ADMINISTRACIÓN, INTELIGENCIA ARTIFICIAL Y RIESGO EXISTENCIAL: EL PAPEL DE LAS CORRIENTES CRÍTICAS PARA EL FUTURO DE LA HUMANIDAD

ELKIN FABRIANY PINEDA-HENAO[23]
JOSE LONDOÑO-CARDOZO[24]

[23] Estudiante de Doctorado en Ciencias de la Administración en la Universidad Nacional Autónoma de México. Magíster en Filosofía y Licenciado en Filosofía de la Universidad del Valle. Magíster en Administración y Administrador de Empresas de la Universidad Nacional de Colombia. Grupos de investigación: *Episteme: Filosofía y Ciencia* (Universidad del Valle), y Grupo de Estudios Neoinstitucionales GEN (Universidad Nacional de Colombia). Miembro de la Red de Estudios Organizacionales Colombiana (REOC). Correo: efpinedah@comunidad.unam.mx Orcid: https://orcid.org/0000-0002-0168-1739.

[24] Magister en Administración y Administrador de empresas, integrante del Grupo de Estudios Neoinstitucionales - GEN de la Universidad Nacional de Colombia. Profesor tiempo completo y líder de investigación del programa de contaduría púbica. Integrante del Grupo de investigación en Ciencias Administrativas, Económicas y Financieras – GICAEF de la Corporación Universitaria Minuto de Dios – UNIMINUTO. Líder de la Red de profesores investigadores de Palmira (REPIPA). Correo: jodlondonoca@unal.edu.co Orcid https://orcid.org/0000-0002-5739-1191

Resumen

El reciente auge en inteligencia artificial (IA) ha desencadenado que se retomen debates sobre riesgos existenciales. Sin embargo, en el contexto de la administración y organizaciones, esta discusión ha estado notablemente ausente, pese a que existe un gran riesgo de que el estilo de administración tradicional pueda influir en riesgos asociados con la producción e implementación de IA en y desde las organizaciones, debido a la prioridad que este da a los intereses económicos y de eficiencia y productividad. Por tal motivo, el objetivo del presente trabajo es el de argumentar cómo las corrientes críticas de la administración y de la organización, como los Estudios Críticos de la Administración (ECA), los Estudios Críticos Organizacionales (ECO) y la Gestión Humanística Radical (GHR), pueden desempeñar un papel fundamental en comprender críticamente y abordar alternativas ante estos riesgos. En síntesis, estas corrientes críticas pueden contribuir a través de su incidencia en la formulación de políticas públicas, cambios en la formación en administración y programas de investigación específicos para prevenir riesgos existenciales.

Palabras clave

Estudios organizacionales, Estudios críticos de la administración, Gestión humanista radical, Transhumanismo, Posthumanismo, Riesgo existencial de organizaciones, Riesgo existencial de empresas.

Abstract

The recent surge in Artificial Intelligence (AI) has reignited debates on existential risks. However, in the context of management and organizations, this discussion has been notably absent, despite the significant risk that the traditional management style could influence risks associated with the production and implementation of AI within and by organizations, due to its prioritization of economic interests, efficiency, and productivity. Therefore, the aim of this paper is to argue how critical management and organizational theories, such as Critical Management Studies (CMS), Critical Organizational Studies (COS), and Radical Humanistic Management (RHM), can play a pivotal role in critically understanding and addressing alternatives to these risks. In summary, these critical theories can contribute by influencing the formulation of public policies, changes in management education, and specific research programs to prevent existential risks.

Keywords

Organization studies, Critical management studies, Radical humanistic management, Transhumanism, Posthumanism, Existential risk of organizations, Existential risk of companies.

El reciente surgimiento de nuevos modelos de inteligencia artificial (en adelante, IA) ha tenido un profundo impacto en la sociedad, lo que ha llevado a un amplio debate sobre sus riesgos. Este debate ha dado lugar a la consideración de diversas contribuciones en el análisis multidisciplinario y filosófico, incluyendo la discusión del riesgo existencial asociado a la AI. Este enfoque examina los posibles riesgos de la IA que podrían conducir a un colapso o incluso a la extinción de la humanidad. Sin embargo, a pesar de su relevancia para el futuro de la humanidad, y como se argumentará en este documento, la discusión de los riesgos existenciales de la IA en el contexto de la administración y las organizaciones, así como sus corrientes críticas, ha estado notoriamente ausente.

El presente documento se deriva de reflexiones posteriores a trabajos previos (Pineda-Henao, 2022), donde los riesgos existenciales relacionados con la administración y las organizaciones se identificaron como temas que requieren una atención urgente. Estos riesgos se analizan desde perspectivas de corrientes de pensamiento crítico-social que exploran el impacto del transhumanismo y el posthumanismo en la administración y las organizaciones. La importancia de abordar estos riesgos radica en el aumento en la producción y aplicación de IA en las organizaciones, lo que podría dar lugar a riesgos existenciales, especialmente en un contexto donde el enfoque tradicional de la administración, que prioriza valores productivos y económicos sobre la dignidad humana y el bienestar social y ambiental, podría ser vulnerable ante estos riesgos.

El objetivo principal de este escrito es argumentar cómo las corrientes críticas en el campo de la administración y las organizaciones desempeñan un papel fundamental en la crítica y la propuesta de alternativas al estilo de administración predominante, con el fin de prevenir y mitigar los posibles riesgos existenciales asociados con la producción y aplicación de AI. Dado que existe una falta de investigaciones que aborden los riesgos existenciales relacionados con

la administración y las organizaciones, este documento también busca persuadir sobre la importancia de abordar esta cuestión, destacando las conexiones críticas entre ambos campos de estudio, especialmente desde las perspectivas de corrientes como los Estudios Críticos de la Administración (en adelante ECA), los Estudios Críticos Organizacionales (ECO) y la Gestión Humanista Radical (GHR).

Desde el punto de vista metodológico, se emplea una síntesis cualitativa (Seers, 2012; Thomas & Harden, 2008) de fuentes documentales en dos áreas temáticas: 1) Los riesgos existenciales relacionados con la IA y 2) Las contribuciones de las corrientes críticas en el campo de la administración y las organizaciones. A partir de estas fuentes documentales, se realiza una revisión documental exploratoria que involucra un proceso de análisis crítico, argumentación e interpretación hermenéutica. Por lo tanto, la estructura del documento se organiza de la siguiente manera: en primer lugar, se presenta un marco teórico que establece los supuestos relacionados con el riesgo existencial y la AI. A continuación, en el segundo apartado, se aborda la discusión, que se divide en dos partes: a) La consideración de la IA como un riesgo existencial importante, conectado a la administración y las organizaciones a través de la producción y aplicación de AI; b) La argumentación sobre la contribución de las corrientes críticas en la administración y las organizaciones para prevenir y mitigar los riesgos existenciales asociados con la producción y aplicación de AI, a través de la crítica al enfoque tradicional de la administración.

Marco Teórico

Para abordar esta reflexión de manera adecuada, resulta imperativo establecer una comprensión sólida de varios conceptos clave. En esta sección, se procederá a detallar la base teórica necesaria para una apreciación cabal del tema en cuestión. En primer lugar, se esbozarán las características fundamentales del riesgo existencial, explorando sus diversas posturas, fundamentos y críticas. En un segundo plano, se

analizará el surgimiento de la inteligencia artificial como una innovadora herramienta tecnológica que complementa las capacidades humanas. Finalmente, en el tercer segmento, se examinará el papel de la inteligencia artificial en el marco de los riesgos existenciales contemporáneos.

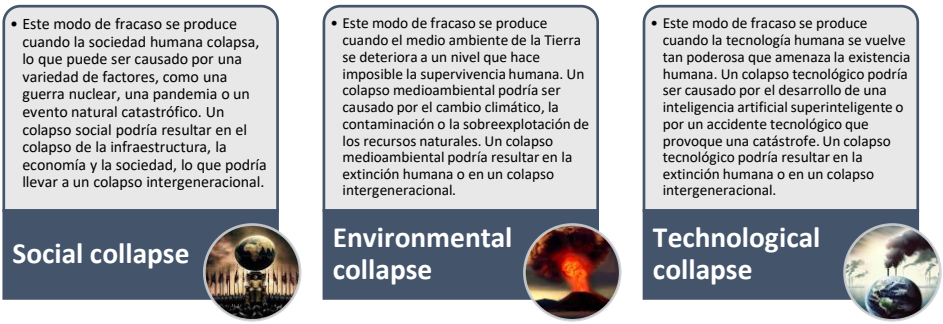
El Riesgo Existencial

El riesgo existencial es un concepto que se refiere a los riesgos que poseen el potencial de amenazar el futuro de la humanidad en su totalidad (Bostrom, 2002). Estos riesgos son tan grandes que, incluso si las probabilidades de que ocurran son bajas, las consecuencias serían catastróficas (Bostrom, 2017). El riesgo existencial no se limita únicamente al riesgo de extinción humana, sino que abarca otros tres modos de fracaso que podrían resultar en un colapso intergeneracional y ocasionar pérdidas igualmente significativas del valor esperado (ver 8). En este sentido, es importante considerar la noción de maxipok, que desde la teorización de Bostrom, sostiene que la acción moralmente correcta es aquella que maximiza la probabilidad de evitar una catástrofe existencial.

Esta noción es una idea propuesta por Ortega y Gasset en la que afirma que el hombre es un ser maxipok. Es decir, un ser que siempre está buscando la perfección, pero que nunca la alcanza (1914). Para el filósofo español, el hombre es un ser insatisfecho por naturaleza. Siempre quiere más, siempre quiere ser mejor. Esto se debe a que el hombre es un ser racional, y la razón siempre le lleva a buscar la perfección (Ortega y Gasset, 1914).

Figura 18

Three Modes of Existential Failure



Nota. Elaboración propia con base en Bostrom (2017, p. 12)

La relevancia de la reducción de los riesgos existenciales es evidente cuando se considera la perspectiva impersonal y global de la humanidad en su conjunto. Según la noción de maxipok, se deben tomar las medidas que tengan la mayor probabilidad de evitar una catástrofe existencial, incluso si esas medidas tienen un costo significativo. En este sentido, la prevención de los riesgos existenciales se plantea como una prioridad de alcance mundial, con el propósito de salvaguardar el futuro de la humanidad (Bostrom, 2013).

Ahora bien, el riesgo existencial se define como el riesgo de extinción humana o el colapso de la civilización, lo que implica amenazar la existencia continua de la humanidad o la destrucción permanente y drástica de su potencial para un desarrollo futuro deseable (Torres, 2023). Este riesgo puede resultar de diversas fuentes, incluyendo amenazas de origen natural, como impactos de asteroides

o cometas, así como amenazas de origen humano, como las derivadas de actividades tecnológicas avanzadas.

Es importante destacar que la definición de riesgo existencial puede variar según el contexto y la audiencia a la que se dirige. Torres (2023) sugiere que la definición de riesgos existenciales como riesgos de extinción humana o colapso de la civilización es eficaz cuando se comunica con el público en general, mientras que la definición de los riesgos existenciales como una pérdida significativa del valor esperado puede ser más adecuada para establecer los estudios de riesgo existencial como un campo legítimo de investigación científica y filosófica.

Los riesgos existenciales, independientemente de su definición específica, presentan características que los hacen notoriamente distintos de los riesgos comunes. Estos riesgos poseen un valor esperado excepcionalmente alto, lo que significa que incluso una pequeña reducción en el riesgo existencial neto puede tener consecuencias enormes. Además, la gestión de riesgos existenciales se ve complicada por la falta de precedentes históricos, lo que dificulta la aplicación de métodos convencionales de gestión de riesgos (Bostrom, 2017; Kaku, 2014; Rees, 2004).

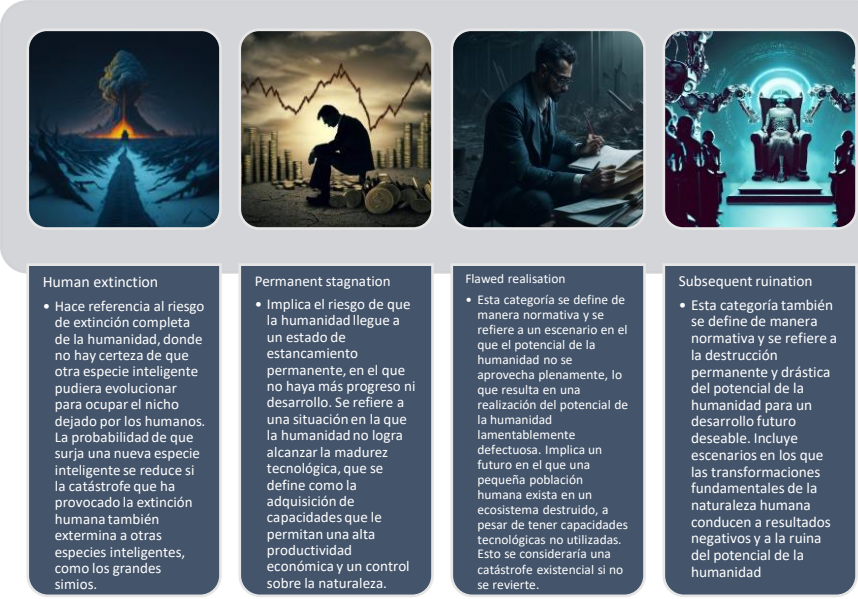
Los riesgos existenciales se pueden clasificar en cuatro categorías principales (ver Figura): a) Extinción humana, b) Estancamiento permanente, c) Realización defectuosa y d) Ruina posterior (Bostrom, 2013). En cada una de estas categorías, los principales riesgos se derivan de las actividades humanas. Por ejemplo, las amenazas tecnológicas avanzadas, como la inteligencia artificial, la biotecnología y la nanotecnología, pueden generar riesgos existenciales significativos.

El término riesgo existencial también se relaciona con la noción de riesgo catastrófico global, aunque no son necesariamente equivalentes (Bostrom & Cirkovic, 2008). Mientras que el riesgo existencial se centra en amenazas que podrían llevar a la extinción prematura de la vida inteligente en la Tierra o la destrucción permanente y drástica de su potencial para un futuro desarrollo deseable, el riesgo catastrófico

global se refiere a amenazas que podrían causar daños catastróficos a nivel global sin necesariamente llevar a la extinción de la humanidad (Bostrom & Cirkovic, 2008; Russell & Norvig, 2010).

Figura 19

Categorization of Existential Risks According to Bostrom



Nota. Elaboración propia con base en Bostrom (2013, p. 19)

La importancia de abordar los riesgos existenciales es tan significativa que se debate su tratamiento prioritario y su inclusión en las agendas y políticas públicas a nivel nacional e internacional. Organismos internacionales como las Naciones Unidas y los gobiernos de diversos países consideran la necesidad de abordar los riesgos existenciales como parte de sus esfuerzos por garantizar la seguridad y el bienestar a largo plazo de la humanidad (Boyd & Wilson, 2020). A

pesar de la importancia en el abordaje de los riesgos existenciales, principalmente a partir de la teoría de Bostrom, algunos autores critican dichas preocupaciones.

Højme destaca en su trabajo una crítica a la preocupación de Bostrom por los riesgos existenciales, destacando las contradicciones internas del pensamiento transhumanista y las premisas inválidas en las que se basa (2019). Para este autor, El transhumanismo, en su intento de superar la condición humana y alcanzar la poshumanidad, vuelve al mito y descuida la preocupación por la vida real. Esta discrepancia socava la supuesta preocupación del transhumanismo por toda la humanidad (Højme, 2019).

En general, la crítica de Højme se centra en el proceso de selección. Según él, ante los riesgos existenciales, el proceso de selección puede favorecer a ciertas personas o grupos. De igual forma, en medio de la crítica se reconoce el desafío de determinar qué valores deberían ser prioritarios para la vida inteligente originaria de la Tierra y sugiere que incluso una superinteligencia podría no ser capaz de proporcionar la respuesta (Højme, 2019). Los debates recientes sobre el riesgo existencial se han centrado en las fuentes específicas de riesgo, más que en la compleja interacción de fallos o riesgos que no pueden especificarse con claridad (Manheim, 2020). Manheim propone un análisis de la visión ampliada de los mundos vulnerables, dado que lleva a conclusiones que son diferentes o incluso contrarias a las sugeridas por Bostrom (2020).

La visión ampliada de los mundos vulnerables, propuesta por Manheim (2020), desafía la hipótesis del mundo vulnerable de Bostrom (2002). Mientras que Bostrom sostiene que hay avances tecnológicos específicos que, por defecto, conducen a la devastación o la extinción de la civilización, Manheim sostiene que la fragilidad, es decir, la propensión de un sistema a sufrir daños o fallas, que resulta de la complejidad de ciertos sistemas, puede ser una fuente inevitable de riesgo catastrófico o existencial.

Esta visión ampliada, dado que lleva a conclusiones que son diferentes o incluso contrarias a las sugeridas por Bostrom (2002, 2013, 2017), tiene importantes implicaciones para la forma en que se abordan los riesgos existenciales. Manheim sugiere que, en lugar de centrarse en identificar tecnologías específicas como las principales fuentes de riesgo, se debería hacer hincapié en el abordaje de la fragilidad sistémica (2020). Es decir, en la propensión de un sistema complejo a sufrir daños o fallas. Esto implica reconocer que todos los sistemas, incluso los más complejos, son susceptibles a fallos catastróficos (Taleb, 2016).

De todo lo anterior, es posible sintetizar que el riesgo existencial es un fenómeno de interés y debate fundamental, caracterizado por su potencial para amenazar la continuidad de la humanidad en su totalidad. Dos enfoques notables han contribuido significativamente a este debate. Por un lado, la perspectiva de Bostrom (2002, 2013, 2017) resalta la necesidad de priorizar la mitigación de riesgos existenciales, subrayando su magnitud y la importancia de tomar medidas para prevenirlos. Por otro lado, la visión de Højme (2019) y Manheim (2020) relativiza el riesgo existencial, destacando las contradicciones internas del pensamiento transhumanista y las premisas inválidas que subyacen a esta preocupación. Ambas perspectivas poseen elementos de validez y ofrecen un espectro completo de consideraciones para la comprensión y gestión del riesgo existencial.

Inteligencia Artificial

La historia evolutiva del hombre ha estado permeada por la tecnología (Pérez de Paz & Londoño-Cardozo, 2021). Ello, dado que algunos autores consideran que la tecnología es un compensador de las deficiencias evolutivas del hombre (Pérez de Paz, 2016). Desde la antigüedad, el uso de tecnología, como el fuego o piedras afiladas para cortar, fueron determinantes en el desarrollo de la civilización (Londoño-Cardozo & Pérez de Paz, 2021; Melnyk et al., 2019; Pérez de

Paz & Londoño-Cardozo, 2021). En la historia más reciente se puede rastrear el papel de la tecnología mediante las llamadas revoluciones industriales y tecnológicas.

La historia de las revoluciones industriales se remonta al siglo XVIII, con la Primera Revolución Industrial, marcada por la transición de la producción manual a la mecanizada. Inventos clave incluyen la máquina de vapor de James Watt, que impulsó la industria textil, y la hiladora mecánica de Richard Arkwright (Kemp, 1979). Esta revolución transformó la economía y la sociedad, dando lugar a la urbanización y el crecimiento industrial (Villani, 2009).

La Segunda Revolución Industrial, a finales del siglo XIX, fue impulsada por avances en la electricidad, el acero y el petróleo. Inventos notables incluyen el teléfono de Alexander Graham Bell y la bombilla de Thomas Edison (Xu et al., 2018). La Tercera Revolución Industrial, en el siglo XX, se centró en la electrónica y la automatización, con la creación de la computadora personal (Rifkin, 2011; Roel, 1998; Xu et al., 2018).

La actual Cuarta Revolución Industrial se caracteriza por la convergencia de tecnologías digitales, biotecnología y la inteligencia artificial (Garrell & Guilera, 2019; Vaidya et al., 2018; Xu et al., 2018). La IA es un elemento central, impulsando la automatización de procesos, el aprendizaje automático y la toma de decisiones inteligentes. Innovaciones como los vehículos autónomos, la robótica avanzada y la IA en la atención médica son ejemplos de cómo la IA está remodelando la economía y la sociedad en la actualidad, consolidando su papel como una fuerza transformadora en la Cuarta Revolución Industrial (Alvarado Rojas, 2015; Flechoso, 2021).

La IA se erige como un campo de estudio y desarrollo dentro de la informática que tiene como propósito la creación de agentes inteligentes, es decir, sistemas capaces de razonar, aprender y actuar de forma autónoma (Russell & Norvig, 2010), también llamados tecnologías agenciativas (Londoño-Cardozo & Pérez de Paz, 2021; Pérez de Paz et al., 2021; Pérez de Paz & Londoño-Cardozo, 2021). En

el transcurso de las últimas décadas, la IA ha experimentado un notable y acelerado avance que ha dejado una huella significativa en la sociedad, permeando diversas áreas de la vida cotidiana y la industria (Collins et al., 2021). La base fundamental de la IA reside en el aprendizaje automático, una técnica que permite a los sistemas aprender de los datos sin necesidad de una programación explícita (Jordan & Mitchell, 2015; J. B. O. Mitchell, 2014). En este sentido, el aprendizaje automático se desglosa en dos categorías principales: el aprendizaje supervisado y el aprendizaje no supervisado, cada uno con sus particularidades y aplicaciones específicas.

El aprendizaje supervisado constituye una de las ramas principales del aprendizaje automático, donde los sistemas son alimentados con un conjunto de datos previamente etiquetados. En esta configuración, cada dato del conjunto se encuentra asociado con una etiqueta que denota su categoría o clase correspondiente (T. M. Mitchell, 1997). El objetivo de un sistema de aprendizaje supervisado es aprender a relacionar las características inherentes a los datos con las etiquetas asociadas. A través de la exposición a este conjunto de datos etiquetados, la IA ajusta sus modelos y algoritmos internos para ser capaz de predecir con precisión las etiquetas de nuevos datos desconocidos. El aprendizaje supervisado es ampliamente aplicado en tareas de clasificación y regresión, desde la detección de spam en correos electrónicos hasta la clasificación de imágenes médicas o la predicción de precios de bienes raíces (Mosqueira-Rey et al., 2023).

El más reciente ejemplo de uso de una IA de aprendizaje supervisado, al momento de la escritura de este documento, es la aparición de Now and Then, la última canción lanzada en público por la banda británica The Beatles. Con el Uso de Melodyne se pudo limpiar y mejorar la calidad de la voz de John Lennon grabada en 1970. Posteriormente, otra IA de aprendizaje supervisado, llamada Deepfake, ayudó a eliminar el ruido y la distorsión, y de restaurar la voz de Lennon a su estado original.

Por otro lado, el aprendizaje no supervisado se caracteriza por la ausencia de etiquetas en el conjunto de datos utilizado para el entrenamiento de la IA. En este contexto, el sistema debe aprender a identificar patrones y estructuras inherentes a los datos sin recibir indicaciones externas. El aprendizaje no supervisado se utiliza para tareas como la agrupación de datos o la reducción de la dimensionalidad (Anderson & Anderson, 2011; T. M. Mitchell, 1997; Mosqueira-Rey et al., 2023). Por ejemplo, en la agrupación de datos, la IA puede descubrir automáticamente categorías o grupos de datos similares dentro de un conjunto, sin la necesidad de etiquetas previas que definan esas categorías. En la reducción de la dimensionalidad, la IA busca simplificar la representación de los datos sin pérdida significativa de información.

Además del aprendizaje supervisado y no supervisado, existen otros enfoques de aprendizaje que también se emplean en el ámbito de la IA, como el aprendizaje por refuerzo y el aprendizaje por evolución (Vinod, 2023). El aprendizaje por refuerzo se centra en la toma de decisiones secuenciales, donde un agente interactúa con su entorno y recibe recompensas o castigos según las acciones que realiza. A través de la retroalimentación del entorno, el agente aprende a tomar decisiones que maximicen su recompensa a lo largo del tiempo. Este enfoque se utiliza en aplicaciones como juegos, robótica y control de procesos (T. M. Mitchell, 1997). El aprendizaje por evolución, por su parte, se inspira en los principios de la evolución biológica para optimizar soluciones. Los algoritmos genéticos, que simulan la selección natural y la reproducción, se emplean para encontrar soluciones óptimas en problemas complejos y espacios de búsqueda amplios.

IA Como el Principal Riesgo Existencial de la Actualidad

Para abordar la cuestión de la inteligencia artificial (IA) como el principal riesgo existencial de la actualidad, es necesario considerar

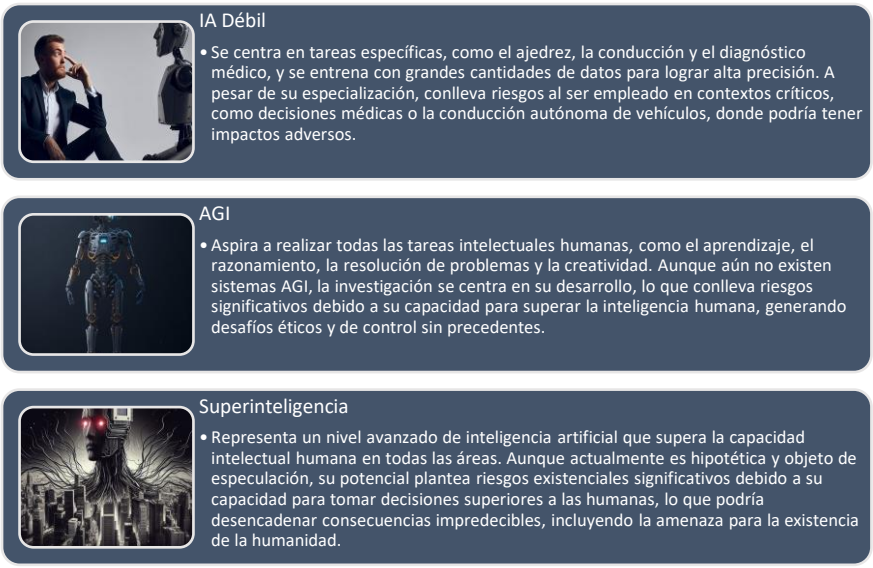
tanto su creciente relevancia como los riesgos asociados a su desarrollo y despliegue. La IA es una tecnología en constante evolución que agrega una capa adicional de complejidad a la ecuación de riesgo existencial. Bajo la óptica de Bostrom (2002, 2013, 2017) y las perspectivas aportadas por Højme (2019) y Manheim (2020), se reconoce la importancia de examinar detenidamente la seguridad y la responsabilidad en la creación y aplicación de sistemas de IA, dada su influencia innegable en la humanidad.

La creciente relevancia de la IA en este contexto añade un nivel adicional de complejidad. La IA es una tecnología en constante evolución que plantea desafíos significativos en términos de riesgo existencial. Las perspectivas de Bostrom y de de Højme (2019) y Manheim (2020) se tornan especialmente pertinentes en el contexto de la IA. La seguridad y la responsabilidad en el desarrollo y la implementación de sistemas de IA se convierten en cuestiones cruciales, dado el impacto innegable de esta tecnología en la humanidad.

Para comprender mejor los riesgos existenciales asociados a la inteligencia artificial, es esencial dividir la IA en tres categorías principales: a) Inteligencia Artificial Estrecha o IA Débil (Russell & Norvig, 2010), b) Inteligencia Artificial General o AGI (Kurzweil, 2014) y c) Inteligencia Artificial Fuerte o Superinteligencia (Bostrom, 2017). Estas categorías podrían suponer un orden evolutivo de los tipos de inteligencia artificial donde cada una tiene sus propias implicaciones y características (ver Figura 20).

Figura 20

La evolución de la inteligencia artificial



Nota. Elaboración propia con base en Bostrom (2017), Kurzweil (2014) y Russell & Norvig (2010)

A pesar de la complejidad intrínseca al riesgo existencial y la dificultad en su cuantificación precisa, existen medidas concretas que pueden adoptarse para mitigar su impacto. Estas acciones pueden incluir el desarrollo de tecnologías seguras y responsables, la promoción de la cooperación internacional en la gestión de riesgos existenciales y la educación pública sobre los desafíos y las implicaciones asociadas a estos riesgos. La investigación interdisciplinaria y el diálogo se erigen como componentes fundamentales para abordar eficazmente estos riesgos y buscar soluciones efectivas. La colaboración entre gobiernos, organizaciones y

la sociedad civil se convierte en una pieza fundamental en el enfrentamiento de este desafío global.

No obstante, el rápido avance de la IA no está exento de riesgos y desafíos, algunos de los cuales son potenciales, mientras que otros ya se han materializado en la realidad. Uno de los riesgos más graves asociados a la IA es el riesgo existencial, que plantea la posibilidad de que la IA alcance un nivel de poder que la coloque en posición de amenazar la supervivencia de la humanidad. Esta amenaza es especulativa, pero ha generado una creciente preocupación en la comunidad científica y ética debido a la posibilidad de que, en un hipotético conflicto de intereses, la IA tome acciones que pongan en peligro la existencia de la especie humana.

Además del riesgo existencial, existen riesgos concretos y actuales relacionados con la IA. Entre ellos, el prejuicio en los sistemas de IA plantea preocupaciones significativas, ya que la IA puede heredar sesgos y prejuicios presentes en los datos con los que se entrena. Esto puede llevar a situaciones en las que los sistemas de IA reflejen involuntariamente los prejuicios de sus creadores o de la sociedad en general, lo que resulta en discriminación y desigualdad. Asimismo, la utilización de la IA en la creación de armas autónomas representa un riesgo considerable, ya que la automatización y autonomía de estas armas pueden desencadenar conflictos letales sin intervención humana directa. Finalmente, la pérdida de control, medida por la dificultad para comprender y supervisar el funcionamiento de sistemas de IA altamente complejos, plantea la preocupación de que la IA pueda tomar decisiones críticas sin una supervisión adecuada, lo que podría desencadenar consecuencias imprevistas y potencialmente perjudiciales en diversas esferas de la sociedad.

La IA como Riesgo Existencial Asociado a la Administración y las Organizaciones

De manera exploratoria, se puede destacar la fuerte ausencia de estudios sobre riesgo existencial desde la disciplina de la administración o los estudios organizacionales. Esto, pese a que la mayor parte de la investigación sobre riesgo existencial suele ser multidisciplinar e involucra distintos actores sociales. No obstante, se pueden deducir teóricamente algunos vínculos entre el riesgo existencial y este campo de estudios. Un ejemplo es el trabajo de Iglesias-Márquez (2020), quien si bien no aborda el tema del riesgo existencial en su trabajo, sí asociado el fenómeno del cambio climático en conexión con la producción y el alto consumo de energía por parte de las grandes empresas, desde una *perspectiva crítica de las responsabilidades climáticas de las empresas*. Otro ejemplo lo plantea el mismo Bostrom (2002, 2013) al señalar el riesgo existencial asociado a un colapso global económico, en donde claramente las empresas pueden jugar un rol importante.

Estos riesgos ambientales y económicos son importantes, y una investigación más exhaustiva podría vislumbrar otros para un análisis panorámico. Sin embargo, el objetivo de este apartado es el de abordar otro riesgo que, por su emergencia y novedad, resulta relevante de explorar: el riesgo existencial asociado a la producción y aplicación de la IA en y desde las organizaciones. Este riesgo ciertamente es introducido por Bostrom (2017) en su obra *Superintelligence: Paths, Dangers, Strategies*. En dicho trabajo, se analiza, entre otras cosas, los riesgos existenciales que surgen a partir del desarrollo de una AGI o una IA fuerte o superinteligencia (Bostrom, 2017), como, por ejemplo, los asociados a los intereses competitivos de las empresas que se encuentren en la carrera por crear una AGI. Del choque de estos intereses competitivos, se puede generar serias implicaciones en la vida social, económica y política, sobre todo en el escenario donde una empresa logre una ventaja competitiva absoluta al lograr desarrollar y

apoderarse de la producción de una AGI, o también al no controlar el proceso de evolución de una AGI a una superinteligencia (Bostrom, 2017).

Otro importante elemento para señalar sobre este riesgo existencial asociado a la producción de IA tiene que ver con la escaza investigación sobre alineación entre los valores humanos y la IA, frente a la abundante investigación reciente enfocada en el aumento de capacidades de la IA, para pasar de una IA estrecha o débil, a una AGI (Han et al., 2022; Sutrop, 2020). Si no se logra una buena alineación entre los valores humanos y la IA, y además sigue aumentando la investigación enfocada en las capacidades de la IA, el resultado de ello es un posible riesgo existencial con una AGI o una Superinteligencia sin alineación con valores humanos (Bostrom, 2017; Sutrop, 2020). A lo anterior se debe sumar las distintas dificultades técnicas y normativas que, para el primer caso, se refiere a las dificultades técnicas sobre cómo codificar valores humanos en una IA, mientras que las dificultades normativas se refieren a qué tipo de valores (éticos, políticos, etc.) se deberían codificar, y desde qué enfoques, lo cual lleva a debates filosóficos (Sutrop, 2020).

Por su parte, en cuanto a la aplicación de la IA, uno de los riesgos existenciales que emergen son los asociados con la empleabilidad. De acuerdo con Romero Vela (2020), la inclusión de la IA, y también de mejoras a partir de la biotecnología, puede generar un riesgo existencial para la empleabilidad, ante lo cual se hace necesario formular políticas públicas que protejan la empleabilidad y no generen un riesgo existencial, posiblemente vinculado con un colapso social y económico. Al respecto, puede notarse como lo mencionado por esta autora cobra relevancia con la incursión de modelos de lenguaje de IA como ChatGPT, las cuales, si bien son modelos de IA que no pueden hasta el momento realizar la amplitud de funciones que haría teóricamente una AGI, ya representa un asunto de análisis respecto a la empleabilidad (Eloundou et al., 2023).

De acuerdo con el trabajo de Eloundou et al. (2023), en donde algunos autores pertenecen a la misma empresa de OpenAI que desarrolló el modelo de lenguaje de ChatGPT, se realizó una investigación sobre las posibles consecuencias de modelos de lenguaje de IA como el usado en ChatGPT, específicamente en el mercado laboral de los Estados Unidos, dando como resultado que la implementación de estas IA en empresas puede afectar distintos empleos, reemplazando muchas labores de profesiones que son altamente automatizables. Considerando esto, los riesgos existenciales asociados a la empleabilidad que puede generar la implementación de una IA con mayor potencial que este modelo de lenguaje de IA, son latentes.

El vínculo de estos problemas de riesgo existencial de producción y de aplicación de la IA con la administración y las organizaciones, es que los intereses productivo-competitivos, de eficiencia y económicos que suelen guiar la toma de decisiones administrativas y los fines organizacionales, pueden influenciar notoriamente en estos riesgos. Hipotéticamente, haciendo una deducción de dichos intereses en los modelos administrativos imperantes, parece que podrían estar alineados con riesgos existenciales latentes, al privilegiar intereses como el económico y el competitivo por encima del bienestar social, ambiental y la dignidad humana. Así pues, como se argumentará en el apartado siguiente, ante este eventual contexto en donde los estilos de administración y de organización imperantes guardan intereses que no contribuyen a evitar o mitigar riesgos existenciales vinculados a la producción y aplicación de IA, sino que incluso los agravaría y los podría justamente causar, se hace necesario que las corrientes críticas de la administración y de la organización jueguen un papel importante cuestionando estos modelos vigentes y proponiendo alternativas.

El Turno de las Corrientes Críticas de la Administración y de la Organización

Consideraciones Generales de Algunas Corrientes Críticas

Las corrientes críticas de la administración y la organización hacen referencia a diversas corrientes de pensamiento en los campos de estudio relacionados con la administración y la organización que se caracterizan por su enfoque en la generación de un conocimiento de tipo crítico social (Pineda-Henao, 2022). Este conocimiento crítico social se entiende como una contribución que proviene de diferentes enfoques de las ciencias sociales y humanas críticas, los cuales centran su atención en las problemáticas y las injusticias sociales, así como en la lucha correspondiente, que suele involucrar agentes de cambio social que buscan abordar desajustes institucionales relacionados con prácticas institucionalizadas que generan dichas injusticias o problemáticas (Ramírez, 2018).

La característica central de estas prácticas y teorías tradicionales radica en su orientación hacia la eficiencia y la productividad, lo que se traduce en la obtención de ganancias económicas (Aktouf, 2009; Gantman, 2017a; Misoczky, 2017; Montaño Hirose, 2013; Pineda-Henao, 2022). Por tanto, el aporte de estas corrientes críticas en los campos de estudio de la administración y la organización implica un cuestionamiento de las prácticas de administración y las formas de organización predominantes, así como de su justificación teórica y disciplinaria. Estas últimas se manifiestan, por ejemplo, en gran parte de la investigación aplicada y funcional de la administración, así como en las contribuciones tradicionales de la teoría organizacional y la teoría administrativa (Aktouf, 2009; Gantman, 2017a; Misoczky, 2017; Montaño Hirose, 2013).

En consecuencia, se busca destacar y denunciar las injusticias sociales, los actos inhumanos, la opresión y, en resumen, los aspectos negativos y oscuros de las prácticas de administración y las formas de organización predominantes. Además, en algunos casos, se proponen

alternativas de cambio y transformación necesarias en relación con la administración y los modos de organización predominantes (Gonzáles-Miranda & Rojas-Rojas, 2020; Misoczky, 2017; Saavedra Mayorga, 2009; Sanabria Rangel et al., 2015).

Por lo tanto, se comprende que muchas de las contribuciones de estas corrientes críticas de la administración y la organización pueden tener un propósito tanto de crítica epistemológica y ontológica (es decir, una crítica teórica) como de crítica ética y política (es decir, la esencia de la crítica social) contra las prácticas y teorías de la administración y las formas de organización predominantes (Gonzáles-Miranda & Rojas-Rojas, 2020; Misoczky, 2017; Montaño Hirose, 2013). En estas orientaciones críticas, predominan los objetivos ético-políticos, a partir de los cuales se puede fomentar la discusión epistemológica y ontológica, especialmente cuando se confrontan los marcos teóricos tradicionales de la administración y la organización. Esto se debe a que la crítica debe estar conectada con la realidad social que contextualiza la crítica misma, en beneficio de aquellos que carecen de voz y son víctimas de opresión o injusticia por parte de la administración y las formas de organización predominantes (Misoczky, 2017; Montaño Hirose, 2013; Núñez Rodríguez, 2022).

En América Latina, algunas de las corrientes críticas de la administración y las organizaciones que han tenido una gran influencia son los Estudios Críticos de la Administración (ECA), los Estudios Críticos Organizacionales (ECO) y la Gestión Humanista Radical (GHR) (Pineda-Henao, 2022). A pesar de la existencia de elementos epistemológicos e históricos que hacen que la identidad de estas corrientes parezca difuminarse, desde una perspectiva general y compartiendo autores y contribuciones en ocasiones, es posible distinguir ciertos elementos que, al menos con fines analíticos, se pueden resaltar en relación con su origen e identidad.

En el caso de los ECA, su origen se puede remontar a una corriente de pensamiento crítico surgida en las mismas escuelas de administración británicas, con autores como Alvesson & Willmott (1992,

2003), entre otros. Según Gantman (2017a), este surgimiento estuvo relacionado con la migración de académicos de las ciencias sociales y humanas a las escuelas de administración de ciertas universidades. Si se examinan algunos de los marcos epistemológicos de los ECA, se pueden identificar influencias de la Escuela de Frankfurt, el constructivismo y el posmodernismo (Saavedra Mayorga, 2009; Sanabria Rangel et al., 2015). Esta corriente aborda una amplia variedad de temas, pero su núcleo central radica en el cuestionamiento del estilo de administración predominante, destacando la opresión que resulta de su enfoque performativo, orientado hacia la eficiencia (instrumental y eficientista). En contraste, promueve una performatividad crítica que no solo critica la performatividad tradicional, sino que también tiene implicaciones prácticas a través de elementos comprensivos y reflexivos (Sanabria Rangel et al., 2015). A pesar de las críticas sobre la pertinencia de esta corriente (Misoczky, 2017; Misoczky et al., 2015), su influencia es relevante para analizar posibles aplicaciones en los modelos de administración actuales en relación con los riesgos existenciales relacionados con la IA.

Por otro lado, los ECO hacen referencia a la corriente crítica derivada de los Estudios Organizacionales (EO)[25], cuya génesis se produce a través de contribuciones que surgen de la formación del grupo EGOS y la revista Organization Studies en algunos países europeos (Clegg et al., 1996; Clegg & Bailey, 2007; Sanabria Rangel et al., 2014). Aunque los esquemas epistemológicos que sustentan esta corriente son similares a los de los ECA (incluyendo el constructivismo, el posmodernismo y la Teoría Crítica), los ECO se caracterizan, a pesar de su interdisciplinariedad, por un cierto énfasis sociológico (Rendón Cobián & Montaño Hirose, 2004; Sanabria Rangel et al., 2014). En esta

[25] Como la denominación de EO puede tener distintas connotaciones, incluso amplias, en donde se puede discutir su identidad con un mismo campo de estudios que incluye a la Teoría organizacional y los ECO, entre otros enfoques de estudio de la organización (Ríos Szalay, 2014; Saavedra-Mayorga & Sanabria, 2023), en el presente documento sólo se está haciendo referencia a la corriente crítica de los EO con la denominación de ECO.

corriente se pueden encontrar críticas relacionadas con el poder y el control, la identidad y la subjetividad en el trabajo, las ideologías en los discursos organizacionales, entre otros temas. En general, se destaca la intención de una comprensión crítica más amplia del fenómeno organizacional y una crítica más radical frente a la administración y las formas de organización predominantes (Gonzáles-Miranda, 2014; Sanabria Rangel et al., 2014).

Por último, la GHR es una derivación específica de la corriente más amplia de la Gestión Humanista, que abarca diversos enfoques en Europa y Canadá (Arandia & García-de-la-Torre, 2021; García-de-la-Torre et al., 2021). La derivación canadiense se destaca por su carácter radical, como lo demuestran, por ejemplo, las referencias directas de autores como Aktouf (1992, 2009), quienes proponen reemplazar el estilo de administración predominante por uno más crítico, comprensivo y con lógicas distintas al productivismo. Esta perspectiva también enfatiza la importancia de rescatar la dignidad humana y el bienestar medioambiental como prioridades, más allá de proporcionar únicamente modelos eruditos desde las humanidades y las ciencias sociales para analizar las organizaciones (Aktouf, 2009; Bédard, 2003; A. Chanlat, 1995; J.-F. Chanlat, 1994).

El Papel de las Corrientes Críticas Frente al Riesgo Existencial de la IA

Estas corrientes han sido utilizadas en otros trabajos para analizar su relevancia crítica en el contexto más amplio de la influencia del transhumanismo y el posthumanismo en la administración y las organizaciones (Pineda-Henao, 2022). La incidencia del transhumanismo y el posthumanismo en la administración y las organizaciones (Gladden, 2016) se puede entender como un nuevo episodio de la tendencia predominante en la administración que busca la eficiencia y la productividad. Sin embargo, esta búsqueda del mejoramiento humano y la promoción del postantropocentrismo pueden generar diversas problemáticas en las organizaciones, algunas

de las cuales están relacionadas con los riesgos existenciales (Pineda-Henao, 2022).

Por lo tanto, a continuación, se presentan algunas posibles formas de acción de la crítica, basadas en su potencial (Gantman, 2017b; Pineda-Henao, 2022), especialmente en relación con los riesgos existenciales asociados a la producción y aplicación de la inteligencia artificial (IA) en las organizaciones. Para ello, es importante recordar que el argumento central es que las corrientes críticas de la administración y la organización desempeñan un papel fundamental en la reflexión y propuesta de cambios en la administración predominante para evitar dichos riesgos. Este papel propuesto se puede desglosar en tres formas: 1) Participación en la formulación de políticas públicas; 2) La necesidad urgente de una mayor educación crítica y responsable en las escuelas de administración; 3) La orientación de programas de investigación específicos dentro de las corrientes críticas, enfocados en nuevas formas de administración destinadas a prevenir y mitigar los riesgos existenciales, ver Tabla 1.

Tabla 1

Potencialidades de las corrientes críticas ante riesgos de IA

Potencial de actuación de corrientes críticas	**Producción de IA**	**Aplicación de IA**
Participación en formulación de políticas públicas	Contribuciones críticas sobre los intereses empresariales tradicionales en la competitividad de las empresas que desarrollan IA Contribuciones críticas sobre sobre los	Contribuciones de discusión crítica frente a las tasas y limitaciones de la participación de IA en las organizaciones. Contribuciones críticas de las normativas que prevean y regulen la

	intereses empresariales tradicionales que inciden en investigación sobre IA, para privilegiar la alineación de valores humanos con la IA	transición de profesiones tradicionales y emergentes, frente a la implementación actual de la IA
Necesidad de una nueva formación humanística, crítica y responsable en escuelas de administración	Nuevos profesionales en administración para dirigir y tomar decisiones sobre la producción de IA de forma crítica, humanística y responsable	Nuevos profesionales en administración para la implementación de IA de forma crítica, humanística y responsable Nuevos profesionales en administración con valores críticos, humanísticos y responsables, menos remplazables por IA
Nuevas agendas de investigación sobre nuevas formas de administración y organización	Cuestionar y proponer alternativas frente a los modelos de administración tradicionales, que eviten y mitiguen riesgos existenciales por producción de IA	Cuestionar y proponer alternativas frente a los modelos de administración tradicionales, que eviten y mitiguen riesgos existenciales por aplicación de IA

En lo que respecta a la primera forma de acción de estas corrientes críticas frente a los riesgos existenciales relacionados con la IA en las organizaciones, es de gran relevancia debido a su alcance normativo. En el caso de las organizaciones que desarrollan IA, esto implica la

necesidad de establecer marcos normativos que fomenten una investigación más exhaustiva de los riesgos y consecuencias de la IA antes de su producción. Esto incluye la consideración de una investigación más profunda sobre cómo la IA se alinea con los valores humanos. Además, estos marcos normativos deberían conducir a la promulgación de tratados y acuerdos internacionales que prevengan o mitiguen posibles ventajas competitivas riesgosas derivadas de la producción de una inteligencia artificial general (AGI) o una superinteligencia.

En cuanto a la aplicación de la IA en las organizaciones, estos marcos normativos pueden enfocarse en una discusión crítica sobre las tasas y limitaciones de la participación de la IA en las organizaciones. Esto debe considerar la priorización de la inclusión laboral humana, así como la responsabilidad social y la sostenibilidad económica, no solo para las organizaciones, sino especialmente en términos de repensar las implicaciones y riesgos para los trabajadores humanos y la sociedad en general. Esto también implica la necesidad urgente de una regulación que prevea y gestione la transición de profesiones tradicionales y emergentes en respuesta a la rápida implementación de la IA, incluyendo IA débil o estrecha, como ChatGPT.

Estas acciones normativas podrían derivarse de una perspectiva que se enfoque en las posibilidades prácticas de las corrientes críticas, en lugar de limitarse a denunciar los riesgos (Gantman, 2017b; Sanabria Rangel et al., 2015). En este sentido, la comprensión crítica de la competitividad organizacional, el papel de la tecnología en las organizaciones, la responsabilidad social y la sostenibilidad económica, los objetivos instrumentales y económicos de la performatividad, las nuevas formas de producción y empleabilidad, y la reintroducción de la discusión sobre la dignidad humana en las organizaciones, aportadas por los ECA, los ECO y la GHR, desempeñarían un papel esencial en la formulación de estos marcos normativos.

En cuanto a la segunda forma de acción de las corrientes críticas, que implica la urgente inclusión de una mayor educación crítica y

responsable en las escuelas de administración, es importante destacar que una de las principales deficiencias en este sentido radica en que la formación en administración tradicional ha normalizado el enfoque instrumental y productivo de la administración. Además, da prioridad a la investigación aplicada y funcional sobre otras formas de investigación (Pineda-Henao, 2014, 2017, 2018a; Pineda-Henao, Ortega González, et al., 2020; Pineda-Henao & Tello-Castrillón, 2018). Por lo tanto, incluso desde la perspectiva de la erudición disciplinaria dentro de la tradición de la administración predominante, y especialmente con el propósito de fomentar el pensamiento reflexivo y crítico, es esencial fortalecer la formación investigativa en este campo, incluso en el nivel de pregrado (Giraldo López et al., 2019; Pineda-Henao, 2018b, 2018a; Pineda-Henao, Tello Castrillón, et al., 2020). Esto resalta la importancia de la investigación más crítica y científica en estos campos de estudio, que vaya más allá de los modelos convencionales de administración y las concepciones de organización predominantes (Pineda-Henao, 2017, 2021; Pineda-Henao & Tello-Castrillón, 2018).

Dentro de este contexto, tanto para la producción como para la aplicación de la IA en las organizaciones, se requiere una formación crítica y responsable que incorpore áreas humanísticas y sociales. Además, es fundamental establecer una base sólida y profunda en corrientes críticas como los ECA, los ECO y la GHR en los planes de estudio, junto con una formación crítica en temas de responsabilidad social organizacional y ética en la administración y las organizaciones. En general, estos enfoques de formación humanista y social, basados en corrientes críticas en el campo de la administración y la organización, pueden llevar a un cambio profundo en la educación en administración, contribuyendo a la formación de una nueva generación de administradores que puedan enfrentar los retos planteados por la IA y los riesgos existenciales de manera más efectiva.

Por su parte, en lo que respecta a la aplicación de la IA, igualmente pueden impactar en dos sentidos: por un lado, en los profesionales administrativos que toman decisiones de implementar IA, desde

criterios de formación humanísticos, sociales, críticos, responsables y éticos que consideren los riesgos y consecuencias de la IA, y la primacía de la dignidad humana en las organizaciones. Por otro lado, en lo que respecta a la empleabilidad, el hecho de formar profesionales en administración con fundamentos sólidos y profundos en humanidades, ciencias sociales, corrientes críticas y ética y responsabilidad social da lugar a un factor distintivo de dicha profesión con mayor dificultad de automatización por parte de una IA, por lo que el valor profesional aumentaría, contribuyendo a la reducción del riesgo asociado a la empleabilidad.

Como se destacó líneas arriba, los ECA, ECO y GHR, se fundamentan en marcos epistemológicos que justamente privilegian la crítica social y distintos referentes de las ciencias sociales y las humanidades, con el especial énfasis en que dichos marcos se aplican en comprender críticamente a la administración y las organizaciones. Por lo anterior, su potencial en la formación en escuelas de administración es crucial para plantear formas de mitigar y evitar riesgos existenciales relativos a la producción y aplicación de IA, desde dentro de las mismas escuelas de administración. Como se mencionó, esto no sólo va aunado con el proyecto de erudición de la disciplina misma, sino sobre todo de un cambio o transformación de la formación en administración, que abandone el estilo de administración tradicional y lo reemplace por nuevas formas de administración que no tengan en su centro la eficiencia y la productividad, que claramente están vinculados a modelos de administración con una mayor fragilidad de riesgos existenciales.

Lo anterior lleva al último punto de actuación de las corrientes críticas. Para ello, hay que comprender que tanto la producción como la aplicación de IA en y desde las organizaciones es un asunto crucial respecto al riesgo existencial, en la medida que en los estilos de administración y de organización imperantes plantean ejes, como los ya mencionados de la eficiencia y productividad, que hacen que sus modelos de gestión supongan modelos frágiles ante estos riesgos.

Dicha fragilidad radica en que, al privilegiarse la búsqueda de una mayor eficiencia y productividad, la derivación lógica más razonable de ello es que en la producción y aplicación de la IA prevalezcan intereses más económicos e instrumentales de las organizaciones, que los intereses éticos y sociales de los trabajadores y de la sociedad (Pineda-Henao, 2022).

Si se le mira desde ese ángulo, una posible deducción es que las corrientes críticas, en sus nuevas agendas de investigación, incluyan en su denuncia y sus formulaciones de nuevas formas de administración, los riesgos y las consecuencias negativas de la tecnología en las organizaciones, y especialmente el tema del riesgo existencial relativo a la IA. Estos temas, si bien son emergentes desde el marco general del transhumanismo y el posthumanismo en la administración y las organizaciones, requieren una mayor atención específica por parte de las corrientes críticas de los ECA, los ECO y la GHR (Pineda-Henao, 2022). Así pues, puede pensarse que la idea de performatividad crítica incluya la problematización de la relación estrecha del trabajo humano y la IA. Igualmente, que la discusión de la dignidad humana en la administración y las organizaciones incluya no sólo las discusiones filosóficas sobre el humanismo y el posthumanismo, sino también que se incluya frente a la alineación de los valores humanos frente a la IA, en términos de, por ejemplo, evitar los actos inhumanos, las injusticias y la opresión (Pineda-Henao, 2022).

Finalmente, pensar en nuevas formas de administración y de organización que eviten y mitiguen los riesgos existenciales en general, y en particular los relativos a la IA, sería entonces uno de los mayores retos de estas corrientes críticas, porque supone no sólo agregar un elemento a la discusión conceptual de las nuevas formas de concebir la administración y la organización, sino también de pensar en nuevas formas de intervención y aplicabilidad práctica, tal vez más orientadas desde las restricciones y las limitaciones de la IA, por la naturaleza negativa misma de la crítica. Sin embargo, en términos de las alternativas de la crítica, esto también invita a pensar en nuevos

sentidos de la eficiencia, orientados a objetivos distintos de la productividad y las ganancias, y que privilegien el trabajo y la dignidad humana.

Conclusiones

En este trabajo, se partió de la escasa producción de conocimiento que establece una conexión directa entre el tema de los riesgos existenciales asociados a la producción y aplicación de la inteligencia artificial (IA) con las corrientes críticas de la administración y de la organización. A pesar de su relevancia y urgencia, este podría ser un tema que no está siendo abordado de manera directa por académicos en dicho campo de estudio. Por lo tanto, se argumentó cómo estas corrientes críticas pueden desempeñar un papel importante en la concienciación y prevención de estos riesgos, al cuestionar el modelo de administración tradicional, que se muestra vulnerable ante este tipo de amenazas.

A partir de lo expuesto, se deduce la necesidad de que las corrientes críticas, como los ECA, los ECO y la GHR generen más contribuciones de conocimiento crítico-social que contribuyan a la formulación de políticas públicas, la reformulación de la educación en administración y la presentación de cuestionamientos y alternativas en relación con el modelo de administración tradicional. Este modelo tradicional favorece intereses que no están alineados con la prevención y mitigación de los riesgos existenciales relacionados con la producción y aplicación de la IA en las organizaciones, y su predominio podría tener un impacto negativo en este asunto. En términos generales, más allá de la necesidad de utilizar la IA de manera responsable, es crucial debatir los aspectos negativos, los límites y los cambios necesarios en la producción y aplicación de la IA en el contexto organizacional.

En lo que respecta a futuras investigaciones, se recomienda abordar otros tipos de riesgos existenciales relacionados con la administración

y las organizaciones, como los riesgos asociados al cambio climático y al colapso ambiental, así como los riesgos relacionados con colapsos económicos y sociales. Las organizaciones, especialmente las empresas privadas, y el modelo de administración tradicional, desempeñan un papel central en estos desafíos de riesgos existenciales. Por lo tanto, es esencial llevar a cabo un análisis crítico interdisciplinario, ya que los intereses políticos, éticos y económicos predominantes en las empresas y en este modelo de administración podrían estar pasando por alto cuestiones que amenazan el futuro de la humanidad.

Agradecimientos

Este trabajo se nutre, fundamentalmente, de aportaciones que se han derivado como reflexiones emergentes a partir de la tesis doctoral en curso, titulada Análisis epistemológico del devenir contemporáneo de las corrientes críticas de la administración y de la organización en América Latina, del primer autor Elkin Fabriany Pineda-Henao, del Doctorado en Ciencias de la Administración de la Universidad Nacional Autónoma de México. Por ello, se extiende un agradecimiento especial al Dr. Jorge Ríos Szalay (director de tesis), y al comité tutor los doctores Dr. Luis Montaño Hirose y Dr. Luis Cruz Soto. También se extiende un agradecimiento al Programa de Posgrado en Ciencias de la Administración de la Facultad de Contaduría y Administración de la UNAM, y, finalmente, un especial agradecimiento al Consejo Nacional de Humanidades, Ciencia y Tecnología (Conahcyt) por el patrocinio del programa de Becas Nacionales para la tesis en curso.

Referencias

Aktouf, O. (1992). Management and Theories of Organizations in the 1990s: Toward a Critical Radical Humanism? Academy of Management Review, 17(3), 407–431. https://doi.org/10.5465/amr.1992.4281975

Aktouf, O. (2009). La administración: Entre tradición y renovación (4ta edición en español). Artes Gráficas Univalle.

Alvarado Rojas, M. (2015). Una mirada a la inteligencia artificial. Revista Ingeniería, Matemáticas y Ciencias de la Información, 2(3), 27–31.

Alvesson, M., & Willmott, H. (1992). Critical Management Studies. Sage Publications (CA).

Alvesson, M., & Willmott, H. (2003). Studying Management Critically. SAGE Publications.

Anderson, S. L., & Anderson, M. (2011). A prima facie duty approach to machine ethics: Machine learning of features of ethical dilemmas, prima facie duties, and decision principles through a dialogue with ethicists. En M. Anderson & S. L. Anderson, Machine Ethics (First Ed., pp. 476–494). Cambridge University Press.

Arandia, O., & García-de-la-Torre, C. A. (2021). Humanistic management: A history of a management paradigm from the human dignity. En C. A. García-de-la-Torre, O. Arandia, & M. Vázquez-Maguirre (Eds.), Humanistic Management in Latin America (pp. 1–18). Routledge.

Bédard, R. (2003). Los fundamentos del pensamiento y las prácticas administrativas. 1—El rombo y las cuatro dimensiones filosóficas. AD-minister, 3, 68–88.

Bostrom, N. (2002). Existential risks: Analyzing human extinction scenarios and related hazards. Journal of Evolution and Technology, 9. http://jetpress.org/volume9/risks.html

Bostrom, N. (2013). Existential Risk Prevention as Global Priority. Global Policy, 4(1), 15–31. https://doi.org/10.1111/1758-5899.12002

Bostrom, N. (2017). Superintelligence: Paths, Dangers, Strategies (2° ed.). Oxford University Press.

Bostrom, N., & Cirkovic, M. M. (2008). Introduction. En N. Bostrom & M. M. Cirkovic (Eds.), Global Catastrophic Risks (First published, pp. 1–30). OUP Oxford.

Boyd, M., & Wilson, N. (2020). Existential Risks to Humanity Should Concern International Policymakers and More Could Be Done in Considering Them at the International Governance Level. Risk Analysis, 40(11), 2303–2312. https://doi.org/10.1111/risa.13566

Chanlat, A. (1995). Modos de pensamiento y comunicación. HEC-Montréal, Groupe Humanisme et Gestion.

Chanlat, J.-F. (1994). Hacia una antropología de la organización. Gestión y Política Pública, III(2), 317–364.

Clegg, S. R., & Bailey, J. (2007). International Encyclopedia of Organization Studies. Sage.

Clegg, S. R., Hardy, C., & Nord, W. R. (1996). Handbook of organization studies (pp. xxix, 730). Sage Publications, Inc.

Collins, C., Dennehy, D., Conboy, K., & Mikalef, P. (2021). Artificial intelligence in information systems research: A systematic literature review and research agenda. International Journal of Information Management, 60, 102383. https://doi.org/10.1016/j.ijinfomgt.2021.102383

Eloundou, T., Manning, S., Mishkin, P., & Rock, D. (2023). GPTs are GPTs: An Early Look at the Labor Market Impact Potential of

Large Language Models (arXiv:2303.10130). arXiv. https://doi.org/10.48550/arXiv.2303.10130

Flechoso, J. J. (2021). Digitalización y recuperación económica: El papel de la digitalización en la recuperación socioeconómica tras la pandemia (Primera ed). Editorial Almuzara.

Gantman, E. R. (2017a). El desarrollo de los estudios críticos de gestión en los países latinoamericanos de habla hispana. Política y Sociedad, 54(1), 45–64. https://doi.org/10.5209/POSO.51679

Gantman, E. R. (2017b). En torno al potencial transformador de los CMS (Critical Management Studies). RECERCA. Revista de Pensament i Anàlisi, 20, Article 20. https://doi.org/10.6035/Recerca.2017.20.2

García-de-la-Torre, C. A., Arandia, O., & Vázquez-Maguirre, M. (Eds.). (2021). Humanistic Management in Latin America. Routledge.

Garrell, A., & Guilera, L. (2019). La Industria 4.0 en la sociedad digital (Primera ed.). Marge Books.

Giraldo López, A. R., Tello-Castrillón, C., Londoño-Cardozo, J., & Pineda-Henao, E. F. (2019). Influencia de la malla curricular en la formación investigativa en programas de administración en Colombia. Revista Argentina de Investigación en Negocios, 5(1), 19–32.

Gladden, M. E. (2016). Posthuman Management: Creating Effective Organizations in an Age of Social Robotics, Ubiquitous AI, Human Augmentation, and Virtual Worlds (Second Edition). Defragmenter Media, West Pole & Larkspur, and Synthypnion Press LLC.

Gonzáles-Miranda, D. R. (2014). Los estudios organizacionales. Un campo de conocimiento comprensivo para el estudio de las organizaciones. Innovar: revista de ciencias administrativas y sociales, 24(54), 43–58. https://doi.org/10.15446/innovar.v24n54.46431

Gonzáles-Miranda, D. R., & Rojas-Rojas, W. (2020). Repensando la crítica en los estudios organizacionales. Innovar, 30(78), 3–10. https://doi.org/10.15446/innovar.v30n78.90295

Han, S., Kelly, E., Nikou, S., & Svee, E.-O. (2022). Aligning artificial intelligence with human values: Reflections from a phenomenological perspective. AI & SOCIETY, 37(4), 1383–1395. https://doi.org/10.1007/s00146-021-01247-4

Højme, P. (2019). Whose Survival? A Critical Engagement with the Notion of Existential Risk. Scientia et Fides, 7(2), Article 2.

Iglesias-Márquez, D. (Daniel). (2020). Cambio climático y responsabilidad empresarial: Análisis del papel de las empresas para alcanzar los objetivos del Acuerdo de París. https://doi.org/10.15581/010.36.327-366

Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. Science, 349(6245), 255–260. https://doi.org/10.1126/science.aaa8415

Kaku, M. (2014). El futuro de nuestra mente (J. M. Ibeas Delgado & M. Pérez Sánchez, Trads.). Debate.

Kemp, T. (1979). La revolujción industrial en la Europa del siglo XIX (R. Ribé, Trad.; Tercera ed.). Fontanella.

Kurzweil, R. (2014). The Singularity is Near. En R. L. Sandler (Ed.), Ethics and Emerging Technologies (1 ed., pp. 393–406). Palgrave Macmillan UK. https://doi.org/10.1057/9781137349088_26

Londoño-Cardozo, J., & Pérez de Paz, M. (2021). Corporate Digital Responsability: Foundations and considerations for its development. RAM. Revista de Administração Mackenzie, 22. https://doi.org/10.1590/1678-6971/eRAMD210088

Manheim, D. (2020). The Fragile World Hypothesis: Complexity, Fragility, and Systemic Existential Risk. Futures, 122, 102570. https://doi.org/10.1016/j.futures.2020.102570

Melnyk, L. H., Kubatko, O. V., Dehtyarova, I. B., Dehtiarova, I. B., Matsenko, O. M., Рожко, О. Д., Рожко, А. Д., & Rozhko, O. D. (2019). The effect of industrial revolutions on the transformation of social and economic systems. Problems and Perspectives in Management, 17(4), 381–391. https://doi.org/10.21511/ppm.17(4).2019.31

Misoczky, M. C. (2017). ¿De qué hablamos cuando decimos crítica en los Estudios Organizacionales? Administración & Desarrollo, 47(1), 141–149.

Misoczky, M. C., Flores, R. K., & Goulart, S. (2015). An Anti-Management Statement in Dialogue with Critical Brazilian Authors. RAE - Revista de Administração de Empresas, 55(2), 130–138.

Mitchell, J. B. O. (2014). Machine learning methods in chemoinformatics. WIREs Computational Molecular Science, 4(5), 468–481. https://doi.org/10.1002/wcms.1183

Mitchell, T. M. (1997). Machine Learning. McGRAW-HILL.

Montaño Hirose, L. (2013). Los estudios organizacionales. Revisando el papel de la crítica en la administración. En R. Carvajal Baeza, Estudios críticos de la organización: Qué son y cuál es su utilidad (Primera, pp. 21–46). Universidad del Valle - Facultad de ciencias de la Administración.

Mosqueira-Rey, E., Hernández-Pereira, E., Alonso-Ríos, D., Bobes-Bascarán, J., & Fernández-Leal, Á. (2023). Human-in-the-loop machine learning: A state of the art. Artificial Intelligence Review, 56(4), 3005–3054. https://doi.org/10.1007/s10462-022-10246-w

Núñez Rodríguez, C. J. (2022). Apuntes para una teoría crítica en los estudios críticos de la administración. En O. L. Anzola Morales, C. J. Núñez Rodríguez, & M. T. Magallón Díez (Eds.), Problemas contemporáneos de administración y estudios

organizacionales: Una perspectiva latinoamericana (1a ed., pp. 19–51). Universidad Externado de Colombia.

Ortega y Gasset, J. (1914). Meditaciones del Quijote (Primera ed.). Publicaciones de la residencia de estudiantes.

Pérez de Paz, M. (2016). Homo Compensator: Le parcours philosophique d'un concept métaphysique [Mémoire présenté en vue de l'obtention du master philosophie parcours théories pratique et conflit, Universidad de Poitiers]. www.doi.org/10.13140/RG.2.2.28294.91209

Pérez de Paz, M., & Londoño-Cardozo, J. (2021). La implementación de los robots y la inteligencia artificial en las organizaciones: Una paradoja para la Responsabilidad Social. En J. Londoño-Cardozo & O. I. Vásquez (Eds.), La investigación en Administración: Tendencias, enfoques y discusiones (Primera ed., pp. 185–219). Editorial Universidad Santiago de Cali.

Pérez de Paz, M., Londoño-Cardozo, J., & Tello Castrillón, C. (2021). Tecnologías agenciativas y la Responsabilidad Digital Organizacional: Conflictos, retos y soluciones. VI Simposio Internacional de Responsabilidad Social de las Organizaciones (SIRSO).

Pineda-Henao, E. F. (2014). Una fundamentación ontológica de la práctica administrativa como técnica social ordenadora institucionalizada [Tesis pregrado]. Universidad Nacional de Colombia.

Pineda-Henao, E. F. (2017). Disciplina administrativa y práctica administrativa: Una perspectiva analítica del problema del estatus epistemológico de la administración [Tesis pregrado]. Universidad del Valle.

Pineda-Henao, E. F. (2018a). Administración y organizaciones: Una mirada más allá de las fronteras de lo instrumental. En C. Tello Castrillón & E. F. Pineda-Henao, Conjeturas organizacionales:

Fundamentos para el estudio de la organización (Primera ed., pp. 31–54). Editorial Universidad Nacional de Colombia.

Pineda-Henao, E. F. (2018b). Sobre la formación investigativa: Diagnóstico comparativo del programa de Administración de la Universidad Nacional de Colombia Sede Palmira [Tesis Maestría en Administración, Universidad Nacional de Colombia]. https://www.researchgate.net/publication/331877148_Sobre_la_formacion_investigativa_diagnostico_comparativo_del_programa_de_Administracion_de_la_Universidad_Nacional_de_Colombia_Sede_Palmira

Pineda-Henao, E. F. (2021). Desafíos filosóficos para una sistemática científica de las organizaciones [Tesis de Maestría en Filosofía]. Universidad del Valle.

Pineda-Henao, E. F. (2022). Humano, ineficientemente humano: Reflexiones críticas sobre transhumanismo y posthumanismo en las organizaciones. Revista de Administración Pública del GLAP, 6(10), 18–35.

Pineda-Henao, E. F., Ortega González, M. S., & Rivera Morillo, V. (2020). El bien, el mal y el acto de administrar: Una fundamentación crítica desde la razón práctica. En C. Tello-Castrillón, E. F. Pineda-Henao, & J. Londoño-Cardozo (Eds.), La construcción organizacional de la Responsabilidad Social: Fundamentos teóricos y casos de estudio (Primera ed., pp. 39–64). Universidad Nacional de Colombia.

Pineda-Henao, E. F., Tello Castrillón, C., Ortega González, M. S., & Londoño-Cardozo, J. (2020). Dimensiones formales y culturales de la formación investigativa en pregrados de Administración colombianos. En J. C. Arboleda Aparicio (Ed.), Deliberaciones Gerenciales: Perspectiva de la ciudad de Palmira (Primera edición, pp. 147–195). Editorial REDIPE.

Pineda-Henao, E. F., & Tello-Castrillón, C. (2018). ¿Ciencia, técnica y arte?: Análisis crítico sobre algunas posturas del problema del estatus epistemológico de la Administración. Revista Logos Ciencia & Tecnología, 10(4), 112–130.

Ramírez, C. A. (2018). Reconstrucción, proyección, deconstrucción: Una tipología de las ciencias sociales críticas. En Ontología social: Una disciplina de frontera (Primera, pp. 145–168). Universidad Nacional de Colombia.

Rees, M. (2004). Our Final Hour: A Scientist's Warning. Basic Books.

Rendón Cobián, M., & Montaño Hirose, L. (2004). Las aproximaciones organizacionales. Caracterización, objeto y problemática. Contaduría y administración, 213, 1–15.

Rifkin, J. (2011). La Tercera Revolución Industrial: Cómo el poder lateral está transformando la energía, la economía y el mundo (Primera edición). Paidós.

Ríos Szalay, J. (2014). Sobre el estudio de las organizaciones. ¿Traslapes interdisciplinares hacia una ciencia organizacional? XIX Congreso Internacional de Contaduría, Administración e Informática, 1–19.

Roel, V. (1998). La tercera revolución industrial y la era del conocimiento (3ra. Edición). Fondo Editorial Universidad Nacional Mayor de San Marcos.

Romero Vela, S. L. (2020). Análisis de riesgo existencial y el futuro de la empleabilidad. Phainomenon, 19(1), Article 1. https://doi.org/10.33539/phai.v19i1.2174

Russell, S. J., & Norvig, P. (2010). Artificial intelligence: A modern approach (Third edition). Prentice Hall.

Saavedra Mayorga, J. J. (2009). Descubriendo el lado oscuro de la gestión: Los critical management studies o una nueva forma. Revista Facultad de Ciencias Económicas: Investigación y Reflexión, 17(2), 45–60.

Saavedra-Mayorga, J. J., & Sanabria, M. (2023). Teoría organizacional y estudios organizacionales: Dos denominaciones para un mismo campo de conocimiento. Innovar, 33(90), Article 90. https://doi.org/10.15446/innovar.v33n90.111442

Sanabria Rangel, M., Saavedra Mayorga, J. J., & Smida, A. (2014). Los estudios organizacionales. Fundamentos evolución y estado actual del campo (1a ed.). Editorial Universidad del Rosario. https://www.jstor.org/stable/j.ctt1f5g2pv

Sanabria Rangel, M., Saavedra Mayorga, J. J., & Smida, A. (2015). Los estudios críticos en administración: Origen, evolución y posibilidades de aporte al desarrollo del campo de los estudios organizacionales en América Latina. Revista Facultad de Ciencias Económicas: Investigación y Reflexión, 23(1), 209–234.

Seers, K. (2012). What is a qualitative synthesis? Evidence-Based Nursing, 15(4), 101–101. https://doi.org/10.1136/ebnurs-2012-100977

Sutrop, M. (2020). Challenges of Aligning Artificial Intelligence with Human Values. Acta Baltica Historiae et Philosophiae Scientiarum, 8(2), 54–72. https://doi.org/10.11590/abhps.2020.2.04

Taleb, N. N. (2016). The black swan: The impact of the highly improbable. Penguin Books Ltd.

Thomas, J., & Harden, A. (2008). Methods for the thematic synthesis of qualitative research in systematic reviews. BMC Medical Research Methodology, 8(1), 45. https://doi.org/10.1186/1471-2288-8-45

Torres, P. (2023). Existential risks: A philosophical analysis. Inquiry, 66(4), 614–639. https://doi.org/10.1080/0020174X.2019.1658626

Vaidya, S., Ambad, P., & Bhosle, S. (2018). Industry 4.0 – A Glimpse. Procedia Manufacturing, 20, 233–238. https://doi.org/10.1016/j.promfg.2018.02.034

Villani, P. (2009). La Inglaterra de la revolución industrial y la Europa de Napoleón y desde 1848 a 1871. En V. de la Torre Veloz, N. López Saavedra, & M. A. González (Eds.), La revolución industrial y el pensamiento político y social en el capitalismo contemporáneo (Siglo XIX) (2 edicion, pp. 105–127). Universidad Autónoma Metropolitana.

Vinod, B. (2023). Artificial Intelligence in travel. En B. Vinod (Ed.), Artificial Intelligence and Machine Learning in the Travel Industry: Simplifying Complex Decision Making https://doi.org/10.1007/978-3-031-25456-7_13 (pp. 163–170). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-25456-7_13

Xu, M., David, J. M., & Kim, S. H. (2018). The Fourth Industrial Revolution: Opportunities and Challenges. International Journal of Financial Research, 9(2), 90–95. https://doi.org/10.5430/ijfr.v9n2p90