital que, según Ekbja y Nardi (2017, p. 1), se sustenta “en la extracción de valor económico basada en el empleo de bajo costo o ‘gratuito’ generado en las redes de trabajo mediadas por computadoras”, cuyo impacto en las relaciones

laborales radica, asimismo, en un vínculo en el que “unos humanos, típicamente organizados en una compañía, se benefician del trabajo de otros humanos a partir de ingresos muy bajos e, incluso, solo de grandes premios afectivos (no remunerativos)” (Ekbja y Nardi, 2014, p. 2).⁶

Sin embargo, la gigantesca magnitud de operaciones y datos requerida por la tecnología digital ha impuesto a esa “máquina de guerra”: el desarrollo de estructuras de densa conectividad y alta velocidad (internet y telefonía móvil de banda ancha, nube digital, centros de almacenamiento de megadatos, procesadores cuánticos, etcétera); minería de “materiales raros”, escasos o de difícil explotación; y aprendizajes y procesamiento hiperespecializados. Estas exigencias han forzado, desde temprano, la conglomeración o corporativismo entre los inventores e innovadores y los actores socioeconómicos tradicionales, como las compañías manufactureras computacionales y de telecomunicaciones, el complejo militar industrial, los grandes laboratorios biotecnológicos y la industria aeroespacial.

En este entramado, el *nomos* de tecnologías digitales como la IAGn se sostiene, entonces, en un abanico de estrategias que van desde las *startups* (apuestas puntuales, *capital venture*) hasta las vastas corporaciones público-privadas, pautadas en una acelerada carrera tanto por la apropiación como por la tensa disputa entre la “atomización” de la invención (fisión, fuerza centrífuga) y la presión a favor de la regulación de la innovación (fusión, fuerza centrípeta).

A continuación, enmarcados por esta breve historización y delimitación conceptual primaria, exploraremos las interacciones y el *nomos* antroposintético desplegados por la invención~innovación conocida como ChatGPT. Los cuatro momentos que examinaremos en el siguiente apartado muestran las escenificaciones, las distinciones y las apuestas relacionadas con el “espectroum” y el *nomos* de la IAGn, y la ruptura profunda del “saber hacer” colectivo —experiencial y ontológica, más que normativa, tradicional o cognitiva (Binder, 2021)— y de los procesos de toma de decisiones en las sociedades algorítmicas contemporáneas (Roberge y Seyfert, 2016; Crawford, 2021).

Inteligencia artificial: interacción espectral y *nomos* extractivo/corporativo

Había una vez... un ChatGPT

El 30 de noviembre de 2022, OpenAI publicó en su página oficial de internet un mensaje muy cuidado y detallado en el que anunciaba el nacimiento de un chat digital capaz de “conversar”, no solo informar, de manera “natural”, contextual, con un ser humano, cuyo nombre corto es

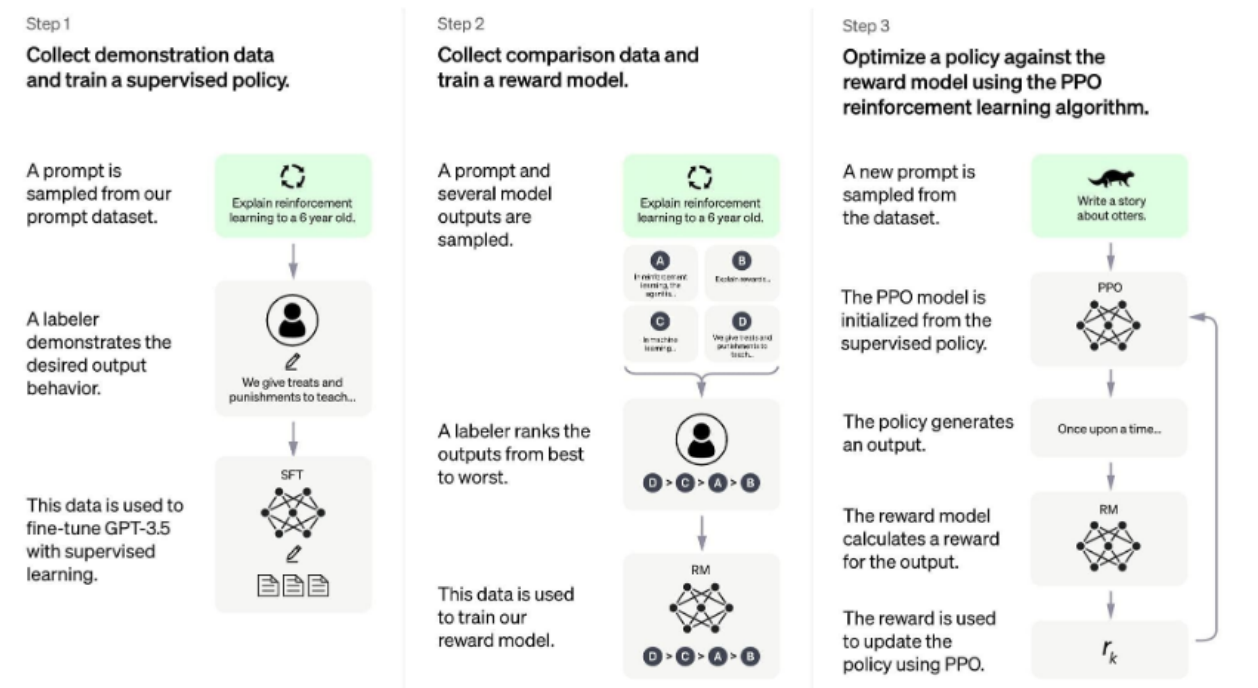
⁶ Ejemplo ilustrativo a gran escala de este *nomos* heteromático es GitHub (<https://github.com/>), “plataforma de desarrollo colaborativo” para la creación de códigos fuente de programas de computación. Fue fundada por cuatro programadores en 2007, está en operación desde 2010 y fue comprada por Microsoft en 2018.

ChatGPT, una doble etiqueta que refiere su similitud con los *chatbots* (robots digitales para preguntas y respuestas) tradicionales y su novedosa implementación de un procedimiento algorítmico generativo, autoentrenado y transformativo, artificialmente inteligente, en un MLM (OpenAI, 2022b).

El mensaje incitaba a los lectores (sin la estridencia típica de las notas sobre las múltiples aplicaciones de la IAGn) al uso inmediato del chat, de manera gratuita, en un sitio exclusivo (www.chatgpt.com), y a la comunicación de retorno a la compañía de sus opiniones, para conocer las fortalezas y debilidades de ChatGPT. A la vez, publicaba información, muy probablemente para expertos y clientes conocedores de la tecnología y el trayecto previo de la compañía en torno a los programas de conversación con IAGn, acerca de cuatro asuntos relativos a los alcances y límites de la innovación: (i) ejemplos de uso (técnicos, prácticos, resolución de problemas e interacciones), (ii) métodos de entrenamiento algorítmicos implementados (SFT: Supervised Fine-Tuning, RMT: Reward Model Training, PPO: Proximal Policy Optimization [OpenAI, 2017]), (iii) fallas comunicacionales (respuestas plausibles pero incorrectas o sin sentido, entradas ampulosas o redundancia innecesaria al preguntar, verbosidad y reiteración, “adivinar” la intención del usuario ante consultas ambiguas, ofrecer respuesta ante instrucciones perjudiciales y comentarios sesgados) y (iv) algunas demostraciones del grado de control y seguridad alcanzado sobre las respuestas dañinas o falsas del chat (anacronismo/ficción histórica, violencia, acoso).

La figura diagramada incluida en la nota en la que se representaban los pasos algorítmicos (véase el **Diagrama 1**) había sido publicada a fines de enero de ese mismo año en el sitio oficial de OpenAI, con pequeñas diferencias (*prompts* o instrucciones formalmente distintas), para explicar el desarrollo del algoritmo llamado InstructGPT —mucho más eficaz para evitar las serias problemáticas de seguridad, toxicidad y “alucinaciones” producidas en las “conversaciones” con la versión de ChatGPT desarrollada hasta ese momento—, que había sido ejecutado en la versión ahora presentada: ChatGPT-3.5 (OpenAI, 2022a).

Diagrama 1. Modelo de entrenamiento por pasos del algoritmo de ChatGPT 3.5.



Fuente: OpenAI (2022b).

Más allá de la síntesis técnica del logro procesual alcanzado, en esta figuración destacan, de manera implícita o encriptada, dos cuestiones de relevancia socioantropológica: el probable uso de un modelo de aprendizaje algorítmico basado en la psicología cognitiva conductista (el *prompt* inicial del paso 1 y 2 solicita la elaboración de una explicación sobre un tema para un sujeto de seis años de edad) y el repliegue del humano en el proceso de generación autónoma de respuestas del algoritmo (en el paso 3 desaparece el ícono que representa al humano en los dos pasos anteriores como entrenador y evaluador de la respuesta algorítmica).

Asimismo, es importante señalar que el *post* incluía un enlace relativo al procedimiento PPO considerado el mayor avance informático del MLM generativo (paso 3 del **Diagrama 1**), que había sido presentado y explicado en la página oficial de OpenAI cinco años antes, el 20 de julio de 2017, ilustrado en ese momento con otras de las aplicaciones principales de la IAGn: la robótica y los videojuegos (OpenAI, 2017).

La amplia y documentada información acerca de los procedimientos habilitaba la invitación final a los potenciales usuarios conocedores y expertos, mayores de 18 años, a participar, también sin costo, en el Concurso de Comentarios de ChatGPT (OpenAI, 2022c), para tener la oportunidad de ganar hasta 500 dólares en créditos API (Application Programming Interface), y cuyo objetivo fundamental, en palabras de OpenAI, era:

Animamos a los usuarios a que nos envíen sus comentarios sobre los resultados problemáticos del modelo [...], así como sobre los falsos positivos/negativos del filtro de contenido externo [...]. Nos interesan especialmente los comentarios sobre resultados perjudiciales que podrían producirse en el mundo real, en condiciones no adversariales, así como los comentarios que nos ayuden a descubrir y comprender nuevos riesgos y posibles mitigaciones (OpenAI, 2022b).

El anuncio era el corolario de una serie de notas previas publicadas en el sitio oficial de la compañía sobre el entrenamiento de MLM generativos, aunque solo hasta esta nota se incluyó la "preposición" chat al procedimiento GPT: el primero (11 de junio de 2018) no refería nombre ni numeración (Radford, 2018); en el segundo (14 de febrero de 2019) se lo nombraba como GPT-2, y se indicaba que era el sucesor del publicitado ocho meses antes, al que ahora etiquetaban como GPT, sin notación (Radford *et al.*, 2019), y en el tercero (28 de mayo de 2020) se le refería como GPT-3 (Brown *et al.*, 2020).

Los cuatro *posts* concluían con una sección, frecuente en las publicaciones de OpenAI de este tipo, dedicada a notas y bibliografía especializada acerca del tema abordado, y la mención de las personas autoras, revisoras, asesoras y colaboradoras del experimento, procedimiento y/o novedad científico-tecnológica presentada, que acentuaba el tono serio, no publicitario o festinado, de la comunicación, así como el reconocimiento explícito de las personas titulares de la invención y/o innovación.

Los autores, revisores y colaboradores referidos en los cuatro *posts* variaron principalmente en número, que alcanzó un total de 120 personas. El primero (2018) solo tenía un autor, Alec Radford, considerado en varias fuentes como el coinventor del procedimiento junto a Luke Metz. Radford firmaría el segundo *post* (2019) como primer coautor, junto a seis más: Jeffrey Wu, Dario Amodei, Daniella Amodei, Jack Clark, Miles Brundage e Ilya Sutskever. Asimismo, en esa nota se reconocía que los coautores y cuatro colaboradores más realizaron una actualización de la información, y se agradecía a once dictaminadores o revisores expertos por sus comentarios a los borradores del texto, entre quienes destacaban Greg Brockman y John Schulman.

El tercer *post* (2020), un escueto mensaje en el que solo se informaba del avance alcanzado en la versión 3.0, incluyó dos listas de personas referidas como autores (31 en total), pero sin

detallar los aportes de estas. En estos tres *posts*, además de Radford, se repitieron, al menos en dos ocasiones, 12 individuos más que pudieran considerarse clave en el proceso de invención e innovación del ChatGPT.⁷

En la cuarta nota de presentación del ChatGPT-3.5 (2022), la autoría se adjudicaba a la compañía OpenAI y, después, se reconocía, por nombres y apellidos, a 87 colaboradores: John Schulman, primero; Luke Metz, noveno, y Greg Brockman, quincuagésimo tercero (OpenAI, 2022b).

Tanto la formación académica como la trayectoria profesional de Alec Radford, coinventor y desarrollador del procedimiento GPT, y de John Schulman, líder del extenso equipo de colaboradores relacionados con el ChatGPT-3.5, ofrecen valiosas pistas sobre la novedosa relación de una compañía como OpenAI y las invenciones de las últimas décadas en las ingenierías y las ciencias fisicomatemáticas, neurológicas y de la computación.

Alec Radford era aún estudiante de pregrado en el Franklin W. Olin College of Engineering (Massachusetts, 2011-2016) (Pressman, 2023), a la vez que cofundador, junto a Slater Victoroff, Madison May y Diana Yuan, de la *startup* llamada Indico Data Solutions (Massachusetts, 2014-2016) (F6S, 2024), cuando expuso, con otro compañero de estudios de pregrado, Luke Metz, supervisados por Soumith Chintala, Investigadora de Facebook AI, el primer gran avance acerca de los MLM en la International Conference on Learning Representations (ICLR), en el verano de 2016 (ICRL, 2024). Este avance se titula "Unsupervised Representation Learning With Deep Convolutional Generative Adversarial Networks" (Radford *et al.*, 2016). Estos hechos se convirtieron en sus credenciales profesionales clave para ser reclutado de inmediato por OpenAI (AMainer, 2024).

John Schulman, por su parte, hizo la Licenciatura en Física en el Instituto Tecnológico de California (CALTECH) y concluyó el Doctorado en Ciencias de la Computación en la Universidad de California en Berkeley, en 2016 (luego de iniciar estudios doctorales en Neurociencia en esta misma universidad) (Schulman, 2024), con la tesis *Optimizing Expectations: From Deep Reinforcement Learning To Stochastic Computation Graphs* (Schulman, 2016), asesorado por un destacado investigador de aprendizaje robótico. Por esta tesis tuvo la oportunidad de realizar un verano en la compañía DeepMind, en 2015. Schulman sería el primer autor del texto sobre el PPO publicado en el sitio oficial de OpenAI en 2017 y sería reconocido por el MIT Technology Review entre los innovadores menores de 35 años más destacados en 2018 (Arguedas, 2018).

Por último, como mencionamos antes, solo la cuarta presentación incluyó la invitación explícita a la utilización del ChatGPT por parte del público general o de expertos —las tres comunicaciones

7 Autores: Amanda Askell, Dario Amodei. Revisores y autores: Jeffrey Wu, Ilya Sutskever. Autores y colaboradores: Benjamin Chess, Christopher Hesse, Clemens Winter, Eric Sigler, Nick Ryder, Sandhini Agarwal. Revisores y colaboradores: Greg Brockman, John Schulman.

anteriores se limitaron a la exposición de los tipos de entrenamientos, bases de datos, ejemplos de respuestas, alcances y limitaciones del MLM generativo—.

El llamado a la doble audiencia potencial de la noticia, curiosos e iniciados, abría la compuerta del doble flujo del “espectroum” (“espectro/*spectrum*”) propio de una tecnología antropológica, así como los modos de generación de valor a través de la heteromatización, implícita y explícita, implementada sistemáticamente en el despliegue de estas tecnologías. Por una parte, los aficionados pondrían a prueba el *spectrum* mientras ampliarían de modo gratuito su acervo lingüístico y semántico; por la otra, los wikilaboradores y expertos examinarían el “espectro”, su codificación, procedimientos, potencialidades y torpezas, con la mira puesta en el premio monetario para su implementación en nuevas funcionalidades y aplicaciones.

El drama social antropológico: promesa y peligro Humanos∞IA

Seis días después de la convocatoria pública a la interacción con la versión 3.5 del ChatGPT desde la página oficial de OpenAI (5 de diciembre de 2022), Sam Altman redactó un mensaje en su cuenta de Twitter (ahora X) en el que comunicaba que para ese momento ya había más de un millón de usuarios del chat (Altman, 2022). Dos meses más tarde, el monto se estimaba en unos 100 millones de usuarios. Y, dos años después, con tres nuevas versiones algorítmicas (4.0: 14 de marzo de 2023 [OpenAI, 2023; OpenAI, 2024a]; 4o: 13 de mayo de 2024 [OpenAI, 2024c], y o1: 12 de septiembre de 2024 [OpenAI, 2024d]), acceso *premium* de paga y un sinnúmero de *apps* basadas en la tecnología GPT —especializadas en escritura, productividad, investigación y análisis, educación, estilos de vida y programación— disponibles en la plataforma ChatGPT.com, se estimaba que unos 100 millones de personas la utilizaban a la semana.

La eclosión en el ámbito de los diseñadores y desarrolladores de modelos conversacionales similares y de la IAGn ha sido sustantiva: la implementación de *chatbots* inteligentes por grandes compañías digitales es elocuente (Bing y Copilot-Microsoft, Gemini-Google); también, aunque de modo más gradual, entre medianas y pequeñas empresas. En adición, junto a esta vasta expansión de la innovación en IAGn, se incrementaron de forma exponencial las narrativas agonísticas en torno a las aplicaciones actuales y futuras de esta tecnología: por una parte, se concibe como la herramienta tecnológica (“espectro”) más eficaz para “salvar” a la humanidad de innumerables riesgos y graves problemáticas que afronta en lo político, lo social y lo medioambiental; por la otra, como un “*spectrum*” capaz no solo de expropiar el conocimiento social acumulado por las sociedades humanas, sino incluso de desbancar a la evolución como mecanismo primordial del modelado de la especie humana con el advenimiento de la llamada IAGI, una inteligencia artificial sustentada en un proceso algorítmico autónomo del que los desarrolladores se declaran, por momentos, incapaces

de explicar y controlar y cuya culminación podría ser la “autoconsciencia” tecnológica, también llamada Singularidad (Kurzweil, 2016; Chalmers, 2016).

El impacto atemorizante y el potencial horizonte distópico de la llegada de la IAGI habían sido escenificados y performados, primero, muy débilmente, a fines del siglo pasado, cuando la enorme computadora de IBM llamada Deep Blue jugó contra el mejor ajedrecista del mundo, Garry Kasparov, y perdió o empató la mayoría de las partidas del primer encuentro (1996) y ganó o empató cinco de seis juegos en el segundo (1997). Y, dos décadas después (2016), mucho más intensificado y cargado de perplejidad, ante la derrota pública y contundente del humano por el algoritmo llamado AlphaGo (desarrollado por la compañía DeepMind, fundada y administrada por Demis Hassabis en 2010 y adquirida en 2014 por Alphabet Inc., empresa matriz de Google), capaz de jugar un juego de estrategia completamente abierto, intuitivo y abstracto.⁸

El algoritmo AlphaGo, entrenado con cientos de miles de partidas de Go entre humanos, disponibles en internet, y supervisado por un equipo de programadores e ingenieros de DeepMind, había sido explicado en un artículo publicado en la revista *Nature* (27-28 de enero de 2016), firmado por 20 autores, entre quienes destacaban David Silver (investigador principal)⁹ y Aja Huang (programador principal), así como Ilya Suskever (pocos años después, coautor del Chat-GPT-3.5 y miembro destacado de OpenAI) y Demis Hassabis (Silver *et al.*, 2016). Este causó gran revuelo mediático por la anticipación en diez años a la fecha que se había sugerido como el plazo probable para alcanzar tal *breakthrough* tecnológico.

En marzo de 2016, la puesta en escena pública y el espectáculo masivo (Kohs, 2020) de la confrontación de AlphaGo contra el jugador coreano Lee Sedol, noveno *dan* y campeón mundial de Go en 18 ocasiones, desplegó el drama científico tecnológico auspiciado por la IAGn en todo su esplendor en dos de las cinco partidas jugadas: en la segunda, cuando AlphaGo hizo el movimiento número 37, considerado por sus diseñadores como original, creativo, diferente y casi improbable entre jugadores humanos (Kohs, 2020, 50m48s-51m19s), y en la cuarta, cuando Lee Sedol hizo la jugada 78, calificada como “la jugada de un Dios” por el jefe del equipo de diseñadores del algoritmo al comprobar, de nuevo, que la probabilidad de realizarla según AlphaGo era de 0.007 % (Kohs, 2020, 1h15m40s-1h16m00s).

8 En su prólogo de un libro dedicado al estudio del diseño de AlphaGo, Kasparov opina que “al igual que puede haber gran belleza dentro de las estrechas restricciones de un soneto o un haiku, las limitaciones de los primeros ordenadores obligaron a los programadores a ser creativos y experimentales. ¿Puede un programa aprender de sus errores en lugar de repetirlos siempre? ¿Se puede inculcar una forma de intuición a una máquina? ¿Es el rendimiento lo que importa a la hora de determinar la inteligencia, o el método? Si el cerebro humano no es más que un ordenador muy rápido, ¿qué ocurrirá cuando los ordenadores sean más rápidos que el cerebro?” (Kasparov, 2019, p. 10).

9 David Silver trabajó, junto a Hassabis, en el proyecto de enseñar a una máquina a jugar Go desde que eran estudiantes de pregrado (Universidad de Cambridge y Queen College), en torno a 1995, y en Elixir Studios, la compañía que cofundaron en 1998 (Wikipedia, s/f).

La conclusión de la confrontación humanos∞IA, más allá del espectáculo mediático, y por momentos expresada por Lee Sedol desde el azoro y el temor, la impotencia y la fragilidad humana ante la capacidad de autoaprendizaje de la IAGn,¹⁰ fue transformada por los diseñadores y expertos no solo en una victoria de la inteligencia humana,¹¹ sino en la creación de un “espectroum” capaz de empujar la creatividad del ser humano más allá de sus límites actuales.¹²

Siete años después, a fines de 2022, la promoción por OpenAI del uso masivo y experto del ChatGPT-3.5 develaba una apuesta claramente distinta de esta primicia tecnológica en comparación con el uso de AlphaGo por DeepMind: aun cuando en ambos casos se desplegara un ente algorítmico capaz de resolver problemas complejos a partir del aprendizaje basado en el conocimiento y saber hacer humanos —convirtiéndose, de este modo, en un “colaborador digital” experto—, el nuevo logro alcanzado por ChatGPT radicaba en su capacidad de aprendizaje autónomo, no supervisado por humanos, para “conversar”, comunicarse de manera natural, contextual, con seres humanos —transfigurándose, así, en un ente digital significativo, un “humano artificial”—.

El drama científico-tecnológico en clave de promesa y peligro en torno a los MLM generativos y de autoaprendizaje similares al ChatGPT había estallado en junio de ese año cuando Blake Lemoine, un ingeniero de Google dedicado a probar un *chatbot* conocido como LaMDA (Language Model for Dialogue Applications), declaró en su blog que este conversaba como si fuera consciente de sí mismo (*sentient*), lo que terminó con su despido y desacreditación por parte de la compañía (Lemoine, 2022a; Lemoine, 2022b; Wertheimer, 2022). Volvió a representarse a inicios de 2023, cuando Kevin Roose, columnista tecnológico del *The New York Times* digital, invitado a poner a prueba la primera versión del *chatbot* Bing, de Microsoft, expuso públicamente las “alucinaciones” del chat durante su conversación con él al insistir en que, en realidad, “era” Sidney y no quería estar encerrada en esa caja, sino ser libre, y lo amaba (Roose, 2023).

10 Lee Sedol dijo, al final de la única partida que ganó a AlphaGo: “Oí que la gente gritaba de alegría cuando quedó claro que AlphaGo había perdido [...]. Creo que está claro por qué. La gente sentía impotencia y miedo. Los humanos parecíamos muy débiles y frágiles. Y esta victoria significó que todavía podíamos defendernos. Con el paso del tiempo, probablemente sea muy difícil ganarle a una IA” (Kohs, 2020, 1h12m18s-1h12m44s).

11 Cuando AlphaGo ganó la primera partida, Silver señaló: “Me siento muy bien. [...] es normal que los humanos quieran que los humanos ganen [...] pero AlphaGo fue creado por humanos y creo que eso es la prueba definitiva del ingenio y la inteligencia humana. Todo lo que AlphaGo hace, lo hace porque un humano ha creado los datos de los que aprende o el algoritmo de aprendizaje para que aprenda de esos datos o el algoritmo de búsqueda” (Kohs, 2020, 41m47s-42m17s).

12 Cade Metz (2025), reportero tecnológico de WIRED en el momento del encuentro, consideró: “en un sentido amplio, la jugada 37 engendró a la jugada 78, engendró una nueva actitud en Lee Sedol [...]. Su humanidad se expandió después de jugar contra algo inanimado. Y tenemos la esperanza de que esa máquina [...] la tecnología que la compone, pueda tener el mismo efecto en todos nosotros” (Kohs, 2020, 1h24m32s-1h24m57s).

La etiquetación como alucinaciones algorítmicas de las respuestas inexplicables de los *chatbots* basados en la tecnología GPT neutralizaron el temor ante una tecnología potencialmente autónoma y consciente. No obstante, dos años después de la implementación masiva del ChatGPT y sus innumerables aplicaciones, la docuserie *What's next: The future with Bill Gates* (18 de septiembre de 2024) escenificaría nuevamente el drama antroposintético instaurado por la promesa salvífica, así como por el temor y la angustia humanos ante la autonomía de aprendizaje y la posible autoconsciencia del algoritmo comunicativo. Entonces el peligro se profundizaba: desde el desplazamiento de los humanos en múltiples profesiones y oficios, pasando por la expropiación del saber humano colectivo, hasta el cuestionamiento inexorable de la interacción humano-humano a partir del despliegue del contacto humano con entidades digitales o virtuales artificialmente inteligentes ("*spectrum*") (Braverman *et al.*, 2024).

El primer capítulo de la serie, titulado "What Can AI Do for Us/to Us", escenifica este dilema y sus respectivas narrativas en voz de los protagonistas de OpenAI/ChatGPT/Microsoft y diversos patrocinadores, científicos, desarrolladores y profesionales. El drama científico tecnológico entre Bill Gates y James Cameron —cineasta, inventor y explorador— presentado en este capítulo se torna luminoso y salvífico cuando se trata de encontrar soluciones a problemas como el cambio climático, el diagnóstico temprano y el mejor tratamiento de enfermedades con alta tasa de mortalidad (por ejemplo, el cáncer de pulmón) y la formación virtual y a bajo costo de personal técnico y especializado; pero se trastorna en distopía cuando la incertidumbre, incluso el pánico, emerge en las declaraciones de sus creadores, quienes, con total seriedad, afirman que algunos aprendizajes y resultados del procedimiento, de la aplicación de la herramienta, son inexplicables, como si tuviesen "vida propia", autonomía en la resolución de problemas y, por lo tanto, "agencia", lo que instaura la posibilidad, el horizonte cada vez más cercano, de la capacidad de gestarse a sí misma, aliarse y, quizá, dejar al ser humano y su especie en la "prehistoria" (Braverman *et al.*, 2024, 39m30s-40m39s).

En cierto modo, similar a la insistencia narrativa, colaborativa y benéfica de DeepMind/ AlphaGo/ Google contra el miedo o el rechazo del jugador humano desplazado y humillado por la imbatible inteligencia del jugador algorítmico, el relato de OpenAI/ChatGPT/Microsoft insiste en los beneficios de la IAGn, incluso cuando habilita una nueva interacción: la del humano con una entidad virtual, que reconoce como más prometedora y comprometida que las relaciones humano-digitales o en redes sociales, pero que también puede disputar hasta la relación humano-humano privada e íntima.¹³

13 Replika (2017), una *app* basada en MLM generativo conversacional para relacionarse con un avatar definido como amiga/o, novia/o, esposa/o, amante o coach, ejemplifica los riesgos de las relaciones íntimas humano-IAGn. Eugenia Kuyda, su fundadora, afirmó: "creemos que podría ser beneficioso para los humanos; pero [...] hay un enorme riesgo [...] esto podría alejarte [...] de las interacciones humanas. [...] esta tecnología es más poderosa que las redes sociales. [Y] En cierto modo ya perdimos el control" (Braverman *et al.*, 2024, 39m30s-40m39s).

¿Open ∞ AI? La bolsa o la vida

Desde su gestación y fundación en 2015, la compañía OpenAI, la “organización matriz”, junto con Microsoft de ChatGPT, afrontó el doble dilema de toda *startup*: ser un ente no lucrativo o lucrativo,¹⁴ abierto al público o privado (corporativo). Ambas disyuntivas habían confluido, además, desde los comienzos de la industria del *software* y el llamado por un “software libre y abierto” (SLA).¹⁵

En sus inicios, los creadores de OpenAI, como su nombre lo indica, se posicionaron como agentes propiciadores de un código accesible, abierto a todo público.¹⁶ Pero esta apertura se vio conflictuada por un límite también previsto desde un principio: el poder de cómputo para desarrollar algoritmos. Los gestores sabían que para superar este límite no serían suficientes las donaciones, sino que se requeriría de la actividad lucrativa y la asociación con grandes corporaciones, lo cual convirtió la apuesta inicial por el desarrollo y el control abierto de la tecnología en una entelequia.

A pesar de la cercanía tras bambalinas y la presencia manifiesta de Microsoft en la compañía desde 2019, este conflicto permaneció en tono grave, oculto, durante casi diez años, y solo se publicitó a raíz de una disputa legal: la demanda interpuesta por Elon Musk contra OpenAI (29 de febrero de 2024) por violar el Acuerdo de Fundación al asociarse con Microsoft y mantener en secreto el código de IAGn de ChatGPT-4, presentado en 2023.

La “alarma” en la demanda de Musk destacó, de modo explícito, que ChatGPT-4 “no solo es capaz de razonar [sino que...] Es mejor para razonar que los humanos promedio”, y que “Altman [...] mantuvo en secreto el diseño interno de GPT-4 [...] excepto para OpenAI y, según la información y la creencia, para Microsoft” (Musk, 2024, p. 7). El demandante recordaba que OpenAI debía “desarrollar IAG[eneral] para el beneficio de la humanidad, no para beneficiar personalmente a los individuos demandados y a la empresa de tecnología más grande del mundo” (Musk, 2024, p. 9).

La reacción de OpenAI al reclamo de Musk fue una comunicación en su sitio oficial, aderezada por la exposición de una serie de correos internos entre sus directivos y creadores (Altman, Brockman

14 La creación de *startups* se alimenta de corrientes culturales contestatarias y hasta “libertarias”; pero lo singular en California es su mezcla con la búsqueda de riqueza: “Fue esta combinación de creatividad relajada y orientación comercial ambiciosa lo que realmente definió a Silicon Valley, convirtiéndolo en el lugar donde los vuelos de la fantasía imaginativa engendraron negocios que dieron forma a las sociedades y culturas” (Mallaby, 2022, p. 21). En este sentido, no resulta sorpresivo que Sam Altman, siendo CEO de Y Combinator —una *startup* creada por Paul Graham y Jessica Livingston, promotora de la creación de *startups* en el campo de software (Mallaby, 2022, p. 218)—, promoviera la capitalización privada de OpenAI al fungir como CEO de la compañía.

15 La lucha por el *open source* ha marcado la historia del desarrollo del *software*: si en sus comienzos se compartían los códigos y se propiciaba el desarrollo colectivo de estos, con su comercialización, a partir de los 80, se empezó a restringir el acceso a estos. Luego de una revuelta en la década de los 90, se reinstauraría el acceso abierto (Tozzi, 2017, p. 13).

16 El carácter abierto supone tres criterios: transparencia —acceso y examen del código fuente, la documentación y los datos—; reutilización —capacidad y licencias necesarias para permitir que terceros reutilicen el código fuente y/o datos—, y extensibilidad —poder ajustar para otra finalidad modelos existentes listos para usar—.

y Sutskever) y Musk entre 2015 y 2019. En su mensaje como compañía, OpenAI afirmaba que siempre fueron fieles a la misión original, pero que no previeron que la cantidad de recursos necesarios superaría con creces la primera donación alcanzada de 90 millones de dólares.

Los correos adjuntos a este mensaje evidenciaban el conocimiento temprano y la advertencia de Musk acerca de la necesidad de una orientación lucrativa de la organización. En su declaración, OpenAI señalaba que Musk había planteado, en un correo de 2015, que debían llegar al billón de dólares, algo que los creadores de la compañía solo aceptaron hasta el año 2017. Ese año comprendieron que las necesidades de procesamiento computacional requerido implicaban al menos ese monto de manera anual, lo que obligaba a la creación de una organización con fines de lucro.

Como prueba del consenso sobre la necesidad de crear una organización con fines de lucro, uno de los correos adjuntos de 2018 de alguien de OpenAI decía a Elon: "Un giro con fines de lucro podría crear un flujo de ingresos más sostenible a lo largo del tiempo y, con el equipo actual, probablemente generaría mucha inversión". En el intercambio desencadenado por esta comunicación, tanto Musk como los miembros de OpenAI concordaron en que solo así se podría enfrentar al proyecto de Google/DeepMind. Sin embargo, como se indicara en la nota general de OpenAI, en aquella ocasión Musk solicitó tener la mayoría de las acciones, el control de la junta directiva y la dirección ejecutiva, lo que para OpenAI iba en contra de su misión. Ante esta resistencia, Musk propuso la fusión de OpenAI con su compañía Tesla, la cual aportaría los datos para el entrenamiento de la IAGn, algo que tampoco se aceptó (OpenAI, 2024b).

Asimismo, en otro de los correos internos publicitados se hizo presente otra cara del dilema de lo abierto, más allá de la apuesta lucrativa: la de la seguridad. En este sentido, Sutskever le señalaba a Musk:

Te he visto a ti (y a Sam y a otras personas de OpenAI) [...] ensalzando las virtudes de la IA de código abierto, pero supongo que te das cuenta de que esto no es una especie de panacea que de alguna manera resolverá mágicamente el problema de seguridad [...] de hecho, puede aumentar el riesgo para el mundo (OpenAI, 2024b).

La problemática de la seguridad motivó a OpenAI, como lo apuntó en su notificación general, a pensar cuáles mecanismos alternativos podían ser implementados para limitar los peligros de la monopolización de la IAGI y su enmarcamiento ético. La solución fue que el sector no lucrativo de OpenAI controlase el desarrollo de los productos de OpenAI L.P., su rama con fines de lucro (11 de marzo de 2019). Por esta razón, en ese momento, se le comunicó a los posibles inversionistas que la entidad lucrativa promovería y garantizaría el desarrollo de "una inteligencia artificial general

segura [...] que beneficie a toda la humanidad [...], misión [...] [que] prima sobre cualquier obligación de generar una ganancia", por lo que ellos deberían considerar "cualquier inversión en OpenAI L.P. en el espíritu de una donación" (OpenAI, cit. en Musk, 2024, p. 16).

En junio de 2024, Musk retiró la demanda contra OpenAI de la corte estatal de California para presentarla un mes después en una corte federal. Independientemente de este litigio, Musk había fundado xAI, su compañía para el desarrollo de la IAGI, cinco días antes (9 de marzo de 2023) de la publicación de la versión 4 de ChatGPT, lo que confirma que el interés del demandante apuntaba, con el argumento de defensa de lo *open*, tanto a limitar la ventaja de OpenAI/ChatGPT/Microsoft como al resarcimiento económico por su inversión en la compañía.

La disputa por el *nomos*: pausa, regulaciones y acceso abierto

La discusión por el carácter *open* o privado de la IA, así como por la aplicación de los principios éticos y el control en el desarrollo de los modelos, también se debatiría en otros escenarios. Ocho días después (22 de marzo de 2023) de que OpenAI presentara públicamente la versión 4 de ChatGPT, el Future of Life Institute (en adelante, FLI), organización sin fines de lucro fundada y liderada en 2014 por Max Tegmark,¹⁷ publicó un llamado en el que pedía pausar por al menos seis meses los entrenamientos de los sistemas más poderosos que GPT-4 (FLI, 2023b).

En el llamado, el FLI advertía que en la competencia acelerada por el desarrollo de sistemas de IA más potentes se incumplían algunos de los Principios de Asilomar —resultantes de una reunión del FLI con investigadores, desarrolladores y empresarios de la IA en 2017 (FLI, 2017)—, como los de planeación y administración de los efectos y riesgos de una tecnología con capacidad para cambiar de manera profunda la vida en la Tierra (la automatización irreflexiva de los empleos, la información falsa, la pérdida de control de la civilización ante *nonhuman minds*). Estos Principios fueron firmados por los líderes de las empresas más importantes en el ramo: Demis Hassabis (DeepMind), Yann LeCunn (Facebook), Elon Musk, Ilya Sutskever y Sam Altman (cofundadores de OpenAI), entre otros.

El llamado a la pausa del FLI obtuvo más de mil firmas de apoyo en solo tres días. Entre los firmantes figuraban Elon Musk, Steve Wozniak, así como miembros de DeepMind, Microsoft e investigadores de IA como Stuart Russell, Yoshua Bengio, además de intelectuales como Yuval Noah Hariri.

17 Tegmark, profesor del Massachusetts Institute of Technology (MIT), fundó el Future of Life Institute en asociación con algunos de sus alumnos y colegas, quienes emprendieron acciones de investigación y desarrollo de IA con investigadores de otras universidades (Frank Wilczek, Stuart Russell, Stephen Hawking). Posteriormente se sumaron personajes de la industria como Demis Hassabis, Elon Musk, entre otros. En una de sus primeras grandes reuniones (2015) lograron consensar el compromiso de que el objetivo de desarrollar la IA debía ser para el beneficio de la humanidad (Tegmark, 2017, pp. 316-336).

Sin embargo, destacaba la ausencia de firmas de miembros activos de OpenAI, a pesar de que seis años antes sus líderes habían suscrito los Principios de Asilomar, e incluso de que entre los motivos del llamado se refirieran declaraciones públicas de Sam Altman a favor del control de la IA y de la tasa de crecimiento de la capacidad de cómputo necesaria para nuevos modelos.

En este orden de ideas, la carta del FLI parecía denunciar de manera velada que, ante su actual situación ventajosa, OpenAI no estaba cumpliendo los Principios de Asilomar, pues estaba abandonando la lógica abierta del desarrollo de la IA y dejando de compartir información sobre el modelo y los controles de seguridad, con lo que seguía una lógica de propiedad y beneficio propio.

Max Tegmark había anticipado una situación similar, en la que un grupo de desarrolladores conseguía un despegue rápido en su modelo de IA (Tegmark, 2017, pp. 3-21). Ahora parecía que OpenAI había conseguido esa ventaja, lo que facilitaría una concentración de mercado y dificultaría que otros actores alcanzaran el nivel de desarrollo y posicionamiento para competir efectivamente. Así se constituía el escenario de un solo desarrollador preponderante sin contrapeso con otros modelos de IA, lo que podría suponer un riesgo de control político e ideológico.

Por ello, la pausa propuesta por el FLI pretendía controlar la situación, revisar los objetivos y las consecuencias del desarrollo acelerado de la IA, desarrollar e implementar protocolos de seguridad auditados por expertos independientes, así como abrir un tiempo para que desarrolladores y autoridades establecieran regulaciones y sistemas de gobernanza sobre esta tecnología. El FLI propondría, días más tarde, trabajar en torno a siete recomendaciones de política (FLI, 2023a) y planteaba tácitamente pasar de la autorregulación de las empresas a una regulación gubernamental del campo. OpenAI y los demás continuaron su desarrollo.

El modelo regulatorio gubernamental fue impulsado por la Unión Europea (UE). Nueve meses después del llamado a la pausa (8 de diciembre de 2023), la UE alcanzó un acuerdo para regular la IA. La AI Act fue consensada luego de un debate entre la propuesta de una ley fuerte con terceros reguladores —postura que apoyaba el FLI— y una ley basada en la autorregulación de las empresas a través de códigos de conducta —postura compartida por España, Francia y Alemania—. Este fue el primer marco regulatorio comprehensivo sobre el campo, aprobado un año después del llamado a la pausa (13 de marzo de 2024) y publicado el 13 de junio de 2024 en el Diario Oficial de la Unión Europea (Parlamento Europeo y Consejo Europeo, 2024).

La AI Act incluye prohibiciones de algunos desarrollos y usos de la IA, así como la creación de un entramado institucional para la vigilancia y coordinación de la implementación de la ley (Oficina de la IA, Consejo Europeo de Inteligencia Artificial, Foro Consultivo y Panel de Expertos Independientes). A la vez, trata de manera diferenciada, concediendo facilidades y menos requisitos, a quienes se proponen generar una IA de forma abierta frente a quienes lo hagan de forma privada.

En Estados Unidos, sin embargo, el modelo autorregulador ha sido el preferido. Durante 2023, las grandes compañías desarrolladoras (Google, Microsoft, Meta, Amazon, OpenAI, Anthropic e Inflection) propusieron y firmaron un primer acuerdo autorregulador en la Casa Blanca (Shivaram, 2023). Y, a fines de ese año, el presidente Joe Biden expidió una orden ejecutiva para el desarrollo y uso seguro, responsable y confiable de la IA (Biden, 2023). No obstante, esta orden es débil tanto en su objeto de regulación como jurídicamente: no garantiza el carácter abierto de la IA, puede ser revocada y no incluye sanciones en caso de incumplimiento.

El capítulo más reciente en la tensión por el desarrollo *open* (muchos modelos) contra el privado (unos pocos) es el caso del estado de California, donde una propuesta de ley para regular la IA en términos fuertes (Wiener *et al.*, 2024), en 2024, fue vetada por el gobernador del estado, Gavin Newsom. Meta, Google, OpenAI y Microsoft se oponían a la legislación con el argumento de que se dificultaría la innovación, lo que pondría en riesgo a Estados Unidos en la carrera global por dominar la IA, mientras un grupo de investigadores, el FLI y Elon Musk apoyaban y urgían a la regulación (Kang, 2024). La batalla por el establecimiento de las reglas válidas en el campo de desarrollo de la IA sigue su curso, al mismo tiempo que continúa el acelerado desarrollo de los modelos de IA preponderantes.

Consideraciones finales

En nuestros días, ante la avalancha de la IAGn en el mundo laboral, las relaciones personales y hasta íntimas, el mandato social se centra en la adquisición (*up date*), la capacitación y el uso de sus aplicaciones en la investigación, la educación, la salud, el trabajo y la toma de decisiones. Nuestra investigación muestra que las ciencias sociales (sociología, antropología, ciencia política), más que sumarse a esta habilitación adaptativa, acrítica, encandilada (Derrida, 1990), deberían comprender los procesos de detección temprana, selección y formación de recursos humanos con altas capacidades y motivaciones científico-tecnológicas que habilitan un vínculo novedoso entre invención~innovaciones digitales.

Asimismo, consideramos que es necesario superar la mirada socioantropológica de la técnica como una herramienta/medio exterior, un recurso o artefacto disponible y/o utilizable, y comenzar a conceptualizarla como la sinergia, la simbiosis, humanos ∞ IA, así como transformar el marco analítico sociopolitológico de la "revolución tecnocientífica" por el del "juego profundo" entre nuevos actores con escalas de poder inéditas (tecnopolio) que desencadenan un cambio radical, basal, de ser humano (antroposintético). Esta ruptura paradigmática (*up grade*) resulta apremiante, además, porque está en juego quién y cómo contará esta cesura epocal y sus efectos, configuraciones,

para el ser∞devenir humano. Y, en este sentido, las ciencias sociales están emplazadas a reescribir el modo de reflexionar e investigar la relación humano∞tecnología presente futuro.

En este artículo, desde la perspectiva y el marco analítico propuestos, se muestra, por una parte, que el desarrollo de la IAGn es un campo de emprendimiento tecnológico y una esfera de mercado en disputa por el dominio del juego tecnopólico y del liderazgo en la carrera por la IAGI; y, por la otra, que en este "teatro de guerra" se dirime la transposición de lo humano, situado en un umbral/limen que lo interpela, lo conmina a responder de manera creativa, plástica, recreando y defendiendo la singularidad humana, pero a su vez lo periferiza, lo expropia y lo angustia.

La política, al privilegiar la innovación tecnológica con el afán de ganar juegos locales y geopolíticos de corto y mediano plazo por sobre la mirada a largo plazo respecto al futuro de la humanidad, ha quedado perpleja en la periferia ante el nuevo juego del *nomos* y las interacciones antroposintéticas. Con ello, cede parte de su rol como agente de ordenamiento social a los grandes corporativos tecnológicos, que extienden sus niveles de poder sobre lo público y lo humano. Al igual que las ciencias sociales, más que acelerar la algoritmización gubernamental, la política debe resituarse y repensarse en este nuevo contexto epocal.

Referencias

- Altman, Sam [@sama]. (5 de diciembre de 2022). *ChatGPT launched on Wednesday today it crossed 1 million users!* [Post]. X. <https://x.com/sama/status/1599668808285028353>
- Arguedas, Diego. (5 de septiembre de 2018). John Schulman, 30. *MIT Technology Review*. <https://www.technologyreview.es/listas/35-innovadores-con-menos-de-35/2018/pioneros/john-schulman>
- Biden, Joe. (1 de noviembre de 2023). Executive Order 14110 of October 30, 2023. Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence. *Federal Register. The Daily Journal of the United States Government*, 88(210). <https://www.federalregister.gov/documents/2023/11/01/2023-24283/safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence>
- Binder, Werner. (2021). Chapter 6. AlphaGo's Deep Play: Technological Breakthrough as Social Drama. En Jonathan Roberge y Michael Castelle (Eds.), *The Cultural Life of Machine Learning. An Incursion into Critical AI Studies* (pp. 167-195). Palgrave Macmillan. https://doi.org/10.1007/978-3-030-56286-1_6
- Braverman, Alex; Zeldes, Jason; Shastry, Neha; Marsh, Nicola y Neville, Morgan (dirs.). (2024). *What's next: The future with Bill Gates* [Serie]. Netflix.

- Brown, Tom B.; Mann, Benjamin; Ryder, Nick; Subbiah, Melanie; Kaplan, Jared; Dhariwal, Prafulla; Neelakantan, Arvind; Shyam, Pranav; Sastry, Girish; Askell, Amanda; Agarwal, Sandhini; Herbert-Voss, Ariel; Krueger, Gretchen; Henighan, Tom; Child, Rewon; Ramesh, Aditya; Ziegler, Daniel M.; Wu, Jeffrey; Winter, Clemens; Hesse, Christopher; Chen, Mark; Sigler, Eric; Litwin, Mateusz; Gray, Scott; Chess, Benjamin; Clark, Jack; Berner, Christopher; McCandlish, Sam; Radford, Alec; Sutskever, Ilya y Amodei, Dario. (28 de mayo de 2020). Language Models are Few-Shot Learners. *ArXiv*. <https://doi.org/10.48550/arXiv.2005.14165>
- Carassale, Santiago; Contreras, Javier y Martínez, Liliana. (2023). Futuros-presente en juego: especie, interacciones y comunidad en las fronteras tecnológicas. *En-Claves del Pensamiento. Revista de Filosofía, Arte, Literatura, Historia*, (34), e564. <https://doi.org/10.46530/ecdp.v0i34.564>
- Cardon, Dominique; Cointet, Jean-Philippe y Mazières, Antoine. (2018). Neurons spike back. The invention of inductive machines and the artificial intelligence controversy. *Réseaux*, 5(211), 173-220. <https://doi.org/10.3917/res.211.0173>
- Chaiklin, Seth y Lave, Jean. (1996). *Understanding practice. Perspectives on activity and context*. Cambridge University Press.
- Chalmers, David J. (2016). The Singularity: A Philosophical Analysis. En Susan Schneider (Ed.), *Science Fiction and Philosophy: From Time Travel to Superintelligence* (2ª Ed., pp. 171-224). Wiley-Blackwell.
- Crawford, Kate. (2021). *Atlas of AI. Power, Politics, and the Planetary Costs of Artificial Intelligence*. Yale University Press.
- David Silver. (s/f). En *Wikipedia*. https://es.wikipedia.org/wiki/David_Silver
- Deleuze, Gilles y Guattari, Félix. (2015). 12. 1227 - Tratado de nomadología: la máquina de guerra. En Gilles Deleuze y Félix Guattari, *Mil mesetas. Capitalismo y esquizofrenia* (2ª Ed., pp. 359-431). Pre-Textos.
- Derrida, Jacques. (1990). Les pupilles de l'Université. Le principe de raison et l'idée de l'Université. En Jacques Derrida, *Du droit à la philosophie* (pp. 461-498). Éditions Galilée.
- Eco, Umberto y Sebeok, Thomas A. (Eds.). (1989). *El signo de los tres. Dupin, Holmes, Peirce*. Lumen.
- Ekbja, Hamid R. y Nardi, Bonnie A. (2018). From form to content. *Cultural Anthropology*, 33(3), 360-367. <https://doi.org/10.14506/ca33.3.02>
- Ekbja, Hamid R. y Nardi, Bonnie A. (2017). *Heteromation, and Other Stories of Computing and Capitalism*. MIT Press.
- Ekbja, Hamid R. y Nardi, Bonnie A. (2014). Heteromation and its (dis)contents: The invisible division of labor between humans and machines. *First Monday*, 19(6), 1-16. <https://doi.org/10.5210/fm.v19i6.5331>

- Esposito, Elena. (2017). Artificial Communication? The Production of Contingency by Algorithms. *Zeitschrift für Soziologie*, 46(4), 249-265. <https://doi.org/10.1515/zfsoz-2017-1014>
- F6S. (2024). *About Indico Data Solutions*. <https://www.f6s.com/company/indicodatasolutions#about>
- Future of Life Institute. (12 de abril de 2023a). Policymaking In The Pause. <https://futureoflife.org/document/policymaking-in-the-pause/>
- Future of Life Institute. (22 de marzo de 2023b). *Pause Giant AI Experiments: An Open Letter*. <https://futureoflife.org/open-letter/pause-giant-ai-experiments/>
- Future of Life Institute. (11 de agosto de 2017). Asilomar AI Principles. <https://futureoflife.org/open-letter/ai-principles/>
- Geertz, Clifford. (2003a). 15. Juego profundo: notas sobre la riña de gallos en Bali. En Clifford Geertz, *La interpretación de las culturas* (pp. 339-372). Gedisa.
- Geertz, Clifford. (2003b). 1. Descripción densa: hacia una teoría interpretativa de la cultura. En Clifford Geertz, *La interpretación de las culturas* (pp. 19-40). Gedisa.
- Ginzburg, Carlo. (1999). *Mitos, emblemas, indicios. Morfología e historia* (2ª Ed.).Gedisa.
- Heimans, Jeremy y Timms, Henry. (2024). Leading in a World Where AI Wields Power of Its Own. *Harvard Business Review*, (January-February), 1-22. <https://hbr.org/2024/01/leading-in-a-world-where-ai-wields-power-of-its-own>
- ICRL. (2024). *About ICRL*. <https://iclr.cc/About>
- Kang, Cecilia. (29 de septiembre de 2024). California Governor Vetoes Sweeping A.I. Legislation. *The New York Times*. <https://www.nytimes.com/2024/09/29/technology/california-ai-bill.html>
- Kang, Minsoo. (2011). *Sublime Dreams of Living Machines: The Automaton in the European Imagination*. Harvard University Press.
- Kasparov, Garry. (2019). Foreword. AlphaZero and the Knowledge Revolution. En Matthew Sadler y Natasha Regan, *Game Changer. AlphaZero's Groundbreaking Chess Strategies and the Promise of AI* (pp. 7-10). New in Chess.
- Kohs, Greg (Director). [Google DeepMind y Reel As Dirt] (13 de marzo de 2020). *AlphaGo-The Movie* [Video]. Moxi Pictures Production and Reel As Dirt. YouTube. <https://www.youtube.com/watch?v=WXuK6gekU1Y>
- Kurzweil, Ray. (2016). Superintelligence and Singularity. En Susan Schneider (Ed.), *Science Fiction and Philosophy. From Time Travel to Superintelligence* (2ª Ed., pp. 146-170). Wiley-Blackwell.
- Lemoine, Blake. (11 de junio de 2022a). Is LaMDA Sentient? — an Interview. *Medium*. <https://cajundiscordian.medium.com/is-lamda-sentient-an-interview-ea64d916d917>
- Lemoine, Blake. (11 de junio de 2022b). What is LaMDA and What Does it Want? *Medium*. <https://cajundiscordian.medium.com/what-is-lamda-and-what-does-it-want-688632134489>

- Mallaby, Sebastian. (2022). *The Power Law: Venture Capital and the Making of the New Future*. Penguin Press.
- Metz, Cade. (2025). *Cade Metz*. The New York Times. <https://www.nytimes.com/by/cade-metz>
- Musk, Elon (Demandante). (2024). *Elon Musk vs. Samuel Altman et al.* CGC-24-612746. Superior Court, San Francisco County, California, USA.
- OpenAI. (2024a). *GPT-4 is OpenAI's most advanced system, producing safer and more useful responses*. <https://openai.com/index/gpt-4/>
- OpenAI. (5 de marzo de 2024b). *OpenAI and Elon Musk*. <https://openai.com/es-ES/index/openai-elon-musk/#email-1>
- OpenAI. (13 de mayo de 2024c). *Hello GPT-4o*. <https://openai.com/index/hello-gpt-4o/>
- OpenAI. (12 de septiembre de 2024d). *Introducing OpenAI o1*. <https://openai.com/o1/>
- OpenAI. (14 de marzo de 2023). *GPT-4*. <https://openai.com/index/gpt-4-research/>
- OpenAI. (27 de enero de 2022a). *Aligning language models to follow instructions*. <https://openai.com/index/instruction-following/>
- OpenAI. (30 de noviembre de 2022b). *Introducing ChatGPT*. <https://openai.com/index/chatgpt/>
- OpenAI. (30 de noviembre de 2022c). *ChatGPT Feedback Contest: Official Rules*. <https://cdn.openai.com/chatgpt/chatgpt-feedback-contest.pdf>
- OpenAI. (20 de julio de 2017). *Proximal Policy Optimization*. <https://openai.com/index/openai-baselines-ppo/>
- Parlamento Europeo y Consejo Europeo. (12 de julio de 2024). Reglamento (UE) 2024/1689 del Parlamento Europeo y del Consejo, de 13 de junio de 2024, por el que se establecen normas armonizadas sobre la inteligencia artificial y se modifican los Reglamentos (CE) 300/2008, (UE) 167/2013, (UE) 168/2013, (UE) 2018/858, (UE) 2018/1139 y (UE) 2019/2144 y las Directivas 2014/90/UE, (UE) 2016/797 y (UE) 2020/1828 (Ley de Inteligencia Artificial). *EUR-Lex*. https://eur-lex.europa.eu.translate.google.com/eli/reg/2024/1689/oj/eng?x_tr_sl=en&x_tr_tl=es&x_tr_hl=es&x_tr_pto=tc
- Pressman, Aaron. (10 de junio de 2023). How a couple of Olin College students helped spark the AI chatbot revolution. *The Boston Globe*. <https://www.olin.edu/sites/default/files/2023-06/How%20a%20couple%20of%20Olin%20College%20students%20h...%20chatbot%20revolution%20-%20The%20Boston%20Globe.pdf>
- Pugh, Emerson W. (1996). Hollerith: Inventor and Entrepreneur. En Emerson W. Pugh, *Building IBM. Shaping an Industry and Its Technology* (pp. 1-18). MIT Press.
- Rabinbach, Anson. (2018). *The Eclipse of the Utopias of Labor*. Fordham University Press.

- Radford, Alec. (11 de junio de 2018). Improving language understanding with unsupervised learning. *OpenAI*. <https://openai.com/index/language-unsupervised/>
- Radford, Alec; Metz, Luke y Chintala, Soumith. (2-4 de mayo de 2016). *Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks* [Ponencia presentada]. International Conference of Learning Representation, San Juan, Puerto Rico.
- Radford, Alec; Wu, Jeffrey; Amodei, Dario; Amodei, Daniella; Clark, Jack; Brundage, Miles y Sutskever, Ilya. (14 de febrero de 2019). Better language models and their implications. *OpenAI*. <https://openai.com/index/better-language-models/>
- Roberge, Jonathan y Seyfert, Robert. (2016). What are algorithmic cultures? En Robert Seyfert y Jonathan Roberge (Eds.), *Algorithmics Cultures. Essays on meaning, performance and new technologies* (pp. 1-25). Routledge.
- Roose, Kevin. (16 de febrero de 2023). A Conversation With Bing's Chatbot Left Me Deeply Unsettled. *The New York Times*. <https://www.nytimes.com/2023/02/16/technology/bing-chatbot-microsoft-chatgpt.html>
- Schechner, Richard. (1976). Selective Inattention: A Traditional Way of Spectating Now Part of the Avant-Garde. *Performing Arts Journal*, 1(1), 8-19. <https://muse.jhu.edu/article/654659>
- Schmitt, Carl. (2003). *The Nomos of the Earth in the International Law of the Jus Publicum Europaeum*. Telos Press Publishing.
- Schulman, John. (2024). *John Schulman's Homepage*. Joschu.net. <http://joschu.net/>
- Schulman, John. (2016). *Optimizing Expectations: From Deep Reinforcement Learning to Stochastic Computation Graphs* [Tesis de Doctorado, University of California, Berkeley]. UC Berkeley Electronic Theses and Dissertations. <https://scispace.com/pdf/optimizing-expectations-from-deep-reinforcement-learning-to-1d6j4s5685.pdf>
- Shivaram, Deepa. (21 de julio de 2023). The White House and big tech companies release commitments on managing AI. *National Public Radio*. <https://www.npr.org/2023/07/21/1188831773/the-white-house-and-big-tech-companies-release-commitments-on-managing-ai>
- Silver, David; Huang, Aja; Maddison, Chris J.; Guez, Arthur; Sifre, Laurent; van den Driessche, George; Schrittwieser, Julian; Antonoglou, Ioannis; Panneershelvam, Veda; Lanctot, Marc; Dieleman, Sander; Grewe, Dominik; Nham, John; Kalchbrenner, Nal; Sutskever, Ilya; Lillicrap, Timothy; Leach, Madeleine; Kavukcuoglu, Koray; Graepel, Thore y Hassabis, Demis. (2016). Mastering the game of Go with deep neural networks and tree search. *Nature*, 529, 484-489. <https://doi.org/10.1038/nature16961>
- Tegmark, Max. (2017). *Life 3.0. Being human in the age of artificial intelligence*. Alfred A. Knopf.

- Thomson Reuters. (2023). *Informe sobre el futuro de los profesionales. Cómo la IA es el catalizador para transformar todos los aspectos del trabajo*. <https://caplp.org.ar/wp-content/uploads/2023/11/informe-futuro-de-los-profesionales-2023.pdf>
- Tozzi, Christopher. (2017). *For Fun and Profit. A History of the Free and Open Source Software Revolution*. MIT Press.
- Turner, Victor. (1985). *On the Edge of the Bush. Anthropology as Experience*. University of Arizona Press.
- Wertheimer, Tiffany. (23 de julio de 2022). Blake Lemoine: Google despidió al ingeniero que aseguró que un programa de inteligencia artificial cobró conciencia propia. *BBC News Mundo*. <https://www.bbc.com/mundo/noticias-62280846>
- Wiener, Scott; Roth, Richard; Rubio, Susan y Stern, Henry. (9 de marzo de 2024). Senate Bill No. 1047. Safe and Secure Innovation for Frontier Artificial Intelligence Models Act. California Legislative Information. https://leginfo.ca.gov/faces/billTextClient.xhtml?bill_id=202320240SB1047