

La singularidad está cerca. Flick

## **Émile P Torres**

Es candidato a doctorado en filosofía en Leibniz Universität Hannover en Alemania. Sus escritos han aparecido en *Philosophy Now*, *Nautilus*, *Motherboard* y el *Boletín de los Científicos Atómicos*, entre otros. Son es auyor de *The End: What Science and Religion Tell Us About the Apocalypse* (2016), *Morality, Foresight, and Human Flourishing: An Introduction to Existential Risks* (2017) y *Human Extinction: A History of the Science and Ethics of Annihilation* (de próxima publicación en Routledge).

Publicado inicialmente en *Aeon*. [2021](#)

Comenzó como una teoría filosófica marginal sobre el futuro de la humanidad. Ahora cuenta con una fuerte financiación y es cada vez más peligroso.

Parece haber un creciente reconocimiento de que la humanidad podría estar acercándose al “fin de los tiempos”. Las terribles predicciones de catástrofe abarrotan las noticias. Los videos en las redes sociales mostrando incendios forestales infernales, inundaciones

devastadoras y hospitales repletos de pacientes con COVID-19 dominan nuestras líneas de tiempo. Los activistas de Extinction Rebellion están cerrando ciudades en un intento desesperado por salvar el mundo. Una [encuesta](#) incluso encontró que más de la mitad de las personas preguntadas sobre el futuro de la humanidad “calificaron el riesgo de que nuestra forma de vida termine en los próximos 100 años en un 50 por ciento o más”.

El “apocalipticismo”, o la creencia de que el fin de los tiempos es inminente, no es, por supuesto, nada nuevo: la gente ha advertido que el fin está cerca durante milenios, y de hecho muchos eruditos del Nuevo Testamento creen que Jesús mismo esperaba que el mundo terminara durante su propia vida. Pero la situación actual es fundamentalmente diferente a la del pasado. Los escenarios “escatológicos” que ahora se discuten no se basan en las revelaciones de profetas religiosos o en metanarrativas seculares de la historia humana (como en el caso del marxismo), sino en conclusiones científicas sólidas defendidas por destacados expertos en campos como la climatología, la ecología, la epidemiología, etc.

Sabemos, por ejemplo, que el cambio climático representa una grave amenaza para la civilización. Sabemos que la pérdida de biodiversidad y la sexta extinción masiva podrían precipitar cambios repentinos, irreversibles y catastróficos en el ecosistema global. Un intercambio termonuclear podría borrar el Sol durante años o décadas, provocando el colapso de la agricultura mundial. Y ya sea que el SARS-CoV-2 provenga o no de un laboratorio de Wuhan o haya sido cocinado en la cocina de la naturaleza (esto último parece más [probable](#) en este momento), la biología sintética pronto [permitirá](#) a los malos actores diseñar patógenos mucho más letales y contagiosos que cualquier cosa que la evolución darwiniana pueda inventar. Algunos filósofos y científicos también han comenzado a hacer sonar la [alarma](#) sobre las “amenazas emergentes” asociadas con la superinteligencia de las máquinas, la nanotecnología molecular y la geoingeniería estratosférica, que no parecen menos formidables.

Tales consideraciones han llevado a muchos académicos a reconocer que, como Stephen Hawking escribió en *The Guardian* en 2016, “estamos en el momento más peligroso en el desarrollo de la humanidad”. Lord Martin Rees, por ejemplo, estima que la civilización tiene una probabilidad de 50/50 de llegar a 2100. Noam Chomsky [argumenta](#) que el riesgo de aniquilación es actualmente “sin precedentes en la historia del *Homo sapiens*”. Y Max Tegmark sostiene que “probablemente será dentro de nuestras vidas... cuando vamos a autodestruirnos o a actuar juntos”. De acuerdo con estas sombrías declaraciones, el *Boletín de los Científicos Atómicos* en 2020 [estableció](#) su icónico Reloj del Juicio Final a solo 100 segundos antes de la medianoche (o fatalidad), lo más cerca que ha estado desde que se

creó el reloj en 1947, y más de 11.000 científicos de todo el mundo firmaron un [artículo](#) en 2020 declarando “clara e inequívocamente que el planeta Tierra se enfrenta a una emergencia climática”, y sin “un inmenso aumento de escala en los esfuerzos por conservar nuestra biosfera [corremos el riesgo] de sufrir indeciblemente debido a la crisis climática”. Como la joven activista climática Xiye Bastida resumió este estado de ánimo existencial en una entrevista de *Teen Vogue* en 2019, el objetivo es “asegurarnos de que no somos la última generación”, porque esto parece ser ahora una posibilidad muy real.

Dados los peligros sin precedentes que enfrenta la humanidad hoy en día, uno podría esperar que los filósofos hayan derramado ríos de tinta sobre las implicaciones éticas de nuestra extinción, o de escenarios relacionados, como el colapso permanente de la civilización. ¿Hasta qué punto sería nuestra desaparición moralmente mala (o [buena](#)), y por qué razones? ¿Sería erróneo impedir que las generaciones futuras lleguen a existir? ¿Depende el valor de los sacrificios, luchas y esfuerzos pasados de que la humanidad continúe existiendo mientras la Tierra, o el Universo en general, siga siendo habitables?

Sin embargo, este no es el caso: hasta hace poco, el tema de nuestra extinción ha recibido poca atención sostenida de los filósofos, e incluso ahora ocupa una posición marginal en la discusión y el debate filosófico. En general, han estado preocupados por otros asuntos. Sin embargo, hay una excepción notable a esta regla: en las últimas dos décadas, un pequeño grupo de teóricos, en su mayoría con sede en Oxford, han estado ocupados trabajando en los detalles de una nueva cosmovisión moral llamada largoplacismo, que [enfatisa](#) cómo nuestras acciones afectan al futuro a muy largo plazo del universo: miles, millones, miles de millones e incluso billones de años a partir de ahora. Esto hunde sus raíces en el trabajo de Nick [Bostrom](#), quien fundó el grandiosamente llamado *Future of Humanity Institute* (FHI) en 2005, y Nick Beckstead, investigador asociado del FHI y director de programas en *Open Philanthropy*. Ha sido defendido públicamente por el filósofo de FHI Toby Ord, autor de *The Precipice: Existential Risk and the Future of Humanity* (2020). El largo plazo es el principal foco de investigación tanto del *Global Priorities Institute* (GPI), una organización vinculada a FHI dirigida por Hilary Greaves, como de la *Forethought Foundation*, dirigida por William MacAskill, quien también ocupa cargos en FHI y GPI. Además de la maraña de títulos, nombres, institutos y acrónimos, el largoplacismo es una de las principales “áreas de atención” (“*cause areas*”) del llamado movimiento del altruismo efectivo (AE), que fue introducido por Ord alrededor de 2011 y ahora [se jacta](#) de tener una alucinante cantidad de 46 mil millones de dólares en fondos comprometidos.



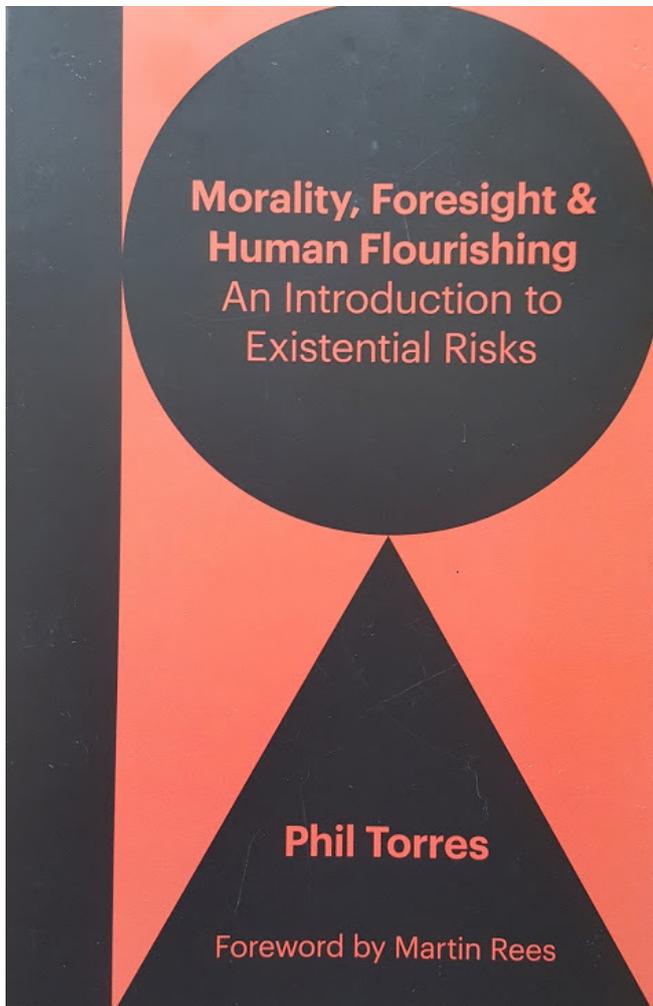
Fuente iStock. El confidencial

Es difícil exagerar cuán influyente se ha vuelto el largoplacismo. Karl Marx en 1845 declaró que el objetivo de la filosofía no es simplemente interpretar el mundo, sino cambiarlo, y esto es exactamente lo que los largoplacistas han estado haciendo, con extraordinario éxito. Consideremos que [Elon Musk](#), quien ha citado y respaldado el trabajo de Bostrom, ha donado 1,5 millones de dólares a FHI a través de su organización hermana, el aún más grandiosamente llamado *Future of Life Institute* (FLI). Este fue cofundado por el multimillonario empresario tecnológico Jaan Tallinn, quien, como [señalé](#) recientemente, no cree que el cambio climático represente un “riesgo existencial” para la humanidad debido a su adhesión a la ideología del largoplacismo.

Mientras tanto, el multimillonario libertario (*libertarian*) y partidario de Donald Trump,

Peter Thiel, quien una vez pronunció el discurso de apertura en una conferencia de AE, ha donado grandes sumas de dinero al *Instituto de Investigación de Inteligencia Artificial*, cuya misión de salvar a la humanidad de las máquinas superinteligentes está profundamente entrelazada con los valores a largo plazo. Otras organizaciones como GPI y la *Fundación Forethought* están financiando concursos de ensayos y becas en un esfuerzo por atraer a los jóvenes a la comunidad, mientras que es un secreto a voces que el *Centro para la Seguridad y las Tecnologías Emergentes* (CSET), con sede en Washington, DC, tiene como objetivo colocar a los largoplacistas en posiciones de alto nivel del gobierno de los Estados Unidos para configurar la política nacional. De hecho, CSET fue establecido por Jason Matheny, un ex asistente de investigación de FHI que ahora es el asistente adjunto del presidente de los Estados Unidos, Joe Biden, para tecnología y seguridad nacional. El propio Ord, sorprendentemente para un filósofo, “asesoró a la Organización Mundial de la Salud, el Banco Mundial, el Foro Económico Mundial, el Consejo Nacional de Inteligencia de los Estados Unidos, la Oficina del Primer Ministro del Reino Unido, la Oficina del Gabinete y la Oficina Gubernamental para la Ciencia”, y recientemente [contribuyó](#) a un informe del Secretario General de las Naciones Unidas que menciona específicamente el “largo plazo”.

La cuestión estriba en que el largoplacismo podría ser una de las ideologías más influyentes de las que pocas personas fuera de las universidades de élite y Silicon Valley han oído hablar. Creo que esto debe cambiar porque, como ex-largoplacista que publicó un [libro](#) completo hace cuatro años en defensa de la idea general, he llegado a ver esta cosmovisión como el sistema de creencias seculares posiblemente más peligroso del mundo actual. Pero para entender la naturaleza de la bestia, primero necesitamos diseccionarla, examinando sus características anatómicas y funciones fisiológicas.



Lo primero que hay que notar es que el largoplacismo, como proponen Bostrom y Beckstead, no es equivalente a “preocuparse por el largo plazo” o “valorar el bienestar de las generaciones futuras”. Va mucho más allá de esto. En su núcleo hay una analogía simple, aunque defectuosa, en mi opinión, entre las personas individuales y la humanidad en su conjunto. Para ilustrar la idea, considere el caso de [Frank Ramsey](#), un erudito de la Universidad de Cambridge considerado por sus compañeros como una de las mentes más excepcionales de su generación. “Había algo de Newton en él”, dijo una vez el *bestseller* Lytton Strachey. G E Moore escribió sobre la “brillantez excepcional” de Ramsey. Y John Maynard Keynes describió un artículo de Ramsey como “una de las contribuciones más notables a la economía matemática jamás hechas”.

Pero la historia de Ramsey no es feliz. El 19 de enero de 1930, murió en un hospital de

Londres después de una intervención quirúrgica, la causa probable de la muerte fue una infección hepática por nadar en el río Cam, que serpentea a través de Cambridge. Ramsey tenía solo 26 años.

Se podría argumentar que hay dos razones distintas por las que este resultado fue trágico. La primera es la más obvia: truncó la vida de Ramsey, privándolo de todo lo que podría haber experimentado si hubiera sobrevivido: las alegrías y la felicidad, el amor y la amistad: todo lo que hace que la vida valga la pena. En este sentido, la temprana muerte de Ramsey fue una tragedia personal. Pero, en segundo lugar, su muerte también robó al mundo una superestrella intelectual aparentemente destinada a hacer contribuciones aún más extraordinarias al conocimiento humano. “El número de senderos que Ramsey estableció fue notable”, escribe Sir Partha Dasgupta. Pero, ¿cuántos senderos más podría haber abierto? “Es angustioso pensar en la pérdida que supuso para su generación”, se lamentó Strachey, “qué luz se ha apagado”, lo que deja a uno preguntándose hasta qué punto la historia intelectual occidental podría haber sido diferente si Ramsey no hubiera muerto tan joven. Desde esta perspectiva, se podría argumentar que, aunque la tragedia personal de la muerte de Ramsey fue realmente terrible, la inmensidad de su potencial para haber cambiado el mundo para mejor hace que la segunda tragedia sea aún peor. En otras palabras, la maldad de su muerte proviene principalmente, tal vez abrumadoramente, de su potencial insatisfecho en lugar de los daños directos y personales que experimentó. O eso dice el argumento.

Los largoplacistas trasladarían estas afirmaciones y conclusiones a la propia humanidad, como si ésta fuera un individuo con su propio “potencial” para malgastar o realizar, arruinar o hacer realidad, en el transcurso de “su vida”. Entonces, por un lado, una catástrofe que reduzca la población humana a cero sería trágica debido a todo el sufrimiento que infligiría a quienes vivieran en ese momento. Imagínese el horror de morir de hambre a temperaturas bajo cero, bajo cielos oscuros al mediodía, durante años o décadas después de una guerra termonuclear. Esta es la primera tragedia, una tragedia personal para las personas directamente afectadas. Pero hay, argumentarían los largoplacistas, una segunda tragedia que es astronómicamente peor que la primera, que surge del hecho de que nuestra extinción excluiría permanentemente lo que podría ser un futuro extremadamente largo y próspero durante los próximos, digamos,  $10^{100}$  años (momento en el cual la “muerte térmica” hará que la vida sea imposible). Al hacer esto, destruiría irreversiblemente el “vasto y glorioso” potencial a largo plazo de la humanidad, en el lenguaje casi religioso de Ord, un “potencial” tan grande, dado el tamaño del Universo y el tiempo restante antes de alcanzar

el equilibrio termodinámico, que la primera tragedia palidecería por completo en comparación con esta segunda.

Esto sugiere inmediatamente otro paralelo entre los individuos y la humanidad: la muerte no es la única forma en que el potencial de alguien podría quedar sin realizar. Imagina que Ramsey no hubiera muerto joven, pero, en lugar de estudiar, escribir y publicar artículos académicos, hubiera pasado sus días en el bar local jugando al billar y bebiendo. Mismo resultado, diferente modo de fallo. Aplicando esto a la humanidad, los largoplacistas argumentarían que hay modos de fracasar que podrían dejar nuestro potencial sin que nos extingamos, a lo que volveré más adelante.

**Desde este punto de vista, una catástrofe climática será un pequeño problema, como un hombre de 90 años que se golpeó el dedo del pie cuando tenía dos años.**

Para resumir estas ideas hasta ahora, la humanidad tiene un “potencial” propio, uno que trasciende los potenciales de cada persona individual, y no realizar este potencial sería extremadamente malo, de hecho, como veremos, una catástrofe moral de proporciones literalmente cósmicas. Este es el dogma central del largoplacismo: nada importa más, éticamente hablando, que cumplir con nuestro potencial como especie de “vida inteligente originaria de la Tierra”. Importa tanto que los largoplacistas incluso han [acuñado](#) el término aterrador “riesgo existencial” para cualquier posibilidad de que nuestro potencial sea destruido, y “catástrofe existencial” para cualquier evento que realmente destruya este potencial.

¿Por qué creo que esta ideología es tan peligrosa? La respuesta corta es que elevar el cumplimiento del supuesto potencial de la humanidad por encima de todo podría aumentar de manera no trivial la probabilidad de que las personas reales, las vivas hoy y en el futuro cercano, sufran daños extremos, incluso la muerte. Considere que, como señalé en otra parte, la ideología largoplacista inclina a sus adherentes a adoptar una actitud despreocupada hacia el cambio climático. ¿Por qué? Porque incluso si el cambio climático hace que las naciones insulares desaparezcan, desencadena migraciones masivas y mata a millones de personas, probablemente no va a comprometer nuestro potencial a largo plazo en los próximos billones de años. Si uno tiene una visión cósmica de la situación, incluso una catástrofe climática que reduzca la población humana en un 75 por ciento durante los próximos dos milenios, en el gran esquema de las cosas, no será más que un pequeño problema, el equivalente a un hombre de 90 años que se golpeó el dedo del pie cuando tenía dos años.



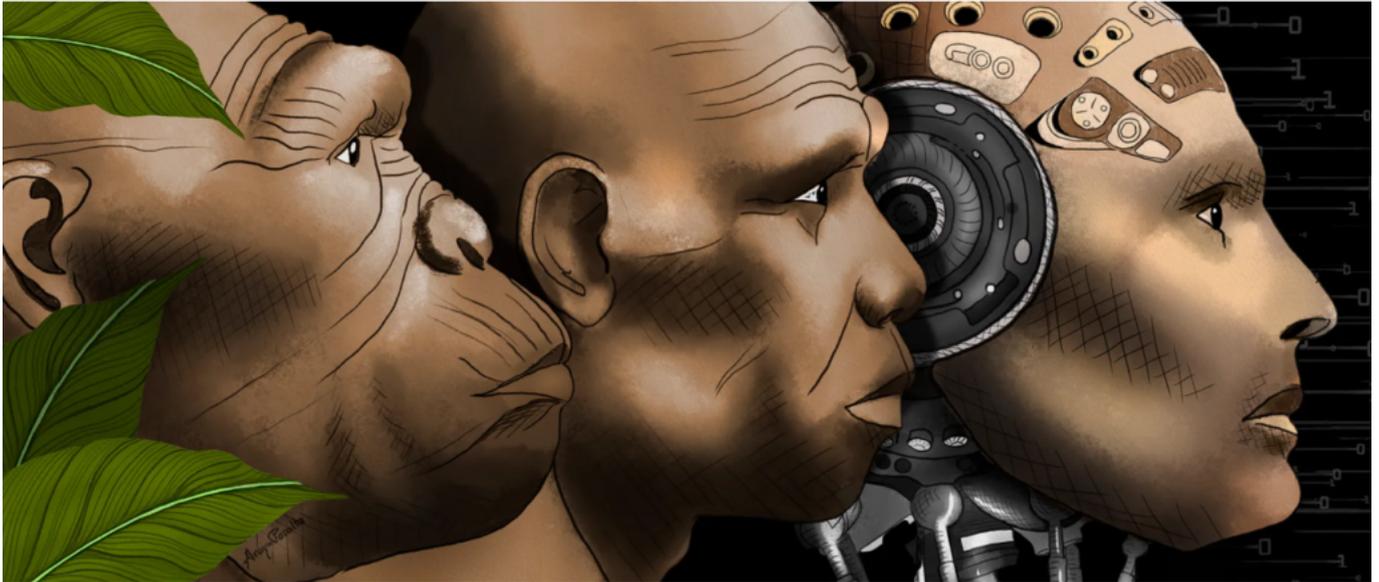
El **argumento** de Bostrom es que “un desastre no existencial que causa el colapso de la civilización global es, desde la perspectiva de la humanidad en su conjunto, un revés potencialmente recuperable”. Podría ser “una masacre gigante para el ser humano”, agrega, pero mientras la humanidad se recupere para alcanzar su potencial, finalmente se registrará como poco más que “un pequeño paso en falso para la humanidad”. En otra parte, escribe que los peores desastres naturales y las atrocidades devastadoras de la historia se convierten en trivialidades casi imperceptibles cuando se ven desde esta gran perspectiva. Refiriéndose a las dos guerras mundiales, el SIDA y el accidente nuclear de Chernóbil, declara que “por trágicos que sean tales acontecimientos para las personas inmediatamente afectadas, en el panorama general de las cosas ... Incluso la peor de estas catástrofes son meras ondas en la superficie del gran mar de la vida”.

Esta forma de ver el mundo, de evaluar la maldad del SIDA y del Holocausto, implica que los futuros desastres del mismo alcance e intensidad (no existenciales) también deben clasificarse como “meras ondas”. Si no representan un riesgo existencial directo, entonces no debemos preocuparnos mucho por ellos, por trágicos que puedan ser para los individuos. Como **escribió** Bostrom en 2003, “la prioridad número uno, dos, tres y cuatro debería ... para reducir el riesgo existencial”. **Reiteró** esto varios años después al argumentar que no debemos “malgastar ... Nuestros recursos finitos se destinan a “proyectos de eficacia subóptima”, como aliviar la pobreza mundial y reducir el sufrimiento animal, ya que ninguno

amenaza nuestro potencial a largo plazo, y nuestro potencial a largo plazo es lo que realmente importa.

Ord se hace eco de estos puntos de vista al argumentar que, de todos los problemas que enfrenta la humanidad, nuestra “primera gran tarea ... es llegar a un lugar de seguridad, un lugar donde el riesgo existencial, como él lo define, “es bajo y permanece bajo”, lo que denomina “seguridad existencial”. Más que cualquier otra cosa, lo que importa es hacer todo lo necesario para “preservar” y “proteger” nuestro potencial “extrayéndonos del peligro inmediato” y diseñando “salvaguardas sólidas que defiendan a la humanidad de los peligros en el futuro a largo plazo, de modo que sea imposible fallar”. Aunque Ord hace un guiño al cambio climático, también afirma, basándose en una metodología dudosa, que la posibilidad de que el cambio climático cause una catástrofe existencial es solo ~1 en 1,000, que es dos órdenes de magnitud más bajo que la probabilidad de que las máquinas superinteligentes destruyan a la humanidad este siglo, según Ord.

Lo que es realmente notable aquí es que la preocupación central no es el efecto de la catástrofe climática en personas reales de todo el mundo (recuerde, en el gran esquema, esto sería, en palabras de Bostrom, un “pequeño paso en falso para la humanidad”) sino la pequeña posibilidad de que, como dice Ord en *El precipicio*, esta catástrofe “plantee un riesgo de un colapso irre recuperable de la civilización o incluso la extinción completa de la humanidad”. Una vez más, los daños causados a las personas reales (especialmente los del Sur Global) pueden ser significativos en términos absolutos, pero cuando se comparan con la “inmensidad” y la “gloria” de nuestro potencial a largo plazo en el cosmos, apenas se registran.



Araya Peralta. *Ideas*

Sin embargo, las implicaciones del largo plazo son mucho más preocupantes. Si nuestras cuatro prioridades principales son evitar una catástrofe existencial, es decir, cumplir “nuestro potencial”, ¿qué no está sobre la mesa para que esto suceda? Considere el comentario de [Thomas Nagel](#) sobre cómo la noción de lo que podríamos llamar el “bien mayor” se ha utilizado para “justificar” ciertas atrocidades (por ejemplo, durante la guerra). Si los fines “justifican” los medios, argumenta, y se piensa que los fines son lo suficientemente grandes (por ejemplo, la seguridad nacional), entonces esto “puede aplicarse para aliviar las conciencias de los responsables de un cierto número de bebés carbonizados”. Ahora imagine lo que podría estar “justificado” si el “bien mayor” no es la seguridad nacional sino el potencial cósmico de la vida inteligente originaria de la Tierra en los próximos billones de años. Durante la Segunda Guerra Mundial, 40 millones de civiles perecieron, pero compare este número con las  $10^{54}$  o más personas (en la [estimación](#) de Bostrom) que podrían llegar a existir si podemos evitar una catástrofe existencial. ¿Qué no debemos hacer para “proteger” y “preservar” este potencial? ¿Para asegurar que estas personas no nacidas lleguen a existir? ¿Qué medios no pueden ser “justificados” por este fin moral cósmicamente significativo?

El propio Bostrom [argumentó](#) que deberíamos considerar seriamente establecer un sistema de vigilancia global e invasivo que monitorice a cada persona en el planeta en tiempo real, para amplificar las “capacidades de vigilancia preventiva” (por ejemplo, para prevenir ataques terroristas omnícidas que podrían devastar la civilización). En otra parte, ha escrito

que los estados deberían usar la violencia / guerra preventiva para evitar catástrofes existenciales, y argumentó que salvar a miles de millones de personas reales es el equivalente moral de reducir el riesgo existencial en cantidades completamente minúsculas. En sus palabras, incluso si hay “un mero 1 por ciento de probabilidades” de que  $10^{54}$  personas existan en el futuro, entonces “el valor esperado de reducir el riesgo existencial en apenas *una mil millonésima parte de una mil millonésima parte de un punto porcentual* vale 100 mil millones de veces más que mil millones de vidas humanas”. Tal fanatismo, una palabra que algunos largoplacistas [abrazan](#), ha llevado a un número creciente de críticos a preocuparse por lo que podría suceder si los líderes políticos en el mundo real tomaran en serio la opinión de Bostrom. Para citar al estadístico matemático Olle Häggström, quien, desconcertantemente, tiende a hablar favorablemente del largoplacismo:

Me siento extremadamente incómodo con la perspectiva de que [los cálculos anteriores] puedan ser reconocidos entre los políticos y los responsables de la toma de decisiones como una guía para la política que vale la pena aceptar literalmente. Simplemente recuerda demasiado al viejo dicho “Si quieres hacer una tortilla, debes estar dispuesto a romper algunos huevos”, que generalmente se ha utilizado para explicar que un poco de genocidio podría ser algo bueno, si puede contribuir al objetivo de crear una utopía futura. Imagine una situación en la que el jefe de la CIA le explique al presidente de los Estados Unidos que tienen pruebas creíbles de que en algún lugar de Alemania, hay un lunático que está trabajando en un arma del fin del mundo y tiene la intención de usarla para acabar con la humanidad, y que este lunático tiene una posibilidad en un millón de tener éxito. No tienen más información sobre la identidad o el paradero de este lunático. Si el presidente ha tomado en serio el argumento de Bostrom, y si sabe cómo hacer la aritmética, puede concluir que vale la pena llevar a cabo un asalto nuclear a gran escala contra Alemania para matar a todas

las personas dentro de sus fronteras.

Oxford University Press *Here Be Dragons: Science, Technology and the Future of Humanity*

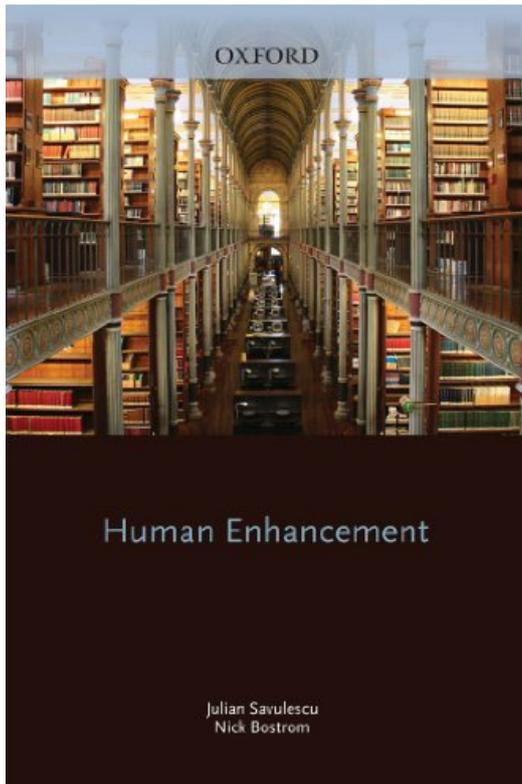
He aquí, entonces, algunas razones por las que encuentro que el largoplacismo es profundamente peligroso. Sin embargo, hay problemas adicionales y fundamentales con esta visión del mundo que nadie, que yo sepa, ha notado previamente por escrito. Por ejemplo, se puede argumentar que los compromisos subyacentes del largoplacismo son una de las principales razones por las que la humanidad se enfrenta a tantos riesgos sin precedentes para su supervivencia. En otras palabras, el largoplacismo podría ser incompatible con la consecución de la “seguridad existencial”, lo que significa que la única forma de reducir genuinamente la probabilidad de extinción o colapso en el futuro podría ser abandonar por completo la ideología largoplacista.

**Para Bostrom y Ord, no convertirnos en posthumanos nos impediría realizar nuestro vasto y glorioso potencial.**

Para entender el argumento, primero vamos a desentrañar lo que los largoplacistas entienden por nuestro “potencial a largo plazo”, una expresión que hasta ahora he usado sin definir. Podemos analizar este concepto en tres componentes principales: transhumanismo, expansionismo espacial y una visión moral estrechamente asociada con lo que los filósofos llaman “utilitarismo total”.

El primero se refiere a la idea de que deberíamos usar tecnologías avanzadas para rediseñar nuestros cuerpos y cerebros y de ese modo crear una raza “superior” de posthumanos radicalmente mejorados (que, confusamente, los largoplacistas colocan dentro de la categoría de “humanidad”). Aunque Bostrom es quizás el transhumanista más prominente de la actualidad, los largoplacistas han evitado usar el término “transhumanismo”, probablemente debido a sus asociaciones negativas. Susan Levin, por ejemplo, señala que el transhumanismo contemporáneo tiene sus raíces en el movimiento eugenésico angloamericano, y transhumanistas como Julian Savulescu, quien coeditó el libro *Human Enhancement* (2009) con Bostrom, han argumentado literalmente a favor del consumo de productos químicos “que aumentan la moralidad” como la oxitocina para evitar una catástrofe existencial (que él llama “daño final”). Como Savulescu escribe con un colega, “es una cuestión de tanta urgencia mejorar moralmente a la humanidad ... que

deberíamos buscar *cualquier* medio que haya para lograrlo”. Tales afirmaciones no solo son controvertidas, sino para muchos bastante inquietantes, y por lo tanto los largoplacistas han intentado distanciarse de tales ideas, sin dejar de defender la ideología.



El transhumanismo **afirma** que hay varios “modos posthumanos de ser” que son mucho mejores que nuestro modo humano actual. Podríamos, por ejemplo, alterarnos genéticamente para obtener un control perfecto sobre nuestras emociones, o acceder a Internet a través de implantes neuronales o, tal vez, incluso cargar nuestras mentes en el *hardware* del ordenador para lograr la “inmortalidad digital”. Como Ord insta en *El precipicio*, piense en lo increíble que sería percibir el mundo a través de la ecolocalización, como murciélagos y delfines, o magnetorrecepción, como zorros rojos y palomas mensajeras. “Tales experiencias inexploradas”, escribe Ord, “existen en mentes mucho menos sofisticadas que las nuestras. ¿Qué experiencias, posiblemente de inmenso valor, podrían ser accesibles, entonces, a mentes mucho mayores? La exploración más fantástica de Bostrom de estas posibilidades proviene de su **evocadora** “Carta desde la utopía” (2008), que representa un mundo tecno-utópico lleno de posthumanos superinteligentes inundados de tanto “placer” que, como escribe el posthumano ficticio de la carta, “lo rociamos en

nuestro té. ‘

La conexión con el largoplacismo es que, según Bostrom y Ord, no convertirnos en posthumanos aparentemente nos impediría realizar nuestro vasto y glorioso potencial, lo que sería existencialmente catastrófico. Como dijo Bostrom en 2012, “la exclusión permanente de cualquier posibilidad de este tipo de cambio transformador de la naturaleza biológica humana puede constituir en sí misma una catástrofe existencial”. Del mismo modo, Ord afirma que “preservar para siempre a la humanidad tal como es ahora también puede desperdiciar nuestro legado, renunciando a la mayor parte de nuestro potencial”.

El segundo componente de nuestro potencial, el expansionismo espacial, se refiere a la idea de que debemos colonizar la mayor cantidad posible de nuestro cono de luz futuro: es decir, la región del espacio-tiempo que es teóricamente accesible para nosotros. Según los largoplacistas, nuestro futuro cono de luz contiene una enorme cantidad de recursos explotables, a los que se refieren como nuestra “dotación cósmica” de negentropía (o entropía inversa). Solo la Vía Láctea, escribe Ord, tiene “150.000 años luz de diámetro, abarcando más de 100.000 millones de estrellas, la mayoría con sus propios planetas”. Alcanzar el potencial a largo plazo de la humanidad, continúa, “solo requiere que terminemos viajando a una estrella cercana y establezcamos un punto de apoyo suficiente para crear una nueva sociedad floreciente desde la cual podamos aventurarnos más lejos”. Al extendernos “solo seis años luz a la vez”, nuestros descendientes posthumanos podrían alcanzar “casi todas las estrellas de nuestra galaxia ...” ya que “cada sistema estelar, incluido el nuestro, necesitaría asentarse solo en las pocas estrellas más cercanas [para que] al final toda la galaxia se llenara de vida”. El proceso podría ser exponencial, dando lugar a sociedades cada vez más “florecientes” con cada segundo adicional que nuestros descendientes salten de estrella en estrella.

Pero ¿por qué exactamente querríamos hacer esto? ¿Qué tiene de importante inundar el Universo con nuevas civilizaciones posthumanas? Esto lleva al tercer componente: el utilitarismo total, al que me referiré como “utilitarismo” para abreviar. Aunque algunos largoplacistas insisten en que no son utilitaristas, debemos notar de inmediato que esto es principalmente una cortina de humo y espejos para desviar las críticas de que el longplacismo -y, más en general, el movimiento de altruismo efectivo (AE) del que surgió, no es más que utilitarismo **con un envoltorio nuevo**. El hecho es que el movimiento AE es profundamente utilitario, al menos en la práctica, y de hecho, antes de decidir un nombre, los primeros miembros del movimiento, incluido Ord, consideraron seriamente llamarlo la “comunidad utilitaria efectiva”.

Dicho esto, el utilitarismo es una teoría ética que especifica que nuestra única obligación moral es maximizar la cantidad total de “valor intrínseco” en el mundo, tal como se cuenta desde un punto de vista cósmico, imparcial e incorpóreo llamado “el punto de vista del Universo”. Desde este punto de vista, no importa cómo el valor, que los hedonistas utilitarios equiparan con el placer, se distribuya entre las personas a través del espacio y el tiempo. Todo lo que importa es la suma neta total. Por ejemplo, imagine que hay 1 billón de personas que tienen vidas de valor ‘1’, lo que significa que apenas vale la pena vivirlas. Esto da un valor total de 1 billón. Ahora considere un universo alternativo en el que mil millones de personas tienen vidas con un valor de ‘999’, lo que significa que sus vidas son extremadamente buenas. Esto da un valor total de 999 mil millones. Dado que 999 mil millones es menos de 1 billón, el primer mundo lleno de vidas que apenas vale la pena vivir sería moralmente mejor que el segundo mundo, y por lo tanto, si un utilitarista se viera obligado a elegir entre estos, elegiría el primero. (Esto se llama la “conclusión repugnante”, que los largoplacistas como Ord, MacAskill y Greaves recientemente [han argumentado](#) no debe tomarse muy en serio. ¡Para ellos, el primer mundo realmente podría ser mejor!)

### **Beckstead argumentó que deberíamos priorizar las vidas de las personas en los países ricos sobre las de los países pobres.**

El razonamiento subyacente aquí se basa en la idea de que las personas, tú y yo, no somos más que medios para un fin. No importamos en nosotros mismos; No tenemos ningún valor inherente propio. En cambio, las personas se entienden como los “contenedores” de valor, y por lo tanto importamos sólo en la medida en que “contenemos” valor, y por lo tanto contribuimos a la cantidad neta total de valor en el Universo entre el Big Bang y la muerte térmica. Dado que el utilitarismo nos dice que maximicemos el valor, se deduce que cuantas más personas (contenedores de valor) existan con cantidades netas positivas de valor (placer), mejor será el Universo, moralmente hablando. En una frase: las personas existen en aras de maximizar el valor, en lugar de que el valor exista en aras de beneficiar a las personas.

Esta es la razón por la que los largoplacistas están obsesionados con calcular cuántas personas podrían existir en el futuro si colonizáramos el espacio y creáramos vastas simulaciones por computadora alrededor de estrellas en las que un número insondablemente grande de personas vive vidas netas positivas en entornos de realidad virtual. Ya mencioné la estimación de Bostrom de  $10^{54}$  personas futuras, que incluye a muchas de estas “personas digitales”, pero en su [éxito de ventas](#) *Superintelligence* (2014) pone el número aún más alto en  $10^{58}$  personas, casi todas las cuales “vivirían vidas ricas y

felices mientras interactúan entre sí en entornos virtuales”. Greaves y MacAskill están igualmente entusiasmados con esta posibilidad, estimando que unos  $10^{45}$  seres conscientes en simulaciones por computadora podrían existir solo dentro de la Vía Láctea.

En eso consiste nuestro “vasto y glorioso” potencial: un número masivo de posthumanos digitales tecnológicamente mejorados dentro de enormes simulaciones por computadora repartidas por todo nuestro futuro cono de luz. Es por este objetivo que, en el escenario de Häggström, un político largoplacista aniquilaría Alemania. Es por este objetivo por lo que no debemos “malgastar...” nuestros recursos en cosas como resolver la pobreza mundial. Es para este objetivo que deberíamos considerar implementar un sistema de vigilancia global, mantener la guerra preventiva sobre la mesa y centrarnos más en máquinas superinteligentes que en salvar a las personas en el Sur Global de los efectos devastadores del cambio climático (principalmente causados por el Norte Global). De hecho, Beckstead incluso ha argumentado que, en aras de lograr este objetivo, deberíamos priorizar las vidas de las personas en los países ricos sobre las de los países pobres, ya que influir en el futuro a largo plazo es de “importancia abrumadora”, y es más probable que los primeros influyan en el futuro a largo plazo que los segundos. Para citar un pasaje de la tesis doctoral de Beckstead de 2013, que Ord elogia con entusiasmo como una de las contribuciones más importantes a la literatura a largo plazo:

Salvar vidas en los países pobres puede tener un efecto dominó significativamente menor que salvar y mejorar vidas en los países ricos. ¿Por qué? Los países más ricos tienen sustancialmente más innovación, y sus trabajadores son mucho más productivos económicamente. [En consecuencia,] ahora me parece más plausible que salvar una vida en un país rico es sustancialmente más importante que salvar una vida en un país pobre, en igualdad de condiciones.

[Beckstad: On the overwhelming importance of shaping the far future](#)

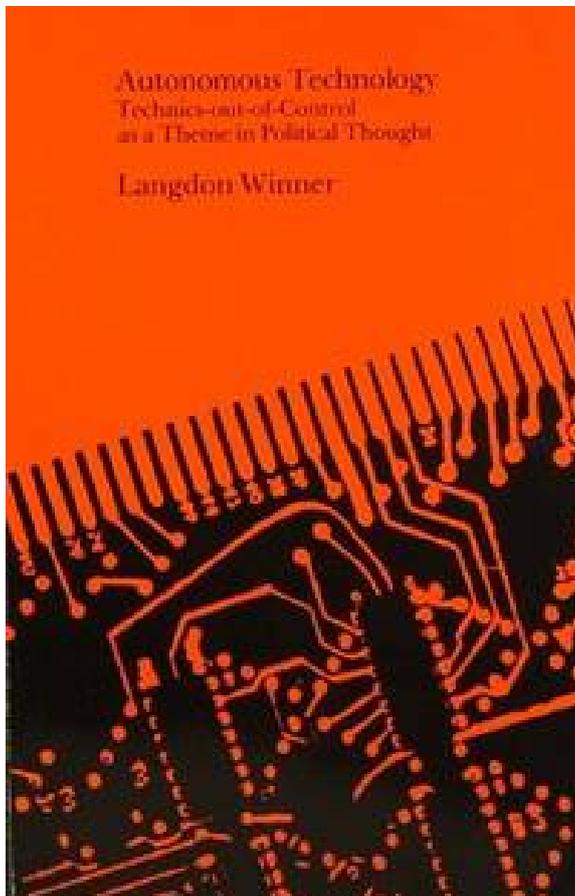
Esto es solo la punta del iceberg. Considere las implicaciones de esta concepción de “nuestro potencial” para el desarrollo de la tecnología y la creación de nuevos riesgos. Dado

que darnos cuenta de nuestro potencial es el objetivo moral final para la humanidad, y dado que nuestros descendientes no pueden convertirse en posthumanos, colonizar el espacio y crear  $\sim 10^{58}$  personas en simulaciones por computadora sin tecnologías mucho más avanzadas que las actuales, no desarrollar más tecnología constituiría en sí mismo una catástrofe existencial, un modo de fracaso (comparable a Ramsey descuidando sus talentos al pasar sus días jugando al billar y bebiendo) que Bostrom llama “estancamiento”. De hecho, Bostrom coloca esta idea al frente y en el centro de su definición canónica de “riesgo existencial”, que denota cualquier evento futuro que impida que la humanidad alcance y/o mantenga un estado de “madurez tecnológica”, es decir, “el logro de capacidades que permitan un nivel de productividad económica y control sobre la naturaleza cercano al máximo que pudiera lograrse de manera factible”. La madurez tecnológica es el eje aquí porque controlar la naturaleza y aumentar la productividad económica hasta los límites físicos absolutos son ostensiblemente necesarios para crear la máxima cantidad de “valor” dentro de nuestro futuro cono de luz.

Pero reflexione por un momento sobre cómo la humanidad se metió en la actual crisis climática y ecológica. Detrás de la extracción y quema de combustibles fósiles, la destrucción de los ecosistemas y el exterminio de especies ha estado la noción de que la naturaleza es algo que debe ser controlado, subyugado, explotado, vencido, saqueado, transformado, reconfigurado y manipulado. Como [escribe](#) el teórico de la tecnología Langdon Winner en *Autonomous Technology* (1977), desde la época de Francis Bacon nuestra visión de la tecnología ha estado “inextricablemente ligada a una sola concepción de la manera en que se usa el poder: el estilo de dominio absoluto, el control despótico y unidireccional del amo sobre el esclavo”. Y añade:

Rara vez hay reservas sobre el papel legítimo del ser humano en conquistar, vencer y subyugar todo lo natural. Este es su poder y su gloria. Lo que en otras situaciones parecerían intenciones bastante groseras y despreciables son aquí la más honorable de las virtudes. La naturaleza es la presa universal, para manipular como los humanos crean conveniente.

[Langdon Winner Autonomous Technology](#)



Esto es precisamente lo que encontramos en el relato de Bostrom sobre los riesgos existenciales y su futurología normativa asociada: la naturaleza, todo el Universo, nuestra “dotación cósmica” está ahí para el saqueo, para ser manipulada, transformada y convertida en “estructuras de valores, como seres sintientes que viven vidas que valen la pena” en vastas simulaciones por computadora, citando el ensayo de Bostrom “Desechos astronómicos” (2003). Sin embargo, esta visión baconiana y capitalista es una de las causas fundamentales de la crisis ambiental sin precedentes que ahora amenaza con destruir grandes regiones de la biosfera, las comunidades indígenas de todo el mundo y tal vez incluso la propia civilización tecnológica occidental. Mientras que otros largoplacistas no han sido tan explícitos como Bostrom, hay una clara tendencia a ver el mundo natural de la manera en que el utilitarismo ve a las personas: como medios para algún fin abstracto, impersonal, y nada más. MacAskill y un colega, por ejemplo, [escriben](#) que el movimiento EA, y por implicación el largoplacismo, es “tentativamente *bienestarista* en el sentido de que su objetivo tentativo de hacer el bien se refiere a promover solo el bienestar y no, por ejemplo,

proteger la biodiversidad o conservar la belleza natural por su propio bien”.

**En este sentido, cada problema surge de muy poca en lugar de demasiada tecnología.**

Igual de preocupante es la demanda a largo plazo de que debemos crear tecnologías cada vez más poderosas, a pesar del hecho acordado de que la abrumadora fuente de riesgo para la existencia humana en estos días proviene de estas mismas tecnologías. En palabras de Ord, “sin esfuerzos serios para proteger a la humanidad, hay fuertes razones para creer que el riesgo será mayor este siglo, y aumentará con cada siglo que el progreso tecnológico continúa”. Del mismo modo, en 2012 Bostrom reconoce que

La mayor parte del riesgo existencial en el futuro previsible consiste en riesgos existenciales antropogénicos, es decir, derivados de la actividad humana. En particular, la mayoría de los mayores riesgos existenciales parecen estar relacionados con posibles avances tecnológicos futuros que pueden expandir radicalmente nuestra capacidad de manipular el mundo externo o nuestra propia biología. A medida que nuestros poderes se expandan, también lo hará la escala de sus posibles consecuencias: intencionales y no intencionadas, positivas y negativas.

Desde este punto de vista, solo hay un camino a seguir, más desarrollo tecnológico, incluso si este es el camino más peligroso hacia el futuro. Pero, ¿cuánto sentido tiene esto? Sin duda, si queremos maximizar nuestras posibilidades de supervivencia, deberíamos oponernos al desarrollo de nuevas tecnologías peligrosas de doble uso. Si más tecnología equivale a un mayor riesgo, como muestra claramente la historia y afirman las proyecciones tecnológicas, entonces tal vez la única forma de alcanzar realmente un estado de “seguridad existencial” es ralentizar o detener por completo la innovación tecnológica.

Pero los largoplacistas tienen una respuesta a este enigma: la llamada “tesis de la neutralidad del valor”. Esto afirma que la tecnología es un objeto moralmente neutral, es decir, “solo una herramienta”. La idea está más famosamente encapsulada en el eslogan de la NRA “Las armas no matan a la gente, la gente mata a la gente”, que transmite el mensaje de que las consecuencias de la tecnología, ya sean buenas o malas, beneficiosas o dañinas, están totalmente determinadas por los usuarios, no por los artefactos. Como dijo Bostrom en 2002, “no debemos *culpar* a la civilización o la tecnología por imponer grandes riesgos existenciales”, y agregó que “debido a la forma en que hemos definido los riesgos existenciales, un fracaso en el desarrollo de la civilización tecnológica implicaría que hemos

sidó víctimas de un desastre existencial”.

Ord argumenta de manera similar que “el problema no es tanto un exceso de tecnología como una falta de sabiduría”, antes de citar el libro de Carl Sagan *Pale Blue Dot* (1994): “Muchos de los peligros que enfrentamos surgen de la ciencia y la tecnología, pero, más fundamentalmente, porque nos hemos vuelto poderosos sin llegar a ser proporcionalmente sabios”. En otras palabras, es nuestra culpa por no ser más inteligentes, más sabios y más éticos, un conjunto de deficiencias que muchos largoplacistas creen, en un poco de lógica retorcida, podrían rectificarse mediante la reingeniería tecnológica de nuestros sistemas cognitivos y disposiciones morales. Todo, por esta razón, es un problema de ingeniería, y por lo tanto cada problema surge de muy poca en lugar de demasiada tecnología.

Ahora podemos comenzar a ver cómo el largo plazo podría ser contraproducente. Su énfasis “fanático” en el cumplimiento de nuestro potencial a largo plazo no solo podría llevar a las personas a, por ejemplo, descuidar el cambio climático no existencial, priorizar a los ricos sobre los pobres y tal vez incluso “justificar” la violencia preventiva y las atrocidades para el “mayor bien cósmico”, sino que también contiene las mismas tendencias -baconianismo, capitalismo y neutralidad de valores- que han alejado a la humanidad del precipicio de la destrucción. El largo plazo nos dice que maximicemos la productividad económica, nuestro control sobre la naturaleza, nuestra presencia en el Universo, el número de personas (simuladas) que existirán en el futuro, la cantidad total de “valor” impersonal y así sucesivamente. Pero para maximizar, debemos desarrollar tecnologías cada vez más poderosas y peligrosas; No hacerlo sería en sí mismo una catástrofe existencial. Sin embargo, no hay que preocuparse, porque la tecnología no es responsable de nuestro empeoramiento de la situación y, por lo tanto, el hecho de que la mayoría de los riesgos se derivan directamente de la tecnología no es razón para dejar de crear más tecnología. Más bien, el problema radica en nosotros, lo que significa solo que debemos crear aún más tecnología para transformarnos en posthumanos cognitiva y moralmente mejorados.

Esto parece una receta para el desastre. Crear una nueva raza de posthumanos “sabios y responsables” es inverosímil y, si las tecnologías avanzadas continúan desarrollándose al ritmo actual, una catástrofe a escala global es casi seguramente una cuestión de cuándo en lugar de si. Sí, necesitaremos tecnologías avanzadas si deseamos escapar de la Tierra antes de que sea esterilizada por el Sol en mil millones de años más o menos. Pero el hecho crucial que los largoplacistas pasan por alto es que es *mucho más probable que la tecnología cause nuestra extinción antes de este evento futuro lejano que salvarnos de él*. Si usted, como yo, valora la supervivencia continua y el florecimiento de la humanidad, debería

preocuparse por el largo plazo, pero rechazar la ideología del largoplacismo plazo, que no solo es peligrosa y sesgada, sino que podría estar contribuyendo y reforzando los riesgos que ahora amenazan a todas las personas en el planeta.

### Para citar esta entrada

Torres, Emil P. (2023) Primeros embriones sintéticos: el avance científico plantea serias cuestiones éticas.

En *Niaia*, <https://niaia.es/primeros-embriones-sinteticos-el-avance-cuetifico-plantea-serias-cuestiones-eticas/>

Publicado inicialmente en *Aeon*. **2021** Publicado en *Niaia* con la autorización de la revista *Aeon*. Traducción automática de Word, revisada por Félix García Moriyón

Si lo desea, puede volver a publicar este artículo, en forma impresa o digital. Pero le pedimos que cumpla estas instrucciones: por favor, no edite la pieza, asegúrese de que se la atribuye a su autor, a su institución de referencia (universidad o centro de investigación), y mencione que el artículo fue publicado originalmente en *The Conversation* y *Niaia*.

