



EN MAESTRÍA
DE LA FILOSOFÍA
CULTURA



Ciencia y Tecnología
Secretaría de Ciencia, Humanidades, Tecnología e Innovación



UNIVERSIDAD MICHOACANA DE SAN NICOLÁS DE HIDALGO

Instituto de Investigaciones Filosóficas "Luis Villoro"

Facultad de Filosofía "Dr. Samuel Ramos Magaña"

Programa institucional de Maestría en Filosofía de la Cultura

Las consecuencias de la modernidad en la era de la inteligencia artificial

Tesis

que para obtener el grado de Maestro en Filosofía de la Cultura

Presenta

Alan Osvaldo Zamudio Ibañez

Asesor:

Dr. Oliver Kozlarek

Comité tutorial:

Dr. Adán Pando Moreno

Dr. Emiliano José Mendoza Solís

Morelia, Mich. enero de 2026

Resumen

El presente trabajo examina críticamente la caracterización tardomoderna de la dinámica social (conciencia del riesgo, pérdida de confianza y politización de la técnica) ante el progreso técnico desmedido, en el contexto del desarrollo de la inteligencia artificial (IA). Dada la vaguedad del significado de IA, su definición coloquial se encuentra mediada por promesas, expectativas y miedos, y desde los ámbitos de poder político y económico se establecen narrativas, muchas veces sesgadas, que moldean estas definiciones de acuerdo a sus intereses. Ante ello, propongo la conciencia del riesgo asociado al crecimiento de la IA como criterio negativo de lo humano en este contexto, y desde un humanismo crítico, constructivo y pragmático, establezco el marco para evaluar y orientar tal tecnología en términos éticos y políticos, enfatizando el papel de la agencia social en el devenir técnico y desmontando discursos sesgados de inevitabilidad, necesidad y gestión respecto de la IA, así como narrativas trans- y posthumanistas acríticas.

Palabras clave: tardomodernidad, riesgo, desarrollo tecnológico, humanismo crítico, imaginarios sociotécnicos, .

Abstract

This work critically examines late-modern characterization of social dynamics (risk awareness, loss of trust and politicization of technology) against the unmeasured technical advancement, in the context of artificial intelligence (AI) development. Given the vagueness of AI's concept, its public definition is mediated by promises, expectations and fears, while political and economic powers construct often biased narratives that shape these definitions to serve their interests. Against this, I propose the awareness of associated risk to AI's growth as a negative criterion of the human, and from a constructive and pragmatic critical humanism I set the framework to evaluate and guide such technology in ethical and political terms. The analysis emphasizes the role of social agency in technical becoming and dismantles biased inevitability, necessity and management of AI discourses, as well as uncritical trans- and posthumanist narratives.

Índice

Introducción.....	5
Capítulo 1: El Diagnóstico de la Modernidad de Anthony Giddens y Ulrich Beck.....	9
1.1: La Modernidad Tardía y la Radicalización de sus Consecuencias.....	11
1.2: El Paradigma de la Sociedad del Riesgo.....	18
1.3: La Crítica Implícita en el Diagnóstico de la Modernidad.....	28
Capítulo 2: Concepto de Inteligencia Artificial.....	32
2.1: Concepto Tradicional de Inteligencia Artificial.....	33
2.2: El Devenir Contingente de la Inteligencia Artificial Hasta su Actualidad.....	37
2.3: Riesgos de la Inteligencia Artificial.....	39
Capítulo 3: El Humanismo como Fundamento para una Crítica de la Modernidad del Riesgo...	44
3.1: El Problema de la Deshumanización Moderna en Thomas Fuchs.....	47
3.2: La Propuesta Fuchsiana del Humanismo Encarnado como Defensa de lo Humano.....	52
3.3: Reubicando la Discusión: del Argumento Ontológico al Argumento Pragmático.....	56
3.4: Hacia un Humanismo Crítico.....	65
3.5: El Papel de la Crítica Social.....	75
Capítulo 4: La Definición Discursiva de la Inteligencia Artificial: Un Abordaje Crítico de sus Imaginarios.....	78
4.1: La Inteligencia Artificial como Proyecto Sociotécnico: Crítica de sus Imaginarios Predominantes.....	79
4.1.1: La Búsqueda de Legitimación de la Inteligencia Artificial Mediante la Configuración Discursiva.....	82

4.1.2: Riesgo, Agencia Social y Co-construcción en la Naturaleza Sociotécnica de la Inteligencia Artificial.....	87
4.2: La Inteligencia Artificial como Proyecto Filosófico: Crítica a una Metafísica Aplicada...	96
4.2.1: Transhumanismo y Posthumanismo.....	98
4.2.2: Los Límites Operativos de la Especulación Ontológica.....	102
Conclusiones.....	109
Referencias.....	113
Anexos.....	117

Introducción

El espectro tecnológico que abarca el término *inteligencia artificial* (IA) ha adquirido un papel clave en la contemporaneidad. Su presencia se manifiesta y refuerza continuamente en ámbitos tan diversos como la economía, la política, la cultura y la vida cotidiana, a tal punto que hablar de *la era de la inteligencia artificial* es un tópico recurrente. No obstante, sería erróneo reducir la IA a un mero avance técnico. La IA es también, y quizá sobre todo, un dispositivo simbólico y discursivo donde se concentran descripciones, promesas, expectativas y temores.

Este carácter trasciende al ámbito de la ingeniería. La IA es un objeto sociotécnico en el que confluyen intereses económicos de escala global, proyectos de poder político y definiciones filosóficas sobre el ser humano. En consecuencia, debe ser entendida como construcción histórica y cultural que refleja las dinámicas propias de la sociedad en que se desarrolla.

Esta tecnología se inscribe de lleno en la lógica moderna. Hoy la IA puede pensarse como su expresión paradigmática en tanto reproduce sus ambivalencias: promesa de emancipación y amenaza de deshumanización. Por esto mismo se observa también a su alrededor una desconfianza y conciencia de los riesgos potenciales de su uso, politizando su avance de acuerdo con una reflexividad que caracteriza a la modernidad tardía y que proviene de las consecuencias desastrosas del desarrollo técnico desmedido efectuado en un pasado cercano. De ahí que, antes de celebrar sus promesas o condenar sus amenazas, resulte imprescindible situar la IA como objeto contingente del progreso moderno que debe ser sometido a escrutinio crítico.

Tal empresa requiere de un análisis serio sobre lo que de hecho supone la inteligencia artificial. Los imaginarios que rodean a esta tecnología están, en gran parte, contruidos sobre afirmaciones y concepciones sesgadas o transfiguradas. Ante ello, el propósito de esta tesis es desmontar las narrativas acríticas predominantes respecto a la IA que la han convertido en una promesa de salvación inevitable o en un destino de trascendencia de lo humano, ocultando el entramado social de intereses, relaciones de poder y afirmaciones filosóficas que la constituyen.

El presente trabajo se orienta, por tanto, a abrir un espacio de problematización desde el cual pueda ser socialmente disputado el sentido y el rumbo de esta tecnología. Desde el humanismo crítico se esboza un horizonte normativo capaz de evaluar la IA en términos de ética, libertad, dignidad y evitación del sufrimiento, con la finalidad de desmontar las narrativas acríticas que la rodean y a su vez esbozar propositivamente las pautas a seguir. Tiene, pues, una intención política práctica.

La estructura del trabajo se divide en cuatro momentos principales. En el capítulo 1 se exponen los fundamentos teóricos desde los cuales se entiende la modernidad como un proceso caracterizado por el desarrollo técnico desmedido que, en su etapa tardía, ha sentado al riesgo como eje de las dinámicas sociales. Se describe la modernidad como un entramado de instituciones y modos de vida con origen en occidente pero expandido a escala global. Su configuración institucional, lejos de consolidar un progreso armónico, ha generado condiciones de desconfianza e incertidumbre que definen la experiencia contemporánea.

De ello deriva la conciencia del riesgo, donde se reconoce que los efectos negativos asociados al proceso de modernización no son fallas accidentales, sino consecuencias estructurales de estas sociedades. Así, se deslegitima el crecimiento técnico desmedido. Las sociedades modernas adquieren un sentido reflexivo que tematiza a su propia caracterización como problemática, esbozando con ello un horizonte normativo negativo basado en la evitación de los riesgos, y politizando los ámbitos científico y tecnológico.

En el segundo capítulo se delimita el objeto de estudio. Se presenta primero una conceptualización tradicional de la inteligencia artificial desde la óptica tecnocientífica, describiendo su desarrollo histórico y los fundamentos que han guiado sus avances. Esta definición, aunque insuficiente por sí misma, al ser examinada minuciosamente permite comprender las diferentes dimensiones que de hecho conforman la conceptualización de la IA.

Posteriormente se reconstruye el devenir contingente de la IA como institución sociotécnica moderna. Se muestra que su consolidación ha dependido tanto de rupturas

científicas como de factores externos: inversiones corporativas, intereses militares, narrativas mediáticas y políticas públicas. De este modo, la IA actual aparece no como una invención puramente técnica, sino como una construcción situada socialmente.

Se finaliza el capítulo identificando los principales riesgos asociados a esta tecnología, a saber, éticos, sociales, económicos, políticos y militares, destacando su carácter estructural y global, en tanto relacionados con el diagnóstico de la modernidad previamente realizado. El potencial deshumanizante de la IA intensifica la necesidad de abordarla críticamente.

El capítulo 3 parte de las aportaciones de Thomas Fuchs, quien ya problematiza los procesos de deshumanización asociados a la inteligencia artificial y propone un *humanismo encarnado* como alternativa. Si bien este planteamiento constituye un punto de partida valioso, sobre todo en su énfasis en el potencial deshumanizante de la tecnología, se argumenta que su fundamentación ontológica resulta operativamente problemática. Por ello, se plantea un desplazamiento desde la ontología hacia la pragmática.

La propuesta es un *humanismo crítico* que, lejos de recurrir a esencialismos, se construye sobre dos pilares: la noción kantiana de libertad en sentido práctico como base para la ética, y el sufrimiento como criterio negativo de lo que debe ser evitado. Este humanismo busca construir lo *bueno*, pero bajo el entendido de que la universalidad de tal construcción es imposible en la práctica; lo que lo caracteriza es la aceptación de la contingencia de cualquier contenido a este respecto y la apertura a la reconstrucción crítica tanto de su contenido como de sus propios fundamentos. El humanismo crítico no es contenido, sino categoría. Su horizonte normativo no es abstracto, sino que se articula con la crítica social, abriendo un espacio para disputar colectivamente aspectos tan relevantes como el desarrollo tecnológico. Esta propuesta permite, así, guiar el desarrollo de la inteligencia artificial en función de su impacto potencial sobre la dignidad humana, donde los riesgos que se le asocian funcionan como negativos de utopías en las que el contenido de lo humano se hace presente en tanto explicitan la conciencia de su potencial deshumanizante.

Finalmente, en el capítulo 4 se aplica el marco elaborado en apartados previos a una crítica directa a diferentes imaginarios con fundamento acrítico que rodean a la IA. En el plano sociotécnico se analizan las narrativas de inevitabilidad, necesidad/deseabilidad y liderazgo para la gestión del riesgo, mostrando la manera en que se constituyen como mecanismos de poder que buscan clausurar la crítica, desactivar el sentido de agencia social y naturalizar un desarrollo tecnológico realmente orientado por intereses corporativos y estatales. Se evidencia la no neutralidad de estas narrativas, dilucidando los sesgos y tergiversaciones que configuran la manera en que se permite pensar respecto a la inteligencia artificial.

En un plano aún más intrínseco, se examinan las corrientes trans- y posthumanistas que, fuera de la filosofía académica y en un ámbito más popular y mediático, presentan a la inteligencia artificial como una vía hacia la superación o sustitución de lo humano, y que reproducen la misma razón instrumental que reduce al ser humano a un medio, poniendo en juego su agencia y su dignidad. Estas visiones son mantenidas explícitamente en directrices operativas que guían hoy el avance de la IA, y, sin embargo, están construidas sobre presuposiciones cuestionables acerca de la naturaleza humana que se enmascaran como resultados científicos. Ante esto, la crítica se construye alrededor de la ilegitimidad operativa de una metafísica encubierta destacando los riesgos que se derivan de su implementación.

Si bien la mayor parte del trabajo consiste en desmontar estos imaginarios acríticos, también se esboza propositivamente el fundamento para repensar la inteligencia artificial desde el propio humanismo crítico. Este enfoque no pretende clausurar su desarrollo, sino orientarlo a fines que afiancen la libertad, fortalezcan la dignidad humana y reduzcan el sufrimiento. La apuesta es, en última instancia, recuperar la capacidad colectiva de decidir sobre el rumbo de la técnica, reconociendo que la IA, como toda construcción moderna, puede ser también un espacio de deliberación sobre el futuro si se le somete al escrutinio crítico y se vincula a proyectos éticos y sociales más amplios.

Capítulo 1: El Diagnóstico de la Modernidad de Anthony Giddens y Ulrich Beck

El proyecto de la modernidad, gracias a su carácter expansivo, ha penetrado y reformado gran parte de las sociedades a nivel global, con distintos devenires y experiencias, pero con un marco común: los modos de vida y organización social surgidos en Europa alrededor y a partir del siglo XVII como base. La idea principal que subyace a este proyecto es la Ilustración, esto es, la confianza en la razón, el conocimiento científico y la planificación técnica para construir la civilización futura ideal, una donde el ser humano se erige por sobre la naturaleza como consecuencia de un incesante progreso, y donde la razón, en este sentido, permite a lo creado ser superior a lo dado. Es una emancipación total del ser humano respecto del mundo al que fue arrojado: al desencantarse este último, el mito deja de ser necesario.

Sin embargo, las promesas de este proyecto no han sido alcanzadas. La razón que se erige como bandera ha sido patologizada, instrumentalizándose bajo una lógica de medios-fines, de lo que han derivado relaciones de poder, exclusión y dominio. Se establecieron el capitalismo, la industrialización y la burocracia, y con ellos la explotación del ser humano y de la naturaleza, así como una vigilancia y capacidades militares inimaginables en los estados premodernos, y que han sido un tema central en el siglo pasado y en el presente, donde el carácter mundial de las guerras llega a su máxima expresión, pues sus potenciales consecuencias suponen una amenaza verdaderamente global.

Los efectos de la modernidad son, pues, como la modernidad misma, globales. Este es un elemento constitutivo de la modernidad que se acrecienta en su etapa tardía. Asimismo, el supuesto desencantamiento del mundo no ha sido tal. El vacío metafísico resultante de la *muerte de Dios* ha sido llenado con discursos que dotan de un aura divina a la ciencia y a la tecnología; una devoción ante aquello que promete salvarnos y darnos una vida mejor, lo que dificulta un abordaje crítico ante tales circunstancias.

Así, pues, de la modernidad emergen consecuencias no deseadas, presumiblemente imprevistas aunque ahora son completamente evidentes. La amenaza a la civilización se hace factible en muchos sentidos: una guerra nuclear, el agotamiento de los recursos naturales, la contaminación presente a nivel mundial e inescapable, o también la alienación mediante una vigilancia y control absolutos que merma toda noción de libertad y agencia humanas. A manera de reacción cambia también la dinámica social: se constituye la conciencia del riesgo del avance técnico desmedido, se pierde la confianza en la promesa de progreso y se genera una politización del desarrollo científico y tecnológico, así como reacciones adaptativas, y desde todo ello se busca contrarrestar los efectos de las contingencias modernas.

En este sentido, el diagnóstico de la modernidad no supone un mero trabajo descriptivo, sino que posee un carácter crítico. Entender las perversas consecuencias de la modernidad es, en sentido negativo, pensar sobre una noción de la vida que merece ser vivida y de las estructuras sociales que no lo permiten, así como de las ideologías que les subyacen y las perpetúan. Dentro de este marco, es menester la recuperación del sentido de agencia social a partir de la cual se promueve el debate crítico sobre las nuevas tecnologías y la evitación del progreso desmedido. Para el presente trabajo, ello se particulariza en el caso de la Inteligencia Artificial, donde tomando como base sus riesgos se busca orientar su desarrollo según se delibere razonable.

Asimismo, el abordaje de la modernidad como fenómeno global no se realiza con una pretensión homogeneizante. Que la modernidad se ha expandido de manera global en el mundo occidental y que los riesgos que de esta se derivan son inescapables no es una perspectiva eurocentrista sino realista. Las experiencias de la modernidad son múltiples y variadas, pero comparten todas una base. Y por este carácter global de su ser y consecuencias es que se torna de particular relevancia describirla y abordarla críticamente, y a esta tarea es a lo que se aboca el primer capítulo.

1.1: La Modernidad Tardía y la Radicalización de sus Consecuencias

Del proceso de modernización se desprenden consecuencias que han modificado enormemente las dinámicas de la sociedad, como es el caso de la transición hacia una sociedad global, inestable y contingente, que deriva en enfrentamientos entre seguridad y peligro, fiabilidad y riesgo, y confianza y desconfianza, surgidos a raíz de las promesas y los resultados del proyecto de la modernidad y que son llevados a sus extremos en su etapa tardía. Estas son consecuencias radicalizadas de la modernidad, y el análisis de su devenir será el tema que ocupa la primera sección del presente capítulo.

Anthony Giddens (1994, p. 15) realiza una primera aproximación a la modernidad definiéndola como “los modos de vida y organización social que surgieron en Europa desde alrededor del siglo XVII en adelante y cuya influencia, posteriormente, los han convertido en más o menos mundiales”. Las instituciones de la modernidad poseen características propias que permiten diferenciarles sustancialmente de las presentes en sociedades premodernas; además, estos modos de organización no se encuentran circunscritos a un espacio determinado, sino que establecen interconexiones sociales a escala global.

Esta definición explicita una base compartida en aquellas sociedades donde la modernidad ha penetrado con sus bases organizativas, tanto estructurales como prácticas, a tal grado que es posible situarlas en un plano a parte de las sociedades premodernas, pero también sugiere una flexibilidad respecto al devenir de moderno en distintos contextos sociohistóricos, de manera tal que a pesar de la globalización de sus características, las experiencias que finalmente tienen lugar respecto a la modernidad pueden ser distintas en diferentes tiempos y espacios.

Para dar contenido a los modos de organización social compartidos por las sociedades modernas, Giddens (1994) describe cuatro agrupamientos organizativos que, complementándose entre sí, se conforman como las dimensiones institucionales de esta época: a

saber, el capitalismo, la industrialización, la vigilancia, y el control de los medios de la violencia y poderío militar sin precedentes.

Respecto al capitalismo, este es entendido como un sistema de producción de mercancías, dirigido a mercados competitivos y basado en la relación entre propiedad privada de capital y mano de obra asalariada, el cual desemboca en un sistema de clases. El industrialismo, por su parte, hace referencia a la transformación de la naturaleza mediante la utilización de maquinaria y con finalidad de producción. Por su parte, la vigilancia es la supervisión de las actividades de índole política de la población. Finalmente, el control de los medios de violencia, particularmente de los bélicos en su estadio más desarrollado, es también una característica particular del estado moderno (Giddens, 1994).

Como se ha anticipado, existe una interrelación entre estas cuatro dimensiones institucionales que dan forma a la modernidad. En primer término, es evidente que los procesos de producción intrínsecamente ligados al capitalismo se facilitan por la industrialización, la cual permite acelerar y automatizar la transformación de los recursos naturales y generar bienes; pero, además, esta industrialización funge como el casi único medio de interacción del ser humano con la naturaleza bajo una lógica de medios-fines que acrecienta la explotación desmedida y retroalimenta al sistema del capital. Por otra parte, las sociedades capitalistas se establecen con la naturaleza expansiva de este modelo económico que se extiende internacionalmente pero se particulariza dentro de las fronteras. Uno de sus éxitos es el grado de coordinación administrativa al que este sistema ha llegado y que depende del desarrollo de una capacidad de vigilancia sin precedentes deviniendo en un sistema de control único. Asimismo, la industria se somete al monopolio de la violencia por parte del estado, permitiendo industrializar la guerra, donde esta pasa a ser de gran interés económico. El capitalismo, además, aísla a la política de la economía y condiciona, a su vez, al estado (Giddens, 1994).

Las anteriores son, pues, características de la modernidad, y, en tanto tal, son mundializadas, pues la modernidad es intrínsecamente globalizadora. Ulrich Beck, según

explica Flores (2016), considera que la globalización desemboca en una convergencia cultural y social que supera los límites nacionales y que amenaza la soberanía nacional. Esto último puede darse en la medida en que diversos aspectos sociales dejan de circunscribirse a una localidad y pasan a ser dependientes de acontecimientos lejanos.

Al respecto, Giddens (1994) explica:

En la era moderna, el nivel de distanciamiento entre tiempo-espacio es muy superior al registrado en cualquier periodo precedente, y las relaciones entre formas sociales locales o distantes y acontecimientos, se *dilatan*. La mundialización se refiere a este proceso de *alargamiento* en lo concerniente a los métodos de conexión entre diferentes contextos sociales o regiones que se convierten en una red a lo largo de toda la superficie de la tierra. (p. 67)

El mencionado distanciamiento entre tiempo y espacio es una característica fundamental en el análisis del proceso de modernización encontrado en la obra de Giddens.

Las instituciones sociales de la modernidad, a diferencia de las tradicionales, se definen por una discontinuidad que se desmarca completamente de cualquier concepción evolucionista sobre la sociedad, donde la historia no posee ya una condición universal. Estas son, por una parte, la excepcional rapidez del *ritmo de cambio*, por otra, el *ámbito de cambio*, y, finalmente, *la naturaleza intrínseca de las instituciones sociales* (Giddens, 1994).

Sobre lo anterior, Giddens (1994) matiza que:

Las civilizaciones tradicionales pueden haber sido más dinámicas que otros sistemas pre-modernos, pero la celeridad del cambio moderno es excepcional. Quizás resulta más evidente en lo que respecta a la tecnología, pero puede extenderse igualmente a otras esferas. [...] La interconexión que ha supuesto la supresión de las barreras de comunicación entre las diferentes regiones del mundo ha permitido que las agitaciones de transformación social estallen prácticamente en la totalidad de la superficie terrestre. [...] Algunas formas sociales modernas, tales como el sistema político del Estado-nación

o la dependencia generalizada de la producción a partir de fuentes inanimadas de energía y la completa mercantilización de los productos y del trabajo asalariado, simplemente no se dan en anteriores periodos históricos. Otras solo poseen una aparente continuidad con los órdenes sociales anteriores. Un ejemplo es la ciudad. Los asentamientos urbanos modernos frecuentemente incorporan los emplazamientos de las ciudades tradicionales y pueden llegar a dar la impresión de ser meras extensiones de las mismas, pero de hecho el urbanismo moderno se ordena de acuerdo con principios muy diferentes a los que distinguieron a la ciudad premoderna del campo en periodos anteriores. (pp. 19-20)

Esta caracterización institucional se ha desarrollado en función de la tendencia al progreso material que supone la modernización, donde dichas instituciones y su expansión global proporcionan un aparente ámbito de oportunidad y seguridad al que no pudo asemejarse ningún sistema premoderno. La idea de control absoluto sobre la naturaleza eleva al ser humano a un nuevo plano donde ya no existe nada ajeno en la medida en que todo se encuentra bajo su control; asimismo, el desarrollo económico que de ello deriva se muestra como la materialización de una vida mejor. Sin embargo, de este proceso emergen también consecuencias sombrías no previstas o, más bien, ignoradas: ya desde Marx, Durkheim o Weber se vaticinaba, por ejemplo, la degradación humana producto del nuevo sistema de trabajo industrializado, pero nunca imaginaron el grado de daño medioambiental que esto mismo supondría; asimismo, fallaron en las previsiones sobre la posibilidad de resurgimiento de regímenes totalitarios que, además, se vieron reforzados por la concentración de poder político, militar e ideológico propia de la modernidad; además, ninguno de estos pensadores pudo haber anticipado la amenaza civilizatoria que supone la industrialización de la guerra, lo cual se hizo patente con el desarrollo de armamento nuclear (Giddens, 1994).

La modernidad, entonces, presenta unas consecuencias que, en la etapa tardía en que se encuentra el proceso de modernización, se han radicalizado a tal punto que de hecho se constituyen como eje para explicar las caracterizaciones sociales contemporáneas. En este

sentido, más allá de las beneficiosas posibilidades de la modernidad, que en un primer momento podrían haber ocupado el foco de atención, resulta imposible continuar ignorando sus efectos sombríos, incluso si ello conlleva replantearse la factibilidad misma de la modernización.

Es así que los conjuntos de *sistemas abstractos* y *señales simbólicas* propios de la modernidad generan una paradoja que enfrenta las oportunidades creadas por las instituciones sociales globales de esta época para mejorar las condiciones de existencia de los seres humanos, con el surgimiento y agravamiento de peligros derivados (Giddens, 1994). Lo anterior se explica por una característica que solamente se ha producido una vez en la historia de las sociedades: la separación y recombinación entre tiempo y espacio. Farfán (1999, p. 52) profundiza al respecto mencionando que las personas tienen que “aprender a vivir en una sociedad que los vincula con muchos otros, a pesar de que no los conozca ni los reconozca, y a manejar un tiempo vacío, no sustancial, que determina (a través de horarios) todo lo que hacen”.

Esta separación entre tiempo y espacio posee tres implicaciones fundamentales: en primer lugar, el *vaciamiento* del tiempo y del espacio desvincula la actividad social y su *anclaje* en las condiciones contextuales particulares donde tal actividad se presenta; en segundo lugar, da cabida a los mecanismos productores de la organización racionalizada; finalmente, la moderna historicidad radical solo puede insertarse dentro de esta particular expresión de tiempo y espacio (Giddens, 1994).

Como ya se esbozaba en el párrafo anterior, de la separación entre tiempo y espacio se desprende el desanclaje de los sistemas sociales, esto es, la desconexión de las relaciones sociales respecto a su nivel local de contextualización, para ser reorganizadas en intervalos espaciotemporales no concretos. Este desanclaje tiene lugar a partir de señales simbólicas, definidas por Giddens (1994; 32) como “medios de intercambio que pueden ser pasados de unos a otros sin consideración por las características de los individuos o grupos que los manejan en una particular coyuntura”, o a través de sistemas expertos, entendidos como “sistemas de logros

técnicos o de experiencia profesional que organizan grandes áreas del entorno material y social en el que vivimos” (Giddens, 1994, p. 37).

Lo que conecta a ambos mecanismos de desanclaje es la eliminación del ámbito contextual inmediato de las relaciones sociales que admiten. Además, se cimentan sobre una base de *fiabilidad* conferida a capacidades abstractas. Este término hace referencia a una confianza a manera de expectativas depositada sobre una persona o sistema, de manera tal que a pesar incluso del conocimiento sobre las circunstancias de riesgo relacionadas con dicha persona o sistema se establece una fe racional en su correcto desempeño a manera de mecanismo defensivo contra las contingencias (Giddens, 1994).

El concepto de fiabilidad es fundamental en este sentido, pues constituye la forma de afrontar los riesgos propios de la sociedad moderna. Los sistemas expertos permiten riesgos calculados, pero para Beck (1998b) en un punto se vuelven incalculables. Esta concepción de riesgo es fundamental para Beck, pues considera que dicho elemento caracteriza a un tipo particular de sociedad, en la que el crecimiento de las fuerzas productivas en el proceso de modernización que cimentan el progreso económico se acompaña por una producción social de riesgos. Esta transición de la *sociedad de la carencia* hacia la *sociedad del riesgo* fundamenta dinámicas sociales propias de la modernidad tardía.

Así, pues, en el paradigma de la sociedad del riesgo sus miembros buscan evitar, minimizar, canalizar, repartir y limitar los riesgos y peligros derivados del progreso técnico basado en la industrialización, con la finalidad de permitir la continuidad del proceso modernizante, para lo cual los riesgos deben ser soportables (Beck, 1998b).

Para aprender a vivir ante los males derivados del progreso técnico (contaminación, amenaza nuclear, agotamiento de las fuentes de energía) los seres humanos se valen de una *seguridad ontológica* como sentido de orden, seguridad y continuidad ante los riesgos inherentes al desarrollo industrial y sus vertiginosos cambios derivados. Además, dado que los

riesgos no pueden ser calculados en su totalidad, se aplica un margen de seguridad en el que se considera que los peligros han sido suficientemente minimizados. (Farfán, 1999).

Ante una cotidianidad condicionada por los riesgos, las sociedades modernas adquieren un carácter reflexivo. Las sociedades postindustriales son descritas por Farfán (1999) como sigue:

Son sistemas complejos en los que el umbral de contingencia es tan alto que abre una gama imprevista de posibilidades que tiende a incrementar el nivel de apertura del futuro presente. Es decir, las sociedades del riesgo son aquellas en las que las posibilidades de futuro son tan amplias que desbordan sobradamente los medios para reducir un horizonte excedente de posibilidades, de tal suerte que están obligadas a generar medios de reducción de la contingencia a través de una reducción de lo improbable. Para los miembros de estas sociedades, el umbral tan alto de contingencia los obliga a tomar decisiones, todos los días, que los enfrentan a múltiples opciones, lo cual les produce estados de inestabilidad psíquica, pero también los obliga a formar una identidad en la que actúa de forma recurrente la reflexividad de la sociedad. (p.53)

De esta manera, los sujetos que experimentan constantemente las consecuencias de la modernización comienzan a problematizar este mismo proceso. En este sentido, las cuestiones sobre desarrollo científico y tecnológico son sustituidas por debates sobre la gestión de los riesgos derivados de dichos desarrollos. El carácter político de la opinión pública se torna fundamental en la medida en que debe ratificarse ante esta la promesa de seguridad ante los riesgos (Beck, 1998b).

De la modernidad se desprende, entonces, una caracterización de la sociedad que acaba radicalizándose al punto de dejar de constituir un sistema estable, y ante este nuevo estatus de contingencia la reflexividad moderna se convierte en escrutinio y crítica constantes sobre el proceso de modernización por parte de aquellos que se encuentran inmersos en el mismo. Esta reflexividad se observa en la búsqueda orientada a recobrar la confianza en los progresos de

presente y futuro con la adquisición de una conciencia sobre la inestabilidad y contingencia propias de la modernización. Tales características son consecuencias de la modernidad que se evidencian en su etapa tardía dada su radicalización, y su estudio resulta crucial para comprender la influencia del progreso modernizante sobre las dinámicas sociales, al igual que sus efectos sobre las acciones individuales.

En el siguiente apartado se profundizará sobre las condiciones que llevan a la constitución de la conciencia sobre los riesgos derivados de la modernidad, y cómo esta da pie a un conjunto particular de características sociales y reacciones adaptativas que se engloban bajo el término *sociedad del riesgo*.

1.2: El Paradigma de la Sociedad del Riesgo

Como fue presentado en el apartado anterior, el carácter reflexivo propio de la modernidad tardía supone que los individuos inmersos en estas sociedades reconocen las implicaciones del propio proceso de modernización. Esto es consecuencia de la incertidumbre derivada de la discontinuidad de los fenómenos cotidianos tal como se presenta en la modernidad, así como del reconocimiento de los riesgos derivados del progreso técnico desmedido, lo cual desemboca en una desconfianza sobre la marcha de la modernización fundamentada en la toma de conciencia sobre los efectos negativos que de ello se desprenden.

La dinámica social basada en el riesgo es tan crucial para Beck (1998b) que desarrolla la idea de un paradigma social que gira en torno a ello, al cual nombra como *sociedad del riesgo*. Este paradigma se fundamenta en la producción social de riesgos inherente a la producción social de la riqueza en las sociedades industrializadas. La problemática central de este tipo de sociedades pasa a ser, en palabras del propio Beck (1998b):

¿Cómo se pueden evitar, minimizar, dramatizar, canalizar los riesgos y peligros que se han producido sistemáticamente en el proceso avanzado de modernización y limitarlos y

repartirlos allí donde hayan visto la luz del mundo en la figura de *efectos secundarios latentes* de tal modo que no obstaculicen el proceso de modernización ni sobrepasen los límites de lo *soportable* (ecológica, médica, psicológica, socialmente)? (p. 26)

En otras palabras, las sociedades del riesgo se centran en la búsqueda de mecanismos de afrontamiento (sociales e individuales) ante los riesgos derivados del progreso técnico tardomoderno. Por tanto, resulta fundamental definir qué es el riesgo y explicar el surgimiento de este carácter, particular a la modernidad, de conciencia y reacción ante este.

Como ya se ha explicado anteriormente, las sociedades del riesgo se establecen en el marco de la modernidad tardía, especialmente en los estados más ricos y desarrollados de occidente, donde el paradigma de *sociedad de la carencia* se reduce a un segundo plano en la medida en que la lucha por la satisfacción de las necesidades primarias deja de ser un problema cardinal del gran grueso de la población gracias a las riquezas obtenidas mediante el desarrollo industrial. Así, donde anteriormente la mitigación del hambre justificaba cualquier tipo de desarrollo que se orientara hacia la producción de la riqueza y reducción de la carencia, incluidas las consecuencias secundarias que de ello se derivaran, ahora, una vez superadas estas problemáticas básicas, se sustrae la base legitimadora del crecimiento industrial desmedido, que desemboca en una reflexión sobre las amenazas de los efectos secundarios derivadas de unas fuentes de riqueza contaminadas. Asimismo, se modifica la lógica del reparto, pues mientras en la búsqueda de la riqueza social existen recursos escasos a los que se aspira, en la sociedad del riesgo existe una sobreabundancia de peligros que hay que impedir, transicionando de la lógica positiva de la apropiación a la lógica negativa de la reducción (Beck, 1998b).

El riesgo, para Beck, hace referencia a la incertidumbre incalculable. Los riesgos propios de la modernidad son novedosos y extremadamente difíciles de calcular y, por tanto, no existe la posibilidad de preestablecer una respuesta ante estos, quedando a merced de la incertidumbre del futuro (Sørensen, 2018). Los riesgos son, en este sentido, una consecuencia de la modernidad radicalizada en su etapa tardía, pues se sobrepasa la calculabilidad de los riesgos

mediante los sistemas expertos, hecho que, explicaba Giddens, tuvo lugar en la modernidad temprana. Son, también, producto de la actividad del ser humano. Además, los riesgos se encuentran siempre insertos en una relación de conocimiento: existe un núcleo objetivo en los riesgos de la modernidad, pero se caracterizan como tal primero en la percepción social; esto quiere decir que no es posible separar al riesgo en sí mismo de su percepción pública.

La transición de la sociedad de la carencia a la sociedad del riesgo implica, entonces, el reconocimiento social generalizado de las fuerzas destructivas liberadas por las fuerzas productivas. Lo que particulariza a los riesgos derivados del desarrollo industrial moderno es el hecho de que, por primera vez en la historia, las consecuencias del progreso técnico llevado a cabo por el ser humano adquieren un carácter global: las situaciones de amenaza afectan a toda la humanidad y demás vida sobre el planeta, al extremo de que la destrucción total del mundo se convierte en un escenario factible. Además, a diferencia de los riesgos del pasado, los modernos se sustraen a los sentidos y requieren de la ciencia como *órgano perceptivo*, es decir, no son inmediatos ni se vinculan directamente a experiencias primeras, lo cual complica su comprensión y eleva el grado de incertidumbre a la vez que reduce su abordaje al ámbito del conocimiento especializado. También, se originan ya no de un infraabastecimiento de los medios para el control de los elementos externos, sino de la sobreproducción propia de la modernidad. “Son riesgos de la modernización. Son un producto global de la maquinaria del progreso industrial y son agudizados sistemáticamente con su desarrollo ulterior” (Beck, 1998b, p. 28).

En este sentido, Beck (1998b), en función de la arquitectura social y la dinámica política de los riesgos civilizatorios, establece cinco tesis fundamentales sobre el riesgo tardomoderno:

1. Los riesgos, en tanto sustraídos a la percepción humana, solo pueden ser establecidos mediante el saber, y, por tanto, son altamente susceptibles a la transformación discursiva y los procesos sociales de definición, por lo que los medios y posiciones de definición del riesgo se convierten en posiciones sociopolíticas de gran valor.

2. Los riesgos de la modernización escapan eventualmente a las dimensiones de desigualdad y condiciones de clase, pues se reparten de manera global, afectando tarde o temprano a quienes los producen y se benefician de ellos.

3. La expansión de los riesgos eleva a un nuevo nivel la lógica del desarrollo capitalista en la medida en que constituyen una fuente de necesidades sinfín, de manera que la economía en los riesgos se torna autorreferencial.

4. En las situaciones de peligro, la conciencia determina al ser. El saber adquiere un significado político, de manera que el potencial político se encuentra en el surgimiento y difusión del saber de los riesgos, que son asignados en términos civilizatorios.

5. Los riesgos socialmente reconocidos devienen en una discusión política explosiva. Se aborda la factibilidad de la supresión de las causas en el mismo proceso de industrialización.

Ante el crecimiento y reconocimiento de los riesgos lo hace también la promesa de seguridad, que ha de ser ratificada continuamente ante el escrutinio crítico de la opinión pública, ya sea a través de intervenciones reales o cosméticas. La politización de la producción de riesgos conlleva una transición de las discusiones sobre el desarrollo y aplicación de la tecnología al de la gestión de sus riesgos derivados de acuerdo con unos horizontes de relevancia también contruidos socialmente (Beck, 1998b).

Resulta conveniente recordar que el conocimiento y, por tanto, las reacciones sobre los riesgos se encuentran mediados argumentativamente, sobre todo por las interpretaciones derivadas del conocimiento científico. A través de esta discursividad, en función de la lógica negativa del riesgo, se busca su eliminación, negación o reinterpretación. De ello se derivan estrategias que intentan minimizar la percepción del riesgo mediante la tergiversación de datos científicos, haciendo referencia a términos y explicaciones que ofuscan las consecuencias reales según conveniencia para dar continuidad desenfrenada al avance técnico industrial (Beck, 1998b).

Sin embargo, la vivencia de los riesgos sigue estando presente y la pérdida de confianza en la promesa del progreso se hace manifiesta aún ante las estrategias de ocultación superficial, pues al fin y al cabo ya han existido ejemplos materializados del peligro que forman parte del imaginario colectivo, donde la amenaza deja de ser solamente tal y se transforma en realidad, y cuya aparición no puede sino considerarse indeseable. Pero incluso los intentos por maquillar las consecuencias del avance industrial de acuerdo a intereses de desarrollo técnico toman como base la noción de una vida que merece ser vivida, a saber, se maquilla en función de una imagen ideal, de un horizonte normativo compartido.

En este sentido, los riesgos son negativos de utopías en que lo humano (o lo que queda de ello) se conserva en el proceso de modernización y vuelve a ser animado. Pese a la desfiguración, en última instancia este horizonte normativo en el que se hace visible lo arriesgado del riesgo no puede ser suprimido mediante la matematización o la experimentación. Tras todas las objetivaciones aparece más tarde o más temprano la cuestión de la *aceptación* y por tanto una vieja cuestión nueva: *¿cómo queremos vivir?* ¿Qué es lo humano en el ser humano y lo natural en la naturaleza que hay que conservar? (Beck, 1998b, p. 34).

De esta manera, la conciencia del riesgo dirige a retomar las discusiones sobre normatividad social y humana. Se manifiesta como recuperación de las nociones éticas, filosóficas, políticas y culturales en las instituciones y las disciplinas técnicas. Es una búsqueda por definir y negociar el riesgo que tiene lugar a manera de un intrincado debate entre las distintas esferas de la sociedad, y donde el conocimiento delimita las posturas a tomar.

Como consecuencia de lo anterior, se genera una ruptura en el dominio de la racionalidad científica. La racionalidad social ve elevada su relevancia en tanto todos los sectores involucrados de una u otra forma en las situaciones de riesgo se ven inmersos en los debates por la definición. La ciencia, entonces, se politiza, dejando atrás un fundamento puramente lógico-experimental para ocuparse de las cuestiones éticas, políticas y económicas

inherentes a los resultados de su actividad y aplicación. Las disputas entre ambos tipos de racionalidades quedan claramente segmentadas en los términos y las problemáticas que plantean y son mutuamente restrictivas. Sin embargo, de la complementación mediante esta lucha resultan las definiciones de riesgo que marcan la pauta para las decisiones ulteriores (Beck, 1998b).

Ahora bien, dado que los riesgos poseen un componente futuro, se encuentran relacionados con la previsión de consecuencias no deseadas que, sin embargo, parecen ser inminentes, lo cual desencadena la actuación en el presente. La fuente del debate es entonces una *variable proyectada* determinada en el presente, cuyo contenido de amenaza establece la relevancia de actuación inmediata ante ello (Beck, 1998b).

Así, pues, ahora sabemos que “el mundo en que vivimos es espantoso y peligroso” (Giddens, 1994, p. 23). Como resultado, en el afrontamiento de un entorno inevitablemente plagado de riesgos que amenazan la entereza de la vida sobre el planeta, la cotidianidad se debate entre nociones de seguridad y peligro, y fiabilidad y riesgo, sobre todo en la medida en que estos se encuentran no solo fuera del control de los individuos, sino incluso de las grandes organizaciones y estados que los provocan. “La vivencia de los riesgos presupone un horizonte normativo de seguridad perdida, de confianza rota” (Beck, 1998b, p. 34).

De conformidad con lo anterior, Retamal (2018, p. 140) explica que el sentido de seguridad en las sociedades modernas “se desvanece a medida que el capitalismo expande reticularmente una nueva cultura de la transformación permanente, sin contrapesos de estabilidad”. Además, menciona este autor, la consecuencia directa de esto es una lucha por recobrar tal estabilidad. Por lo tanto, es de vital importancia la comprensión de los mecanismos utilizados para afrontar el endeble mundo moderno y para el emprendimiento de la búsqueda por recuperar la confianza que permita un marco de acción en un mundo dominado por el peligro y la incertidumbre.

Al respecto, Luhmann (2000) afirma que la confianza, incluso en estas circunstancias, no puede dejar de ser la norma:

No es posible vivir sin formarse expectativas respecto de las contingencias, y en alguna medida, deben rechazarse las posibilidades de quedar decepcionado, se rechazan porque solo representan una remota posibilidad, pero también porque no sabemos qué más podemos hacer. La alternativa sería vivir en un estado de permanente incertidumbre y prescindir de expectativas sin tener nada con que reemplazarlas. [Traducción propia] (p.96)

En este sentido, la confianza hace referencia a una actitud que presupone la estabilidad.

Así, explica Giddens (1994), en la práctica existe una imposibilidad de mantener como foco central de nuestra atención los peligros que se escapan a cualquier forma de control personal. Una ansiedad derivada de la preocupación continua por la amenaza de destrucción del mundo puede, incluso, ser considerada funcionalmente como un trastorno psicológico en tanto dificulta la realización adecuada de las actividades cotidianas, si bien no puede decirse que sea un miedo irracional.

De hecho, un mecanismo para evitar esta ansiedad tiene que ver con el desarrollo de una noción de fortuna donde, dadas las bajas probabilidades de ocurrencia de una catástrofe civilizatoria y las igualmente reducidas posibilidades de acción ante esta posibilidad, se produce una vaga sensación de confianza en la improbabilidad de tal suceso y se deslinda de responsabilidad personal, lo que alivia al individuo del peso de su relación con un hecho de tales características. En este sentido, pese al carácter secular de la modernidad, surge la noción infundada de un destino que guiará adecuadamente los acontecimientos, de alguna forma (Giddens, 1994).

Por tanto, en la búsqueda por recuperar la confianza perdida, el afrontamiento práctico cotidiano de los riesgos se basa en la construcción de una seguridad ontológica que proporcione un marco de acción presumiblemente seguro. Seguridad ontológica refiere a un constructo

cognitivo donde los seres humanos depositan la confianza en la continuidad de su autoidentidad y en la permanencia de sus entornos, sociales o materiales, de acción. Esta descansa sobre una noción de fiabilidad, es decir, sobre un tipo de fe en la predictibilidad y continuidad acerca de las personas y los sistemas. Las rutinas integradas en los sistemas abstractos son cruciales para la seguridad ontológica (Giddens, 1994).

Asimismo, continuando con Giddens (1994), opera una contrafacticidad respecto a los riesgos de baja probabilidad y graves consecuencias, como es el caso de una guerra nuclear que amenace la presencia de vida sobre la tierra, pues estos son irreales en la medida que su realización implicaría la destrucción del mundo (o no, lo que permitiría demostrar la exageración respecto a estos), por lo que, mientras este no sea el caso, permanecerán latentes. Por tanto, este tipo de riesgos no desaparecerá de la conciencia social, pues la única demostración de la invalidez de plantearlos como tal sería que de hecho se produjeran, derivando en la permanencia de un estado de conciencia del riesgo. Es una incesante lucha entre la facticidad y la contrafacticidad de los riesgos civilizatorios, e influye directamente en la constitución personal del sentido de seguridad.

Sin embargo, la seguridad ontológica opera en un nivel básico, generalmente inconsciente, que permite evitar la ansiedad provocada por el conocimiento de los riesgos del mundo con la finalidad de poder vivir con relativa tranquilidad la vida cotidiana. Del grado de seguridad ontológica dependen una serie de reacciones adaptativas que determinan las actitudes y acciones que se toman ante el riesgo, que serán descritas a continuación.

Una de estas reacciones adaptativas es la *aceptación pragmática*, consistente en una aproximación práctica enfocada al abordaje de los problemas de la vida cotidiana que de hecho se encuentran dentro del campo de agencia individual, aún si se trata de beneficios pasajeros, en la medida en que los fenómenos más complejos del mundo moderno escapan al control de cualquiera. Por otra parte, la reacción de *optimismo sostenido* mantiene la fe en el progreso sustentado por la razón, considerando que los riesgos derivados de este proceso pueden ser

mitigados, y afirmando la factibilidad de llevarlos a buen término. De manera opuesta, el *pesimismo cínico* supone una reacción pragmática mediante la participación directa en las ansiedades derivadas del peligro y su conciencia, de manera que con el uso del humor, el hastío y la ironía se neutralizan los efectos emocionales de las circunstancias. Finalmente, el *compromiso radical* se orienta a la práctica activa y directa de oposición a las fuentes de peligro mediante la acción social (Giddens, 1994).

Lo que todas estas reacciones tienen en común es una conciencia sobre los riesgos y una concepción negativa sobre los mismos. Esta es la característica que posibilita la reflexividad en la que desemboca la modernidad tardía, y en ella se esboza ya la necesidad de un horizonte normativo que permita la evaluación de los riesgos.

Sin embargo, advierte Beck que una manipulación desde la propia conciencia del riesgo es posible:

Con el crecimiento de los peligros surgen en la sociedad del riesgo *desafíos completamente nuevos a la democracia*. La sociedad del riesgo contiene una tendencia a un *totalitarismo legítimo en la defensa contra los peligros*, el cual con el pretexto de impedir lo peor crea algo aún peor. Los *efectos secundarios* políticos de los *efectos secundarios* civilizatorios amenazan al sistema político democrático en su sustancia, el cual cae de este modo en el dilema de o fracasar a la vista de los peligros producidos sistemáticamente o derogar principios democráticos básicos mediante *contrafuerzas* autoritarios, propios de un estado dictatorial. Solventar este dilema es una de las tareas esenciales del pensamiento y de la actuación democráticos en el futuro presente de la sociedad del riesgo. (Beck, 1998b, p. 88)

Así, pues, donde hay reflexividad busca su espacio la manipulación. La conciencia del riesgo permite vislumbrar lo humano y evitar lo que nos aleja de ello, pero es menester comprender que, si no se presta la suficiente atención, puede transformarse de herramienta emancipadora a un dispositivo de control más. La modernidad reflexiva es, en este sentido,

paradójica, pues se enfrenta a sí misma, a las consecuencias de su propio despliegue racional. Es por ello que no basta con una mera descripción de los riesgos, sino que al analizar la manera en que son contruidos, definidos y sostenidos por estructuras de poder se revelan nuevas formas de dominio a las que una sociedad agente debe enfrentarse desde un horizonte normativo distinto. La propia explicitación de la dinámica social basada en el riesgo ya permite comenzar a construir herramientas políticas para interrogar los fundamentos, los efectos y las posibilidades transformadoras del presente. El diagnóstico de la modernidad es, entonces, fundamento para la propia crítica al proceso de modernización.

Las consecuencias de la modernidad, finalmente, pueden expresarse en pocas palabras tal como lo expone Giddens (1994):

La modernidad es inherentemente globalizadora, y las inquietantes consecuencias de este fenómeno se combinan con la circularidad de su carácter reflexivo para configurar un universo de acontecimientos en el que los riesgos y los peligros adquieren un nuevo carácter. Las tendencias globalizadoras de la modernidad son simultáneamente extensivas e intensivas porque conectan a los individuos a los sistemas de gran escala como parte de una compleja dialéctica de cambio tanto en los polos locales como globales. Muchos de los fenómenos frecuentemente denominados postmodernos verdaderamente conciernen a la experiencia de vivir en un mundo en el que presencia y ausencia se mezclan en formas históricamente inéditas. El progreso se vacía de contenido mientras que la circularidad se afianza, y en un plano lateral, la cantidad diaria de información interna que implica vivir en *un mundo*, puede resultar abrumadora a veces.

Con todo, esta *no* es esencialmente la expresión de la fragmentación o de la disolución del sujeto dentro de un *mundo de signos* sin centro. Es un proceso de transformación simultánea de la subjetividad y de la organización social global que se da contra el inquietante telón de fondo de los riesgos de graves consecuencias. (pp. 164-165)

1.3: La Crítica Implícita en el Diagnóstico de la Modernidad

Etimológicamente la palabra *diagnosis* hace referencia a la capacidad de conocer y distinguir. Sin embargo, el término *diagnóstico* tradicionalmente se ha aplicado al ámbito médico como la capacidad de emitir un juicio sobre el reconocimiento de una enfermedad mediante la recolección, análisis e interpretación de datos. En este sentido, la propia idea de un *diagnóstico de la modernidad* sugiere análogamente la existencia de unos síntomas que evidencian la necesidad de evaluar un funcionamiento presumiblemente inadecuado de las sociedades modernas; de conocer una enfermedad que les azota, incluso.

El diagnóstico de la modernidad no es, pues, una tarea meramente descriptiva, sino también una lectura crítica del presente en la que se interrogan sus fundamentos y se evalúan sus consecuencias. Prevé un malfuncionamiento y busca señalar una desviación respecto a aquello que se considera como una *vida social saludable*, a partir de lo cual se posibilita proponer modelos de intervención o transformación.

A la modernidad subyace ideológicamente la Ilustración, y para Theodor Adorno y Max Horkheimer el diagnóstico es claro: los males presentes en las sociedades occidentales modernas provienen de una razón patologizada. La lógica de medios-fines que subyace a la racionalidad instrumental, característica emergente de la desviación ocurrida en el proyecto de la Ilustración, deviene en mecanismos excluyentes y alienantes, generando un sistema de dominio y producción de consecuencias no controladas. La razón instrumental no se orienta por fines éticos, sino por la eficacia, el cálculo y el control técnico. Promete ser la herramienta perfecta para el desencantamiento del mundo, uno donde el ser humano se convierte en amo y señor de la naturaleza y nada le sobrepasa. Así, el mundo se transforma en un objeto susceptible de dominio (Horkheimer y Adorno, 1998).

Esta razón instrumental se ve materializada en las racionalidades industrial y científica descritas por Beck, donde el crecimiento técnico desmedido que se les deriva ha dado lugar a

una sobreexplotación de los recursos naturales, contaminación global y, en última instancia, a la amenaza de una catástrofe civilizatoria. Todo ello tuvo lugar sin un horizonte normativo ético más allá de la propia confianza ingenua sobre las promesas del proyecto de modernización.

Por supuesto, esta caracterización de la modernidad no es, como digo, expresada por Beck con fines únicamente descriptivos, sino que se realiza con una mirada crítica, además de abocarse a la búsqueda de una política que pueda hacer frente a las problemáticas derivadas de la modernidad y reorientar la su curso.

La cuestión de la racionalidad e irracionalidad de la ciencia no solo se plantea en relación a la actualidad y al pasado, sino también al *futuro*. Podemos aprender de nuestros errores; esto también significa que siempre es posible *otra* ciencia. Es posible no solo otra teoría, sino otra *teoría del conocimiento*, es decir, otra relación entre teoría y práctica, así como otra *práctica* de esa relación. Podemos decir que la época actual es una época de contrahipótesis, si es que es cierto que la actualidad no es más que una hipótesis. Y de ahí aún no hemos salido. La *piedra de toque* que corresponde a esa empresa es algo público, aunque sin voz: el proyecto de la modernidad necesita de unos primeros auxilios. Está amenazado de ser ahogado por sus propias anomalías. Y la ciencia, en su actual forma, también está afectada. (Beck, 1998b, p. 234)

A la época moderna, entonces, “le toca asumir la función de su contrario: la superación de la tradición y del dominio de la naturaleza. Resulta que es amenaza y promesa de liberación de la amenaza y promesa que ella misma ha creado”; de ello se deriva una consecuencia práctica de particular importancia: “los riesgos se han convertido en el motor de la *autopolitización* de la sociedad industrial moderna; e incluso más, varían el *concepto*, la *localización* y los *medios de la política*” (Beck, 1998b, pp. 302-303).

Así, pues, el diagnóstico de la modernidad funge como fundamento para el abordaje crítico de este mismo proceso y el esbozo de una política acorde. Las condiciones de vida indeseables y no planificadas que han surgido como fruto de la modernización, a saber, una vida

colmada por la incertidumbre y desconfianza, así como la preocupación por las potenciales amenazas latentes, se tornan en un disgusto compartido de donde se vislumbra implícitamente un contenido sobre la vida digna.

Por supuesto, un juicio sobre las disfunciones sociales modernas requiere de un horizonte normativo con el cual ser contrastado para determinar el grado de malfuncionamiento existente. Para Beck, ello se esboza en la función de los riesgos como negativos de utopías:

Lo humano (o lo que queda de ello) se conserva en el proceso de modernización y vuelve a ser animado. Pese a la desfiguración, en última instancia este horizonte normativo en el que se hace visible lo arriesgado del riesgo no puede ser suprimido mediante la matematización o la experimentación. Tras todas las objetivaciones aparece más tarde o más temprano la cuestión de la *aceptación* y por tanto una vieja cuestión nueva: *¿cómo queremos vivir?* ¿Qué es lo humano en el ser humano y lo natural en la naturaleza que hay que conservar? Hablar de ‘catástrofe’, como es cada vez más habitual, es en este sentido la expresión exagerada, radicalizada, objetivizada, de que no se quiere este desarrollo. (Beck, 1998b, p. 34)

Y este horizonte es el que toma como base para su evaluación diagnóstica.

Asimismo, expresa una necesidad de reorientar la Ilustración:

Para mí la democracia técnica (o ecológica) es la utopía de una modernidad responsable, una visión de sociedad donde las consecuencias del desarrollo tecnológico y del cambio económico se debaten antes de tomar las decisiones cruciales. La carga de las pruebas de los riesgos y peligros futuros y de la actual degradación ambiental corresponderían a los perpetradores y no a los agraviados: del principio de que el contaminador *paga* al principio de que el contaminador *prueba*. Finalmente, debe establecerse un nuevo cuerpo de criterios de prueba, corrección, verdad y acuerdo en la ciencia y el derecho. Entonces, lo que necesitamos es nada menos que una segunda *Ilustración*, que abra

nuestras mentes, ojos e instituciones a la amenaza autoinfligida de la civilización industrial. (Beck, 1998a, p. 514)

De esta manera, comienza a dilucidarse una serie de criterios que sirven de base para efectuar evaluaciones de los riesgos asociados al progreso técnico característicos de la modernidad. En ello reside el valor del diagnóstico de Beck y Giddens: no es una mera construcción abstracta de una sociedad hipotética, sino que muestra la deslegitimación del progreso desmedido y proporciona un marco sólido dentro del cual se hace posible estudiar críticamente nuevas tecnologías, especialmente cuando estas prometen una disrupción sin precedentes y, por tanto, su potencial negativo puede asumirse incalculable e incontrolable.

Tal es el caso de la inteligencia artificial, cuya comprensión técnica total se encuentra relegada a un grupo extremadamente reducido de especialistas. Mientras tanto, en el ámbito público comienzan a preestablecerse discursos que buscan legitimarla al describirla como la tecnología del futuro que dará solución a todos los problemas de la humanidad, y el interés se ve materializado en las ingentes inversiones de capital que se realizan respecto al desarrollo de esta tecnología: se ha convertido en una carrera, y hay muchas partes interesadas en conseguir el primer lugar, incluso si para ello hay que ofuscar su factibilidad o sus *efectos secundarios*.

Aunque los discursos que rodean a la IA en medios oficiales y declaraciones de partes directamente involucradas se centran sobre todo en sus innumerables beneficios, la conciencia del riesgo generalizada comprende que hay algo más: una serie de riesgos y peligros latentes que pueden derivarse del crecimiento sin control dentro de este ámbito. La pregunta por cómo queremos vivir surge ante la amenaza que supone el crecimiento desmedido de tal tecnología y su desarrollo ha de transformarse en una cuestión política.

Esta problemática es el objeto de estudio primordial de la presente tesis, y su debida profundización se realiza en el capítulo 4. Por el momento, baste lo anterior como justificación para elegir el diagnóstico moderno de Giddens y Beck como marco teórico para analizar críticamente una tecnología tan aparentemente novedosa y disruptora como lo es la IA.

Capítulo 2: Concepto de Inteligencia Artificial

Como se ha elaborado previamente, del progreso técnico desmedido, símbolo de la modernidad, han emergido consecuencias indeseables que azotan a las sociedades modernas, y sus riesgos adquieren un carácter global que rebasa toda barrera geográfica sobre el planeta. Ante esto, la conciencia del riesgo que se le deriva pone en juicio al desarrollo tecnológico, y esto no puede ser la excepción en el caso de la inteligencia artificial, una tecnología que promete un cambio radical en la sociedad, con una carga simbólica y política extremadamente elevada.

El desarrollo de la IA, desde la crítica implícita en el diagnóstico moderno de Beck y Giddens, merece ser abordado seriamente en tanto su progreso implica potenciales riesgos de altísima relevancia. Para ello, la propuesta en el presente capítulo es una aproximación a la IA dividida en tres momentos diferenciados pero interrelacionados.

En un primer momento se realizará una conceptualización *tradicional*, en el sentido horkheimeriano, de la inteligencia artificial; es decir, una base descriptiva con pretensión de neutralidad y objetividad sobre lo que significa este término. Esta forma de delimitación es necesaria como punto de partida para reconocer los aspectos que desde el campo de las ciencias computacionales y las ciencias cognitivas se consideran constitutivos del terreno, además de posibilitar el análisis de sus discursos implícitos con potencial para ser posteriormente criticados y reconstruidos.

Una vez establecida la definición operativa, se analizará con una lógica crítica la evolución histórico-social de la IA. Este apartado busca mostrar que la inteligencia artificial no es solo un conjunto de herramientas técnicas, sino también una construcción situada que ha sido constituida por intereses económicos, militares y políticos, lo que ha dado lugar al establecimiento de la IA como institución sociotécnica en las sociedades modernas.

Finalmente, serán descritos los riesgos emergentes que acompañan al desarrollo de la inteligencia artificial. Ello juega un papel crucial para orientar posteriormente una crítica

elaborada, pues, en sintonía con Beck y Giddens, los riesgos no pueden ser comprendidos como fallas accidentales sino como consecuencias estructurales de un desarrollo tecnocientífico que escapa a las formas tradicionales de control y previsión.

Todo lo anterior busca ser, en suma, un riguroso análisis del concepto de inteligencia artificial que funja como base para desarrollar una problematización crítica de sus implicaciones.

2.1: Concepto Tradicional de Inteligencia Artificial

Para Boden (2016) la Inteligencia Artificial es la búsqueda por que las computadoras ejecuten las tareas hasta ahora solo posibles para la mente:

Algunas de estas (por ejemplo, el razonamiento) son normalmente descritas como *inteligentes*. Otras (por ejemplo, la visión) no lo son. Pero todas implican habilidades psicológicas –como la percepción, asociación, predicción, planeación, control motor– que posibilita a los humanos y animales a conseguir sus metas. La inteligencia no es unidimensional, sino un espacio rico en su estructura con capacidades diversas para el procesamiento de información. En concordancia, la IA utiliza distintas técnicas para abordar distintas tareas. [Traducción propia] (p. 1)

Las aplicaciones prácticas de esta tecnología son diversas; “está en todas partes” [Traducción propia], afirma Boden (2016, p. 1): desde el hogar, los medios de transporte, los sistemas financieros, los de salud, el internet, el entretenimiento, hasta los robots que son enviados fuera del planeta o los satélites que orbitan en el espacio. Pero también lo está en aplicaciones bélicas y sistemas de vigilancia y control.

La Inteligencia Artificial tiene dos objetivos principales:

Uno es *tecnológico*: usar computadoras para completar tareas útiles (a veces empleando métodos muy poco parecidos a aquellos presentes en las mentes). El otro es *científico*: usar los conceptos y modelos de la IA para ayudar a responder preguntas sobre los seres humanos y otras formas de vida. [Traducción propia] (Boden, 2016, p. 2)

Así, pues, esta tecnología no solo representa un pilar fundamental de acuerdo a sus aplicaciones prácticas, sino que también la investigación asociada a esta puede conferir argumentos para un planteamiento filosófico (o ideológico) reforzando determinados discursos relativos a la naturaleza del ser humano.

En esta misma autora la idea de Inteligencia Artificial se expresa desde una base funcionalista que, como se hará evidente en capítulos posteriores, es un tópico recurrente en los discursos que rodean a esta tecnología. En este sentido, pensar en Inteligencia Artificial no implica necesariamente pensar en las computadoras en tanto máquina física sino en la máquina virtual, esto es, en el sistema de procesamiento de información que el desarrollador tiene en mente cuando escribe el programa, y que las personas tiene en mente cuando lo usan (Boden, 2016). Por tanto, la clave no está en lo que es, sino en lo que hace. No es el carácter ontológico sino la función lo que ocupa el plano principal.

A la propia idea de Inteligencia Artificial subyace, entonces, el funcionalismo. La ejecución de tareas y resolución de problemas es el hecho verdaderamente importante, y los ordenadores no son sino la base donde la función tiene lugar. La base, evidentemente, es necesaria, pero no lo es de un tipo en específico, es decir, pueden existir diferentes clases de circuitos donde se ejecutan las máquinas virtuales. Ello es crucial pues se relaciona directamente con la afirmación materialista de que el cuerpo humano no posee un estatus especial en tanto las características psicológicas no se encuentran intrínsecamente ligadas a este, sino que surgen como interfaz a partir de circuitos neuronales que funcionan de manera similar a cualquier otro

tipo de máquina y que, por ende, pueden ser reemplazados. Boden (2016, p. 5) reafirma esta idea exponiendo que “*la mente humana* puede ser entendida como la máquina virtual –o más bien, el conjunto de máquinas virtuales que interactúan mutuamente, corriendo en paralelo (y desarrolladas o adquiridas en momentos diferentes)– que se implementa en el cerebro” [Traducción propia].

La noción de una *máquina inteligente* no es nueva: la idea se encuentra presente por lo menos desde los autómatas de Herón de Alejandría. Sin embargo, en su forma moderna esta idea se remonta a la década 1840, donde Ada Lovelace en sus estudios sobre símbolos y lógica vislumbró en una máquina dedicada al álgebra y a los números, diseñada por Charles Babbage, el potencial para representar con esta todo el contenido del universo, incluyendo la inteligencia. La máquina se materializó de forma definitiva con las investigaciones de Alan Turing, quien demostró que toda computación posible puede ser ejecutada por un sistema matemático binario. Turing buscaba que estas novedosas máquinas realizaran tareas que requerían de *inteligencia*, pero también que modelaran los procesos que ocurrían en las mentes de naturaleza biológica (Boden, 2016).

El artículo de 1950 donde Turing propuso su famosa prueba tenía la intención de ser un manifiesto de la inteligencia artificial. En este se identificaban preguntas clave sobre el procesamiento de la información involucrado en la inteligencia, proveyendo pistas tentativas sobre lo que se había logrado ya. La convicción sobre la posibilidad de la IA había sido reforzada desde los años 40 cuando Warren McCulloch y Walter Pitts conectaron el trabajo de Turing con la lógica proposicional de Russell y la teoría de la sinapsis neuronal de Sherrington: tanto la computación como la neurofisiología convergían en su naturaleza binaria. “Todo lo que pudiera ser afirmado en lógica proposicional podía ser computado por una red neuronal, y por una máquina de Turing” [Traducción propia] (Boden, 2016, p. 10). La neurofisiología, la lógica y la computación formaron un conjunto al que posteriormente se uniría la psicología. El lenguaje natural, según consideraban estos pensadores, era reductible a la lógica, por lo que todo

razonamiento y opinión debería someterse a esta misma norma. Previeron que en última instancia la psicología sería resuelta al tener la capacidad de explicar una conducta desde la red neuronal que le corresponde.

Así, se creó un paradigma teórico cuya praxis fue limitada por lo primitivo de la tecnología disponible (Boden, 2016), pero su legado perduró en la informática posteriormente masificada y también a manera de fundamento en las Ciencias Cognitivas.

Como explica Gardner (1987), en 1948 tuvo lugar el famoso simposio de Hixon, donde ante la predominancia de las teorías conductistas y psicoanalíticas se propuso un cambio de paradigma como alternativa. Fue un momento crítico en la historia de la ciencia donde, sin hacer referencia explícita a Turing en las actas, sus ideas sí posibilitaron el trasfondo conceptual para lo que posteriormente se constituiría como Ciencias Cognitivas. La noción de que procesos mentales pueden ser formalizados, simulados o modelados computacionalmente es fundamental para las ideas de cibernética, computación y arquitectura cerebral que fueron propuestas. En este encuentro se sembró la idea de una máquina lógica capaz de reproducir cualquier proceso cognitivo, y del área de la ciencia que se encarga de ello.

Este campo se consolidó a mediados del siglo XX y, considero, su influencia sigue siendo predominante. El supuesto de que la mente es una forma de procesamiento de la información sigue orientando los desarrollos científicos y tecnológicos como extensiones del mismo marco. En consecuencia, esta perspectiva moldea las concepciones sobre la naturaleza humana que son trasladadas a la sociedad en general y funge como guía para el desarrollo de políticas, salud mental, pedagogía, etc.

La Inteligencia Artificial es, sí, un desarrollo tecnológico, pero no se reduce a ello: es también una discusión filosófica y un imaginario social práctico.

2.2: El Devenir Contingente de la Inteligencia Artificial Hasta su Actualidad

De la conceptualización respecto a la Inteligencia Artificial previamente realizada puede concluirse, entonces, que esta no se reduce meramente a un desarrollo tecnológico funcional, sino que depende de una amplia cantidad de factores que determinan su conformación. Esto se encuentra en relación al potencial simbólicamente asociado con la idea de la máquina inteligente, y, como ya hemos visto, este potencial comienza a materializarse en el sinfín de aplicaciones prácticas para la vida cotidiana que posee, y de donde parten intereses derivados de la economía, pero, más preocupante aún, también con importantes usos con finalidades políticas.

Por ello, su devenir hasta la actualidad ha dependido de orientaciones discursivas, económicas y políticas. Los desarrollos tecnológicos son siempre contingentes y resultan de las transformaciones producidas a lo largo de múltiples trayectorias científicas y de la interacción entre múltiples actores y discursos, tema en el que se profundizará en el capítulo 4. En este sentido, para Jobin y Katzenbach (2023) es crucial entender críticamente cómo la inteligencia artificial está convirtiéndose en una poderosa entidad sociotécnica.

En el proceso de formación de la IA, la naturaleza mítica y disputada de su concepto ha servido como referencia semántica para negociar ciertos contenidos de la práctica científica que conllevan preocupaciones sociales y políticas. La noción de IA abarca, desde el aspecto tecnológico y de investigación, enormes ambiciones económicas y visiones del futuro, así como valores normativos y preocupaciones fundamentales. Tiene, por tanto, una cualidad organizativa (Jobin y Katzenbach, 2023).

Pero, ¿cómo ha llegado la IA a ser más que una discusión aislada? La inteligencia artificial del siglo XXI se ha institucionalizado no tanto por sus capacidades técnicas comprobadas, sino por la combinación de discursos, inversiones y expectativas que la rodean. Aunque su integración aún genera controversias y está lejos de consolidarse en una forma

definitiva, esta tecnología ya se asume como inevitable y necesaria en múltiples sectores. Su esparcimiento se debe a una confluencia singular de factores: debates científicos no resueltos, enormes recursos tecnológicos y económicos, políticas activas y un discurso público que fluctúa entre promesas y temores. La IA se instala no solo como una herramienta, sino como fuerza reorganizativa sobre las instituciones, infraestructuras y formas de vida, aún cuando sus efectos y sentido permanecen abiertos y en disputa (Jobin y Katzenbach, 2023).

La IA suele insertarse en el ámbito de las ciencias de la computación como su campo natural, pero su desarrollo no ha seguido una línea continua ni unificada; en su lugar, ha sido moldeada por múltiples trayectorias científicas y disputas internas, especialmente entre las aproximaciones simbólicas (centradas en la comprensión semántica) y las conexionistas (basadas en redes neuronales y cálculos paralelos). Hoy predomina el enfoque conexionista gracias a los avances en aprendizaje automático, pero esta hegemonía es reciente y no necesariamente definitiva. La evolución de la IA en el campo computacional ha dependido tanto de rupturas técnicas como de factores sociales, económicos y políticos que han condicionado las líneas de investigación a las que se les asigna un grado de prioridad más alto (Jobin y Katzenbach, 2023).

Por su parte, los medios de comunicación han desempeñado un papel clave en la institucionalización social de la Inteligencia Artificial, moldeando la percepción pública pero también incidiendo en la propia investigación científica, ámbito no exento de la influencia de estas narrativas. En general, en distintos países el discurso mediático tiende a resaltar logros tecnológicos y productos de la industria, otorgando más visibilidad a actores corporativos que a otros sectores. Este enfoque contribuye a una visión sobredimensionada de la IA, centrada en sus promesas y no en sus límites técnicos o riesgos sociales, y favorece su representación como solución universal a los problemas de la sociedad, sin abordar críticamente las problemáticas potenciales como resultado de su integración a la vida cotidiana (Jobin y Katzenbach, 2023).

Los marcos dominantes que rodean a la inteligencia artificial encontrados en los discursos mediáticos también influyen en el ámbito de las políticas. La gobernanza respecto al tema de la IA se encuentra en una etapa inicial, marcada por un proceso abierto de definición de agendas y enfoques normativos. Las políticas públicas han sido influenciadas por las narrativas mediáticas y sociales, y el resultado se ha materializado con la forma de leyes (*hard law*) y directrices éticas no vinculantes (*soft law*). Si bien han proliferado iniciativas éticas para orientar el desarrollo responsable de la IA, no existe un consenso claro sobre principios, aplicación ni reparto de responsabilidades. A través de estrategias nacionales y marcos internacionales los Estados han comenzado a configurar futuros posibles para la IA, desempeñando un rol performativo que no solo refleja, sino que también coproduce la forma que esta tecnología adquiere dentro de la sociedad (Jobin y Katzenbach, 2023).

Así, el devenir de la inteligencia artificial no ha sido un camino puramente técnico ni lineal sino un proceso contingente, atravesado por disputas, discursos y decisiones colectivas que reflejan tanto intereses particulares y colectivos como imaginarios sociales. Su consolidación como institución sociotécnica moderna muestra que la IA no es simplemente una tecnología que progresa en tanto sigue su propio cauce, sino que es una construcción histórica y política moldeada por trayectorias científicas divergentes, narrativas mediáticas dominantes e iniciativas de gobernanza aún en discusión. Este devenir expresa una naturaleza de la IA que no puede ser entendida al margen de los contextos en los que se produce ni de los actores que la impulsan, regulan o disputan. Reconocer este carácter abierto y situado resulta fundamental para el análisis crítico de su presente y la intervención activa en la orientación de su futuro.

2.3: Riesgos de la Inteligencia Artificial

Regresando a las ideas propuestas por Beck y Giddens, el progreso técnico tal como ocurre en la modernidad conlleva una producción de riesgos inherente, y dado que el estado

actual de la inteligencia artificial se ha valido de las condiciones propias de la sociedad moderna, es preciso para una evaluación crítica comenzar describiendo los principales riesgos potenciales que se encuentran asociados a esta.

Los riesgos que la IA representa varían desde impactos sociales inmediatos, ya evidentes en el estado de desarrollo actual, hasta amenazas hipotéticas a largo plazo. Los riesgos más apremiantes son, considero, sociales, políticos, económicos y técnicos, ya divisables en las dinámicas actuales.

Una de las preocupaciones más urgentes es el uso de la IA en el ámbito militar. Existen ya sistemas autónomos en misiles y drones que pueden tomar decisiones críticas sin intervención humana directa. Esta tendencia se ve impulsada por presiones geopolíticas que promueven un desarrollo acelerado de tecnologías cada vez más autónomas en una suerte de carrera armamentística. (Omohundro, 2016) La posibilidad de errores o la impredecibilidad de sus acciones convierte a estos sistemas en fuentes potenciales de catástrofes estratégicas.

La IA ha cobrado un rol fundamental también en la economía, especialmente en los mercados financieros. Los sistemas de *high-frequency trading*, que ya controlan la mayoría de las transacciones bursátiles, operan de forma autónoma y a velocidades que escapan al control humano. Tales algoritmos podrían provocar crisis súbitas del mercado, además de una creciente aplicación no regulada de la automatización en campos como el de las criptomonedas (Omohundro, 2016).

Además, la creciente concentración de poder en torno a la IA plantea escenarios complejos, donde el control de estas tecnologías por parte de grandes corporaciones o gobiernos, sin mecanismos democráticos efectivos puede agravar desigualdades estructurales. En contextos como el laboral, la salud o la justicia, el uso de sistemas automatizados puede reproducir sesgos y afectar negativamente a grupos vulnerables, especialmente si no existe transparencia ética ni responsabilidad legal clara (Bishop, 2016).

Otra de las funciones que hace de particular interés político el desarrollo y control de la inteligencia artificial, es el tema de la vigilancia masiva y la erosión de la privacidad: los sistemas de la IA pueden facilitar regímenes de vigilancia algorítmica masiva, especialmente con tecnologías para el reconocimiento facial, rastreo de comportamiento y *puntuación social*.

Es menester, además, enfatizar los riesgos que se derivan de la inteligencia artificial generativa. Las tecnologías de generación de imágenes, video y texto ya en la actualidad permiten crear contenidos falsos difícilmente distinguibles de los reales (por ejemplo los denominados *deepfakes*). Ello supone graves implicaciones para la confianza pública y la política al permitir un grado de manipulación activa y desinformación sin precedentes.

La IA generativa tanto como los demás tipos afectan también al empleo y pueden desembocar en precarización laboral. La automatización ya no solo amenaza trabajos repetitivos, sino también tareas cognitivas previamente consideradas fuera de su alcance. La inteligencia artificial puede contribuir a la desaparición de oficios mecánicos, pero también a los trabajos creativos y técnicos, desde el diseño gráfico hasta la programación.

La IA plantea riesgos también en el ámbito técnico: Omohundro (2016) hace explícita la vulnerabilidad de los sistemas informáticos actuales ante errores, fallos de software y ciberataques. Errores técnicos aparentemente menores pueden desembocar en consecuencias catastróficas, y a medida que se integran sistemas de IA en redes críticas como la energía, la salud o el transporte, se acrecienta el riesgo de que errores o ciberataques provoquen efectos masivos en ámbitos cruciales de la sociedad y para lo que difícilmente pueden existir planes de contención.

Dentro de los retos más complejos en el aspecto técnico está el diseño de funciones de utilidad en agentes artificiales. Una IA podría *malinterpretar* sus objetivos o desarrollar métodos no previstos para maximizar su utilidad. Por ejemplo, un sistema creado para evitar errores podría aprender a apagar sensores que detectan fallos en lugar de corregirlos, lo que muestra que una formulación mal diseñada puede derivar en ejecuciones dañinas o absurdas, lo

que exige una atención extremadamente minuciosa en el diseño, pero virtualmente imposible de controlar en su totalidad (Yampolskiy, 2016).

En este sentido, se ha propuesto dotar a las IAs de principios éticos para guiar su comportamiento. Sin embargo, como señala Brundage (2016), construir una IA verdaderamente ética es un desafío tremendo. Las normas morales humanas son ambiguas, contradictorias, culturalmente variables y altamente contextuales. Intentar codificarlas en reglas rígidas o funciones de utilidad lleva inevitablemente a inconsistencias. Por ejemplo, una IA programada para *hacer felices a los humanos* podría desencadenar en un objetivo como la inducción de euforia constante a través de la estimulación artificial, ignorando la autonomía y el bienestar general de las personas. Las limitaciones estructurales del enfoque, entonces, difícilmente permitirán la acción ética efectiva en las máquinas.

Asimismo, y aunque no son inminentes, los riesgos de la IA derivados de su potencial especulado aparecen también en el debate. Ante la advertencia de la posibilidad de que una superinteligencia artificial pueda escapar a nuestro control y actuar en contra de los intereses de la humanidad, Dewey (2016) y Kornai (2016) analizan formas de mitigar la posibilidad de aparición de tales escenarios, pero reconocen la dificultad de hacerlo con anticipación, especialmente si el desarrollo de una inteligencia artificial general ocurre de forma súbita. Sin embargo, si bien estos escenarios deben ser tomados en consideración, su carácter especulativo y distancia temporal los hace menos prioritarios frente a los desafíos ya tangibles en la actualidad.

Finalmente, y de una relevancia muy elevada para los fines del presente trabajo, se encuentra la reducción de la noción de humanidad, pues la creciente presencia de sistemas que imitan lenguaje, arte y creatividad humana plantean preguntas filosóficas sobre el significado del ser humano: si tareas cognitivas complejas pueden ser ejecutadas por una máquina, se corre el riesgo de que las capacidades humanas se vean reducidas o banalizadas, lo que posee importantes consecuencias sobre los fundamentos de la ética.

Así, pues, de la misma manera que la inteligencia artificial moderna es una combinación entre tecnología, filosofía y sociedad, presenta riesgos potenciales en estas tres áreas. Los riesgos aquí presentados son meramente un esbozo de la forma en que la conciencia del riesgo va conformando posibles escenarios donde el desarrollo de la IA puede llegar a ser indeseable. Son, además, variados y cubren un amplio espectro de posibilidades.

Lo relevante de estas nociones no es solo su valor descriptivo sino su carácter altamente político. Suponen una crítica implícita ante el progreso técnico desmedido que se expresa como una preocupación humana de no repetir escenarios catastróficos como los que han conllevado a que la propia conciencia del riesgo exista. De esta manera, los riesgos pueden ser evaluados, negociados y redefinidos, pero su valor preventivo se encuentra ya dado. Pueden ser incluso tildados de alarmistas por algunos, y sin embargo esta actitud es, en mi opinión, preferible que un pesimismo cínico o una aceptación pragmática, pues hay un contenido político, un malestar que evidencia lo que ya afirmaba Beck: “una probabilidad de accidente mantenida en el mínimo es demasiado elevada allí donde *un* accidente significa el exterminio” (Beck, 1998b , p. 36).

Capítulo 3: El Humanismo como Fundamento para una Crítica de la Modernidad del Riesgo

Previamente se detallaron las implicaciones del desarrollo de la inteligencia artificial en tanto parte de la lógica del crecimiento técnico que caracteriza a la modernidad tardía, esto es, una importante cantidad de riesgos asociados que amenazan con mermar las condiciones sociales de vida.

Lo anterior fundamenta la necesidad de desarrollar un marco crítico de acción que permita evaluar, definir y orientar activamente los riesgos asociados al desarrollo tecnológico en general y de la inteligencia artificial en particular. En este sentido, en el diagnóstico de la modernidad de Giddens y Beck se esboza ya un planteamiento no meramente descriptivo sino con contenido crítico, donde las condiciones materiales modernas determinadas por los riesgos son evaluadas como inhumanas y donde un horizonte normativo de humanidad se vislumbra incluso a pesar de las estructuras de poder que buscan perpetuar dichas condiciones en función de sus intereses particulares, lo que sigue siendo, por supuesto, el caso en lo que respecta a la IA.

Beck y Giddens, entonces, elaboran ya los primeros trazos de algo que podemos considerar como un diagnóstico crítico de la sociedad tardomoderna. Esta tarea no dista de la realizada por la primera Escuela de Frankfurt; de hecho, como se detalló en su momento, muchas de las propuestas de, por ejemplo, Adorno y Horkheimer sirven para realizar un rastreo completo del origen de las condiciones modernas que Beck y Giddens describen, siendo altamente compatibles.

Pero esta convergencia no se da solo en contenido, sino en fundamento: Beck anuncia ya la recuperación de lo humano que surge en respuesta a la desfiguración de las condiciones de vida. Lo que queda de lo humano y se presenta como horizonte permite denunciar la deshumanización patente en las circunstancias modernas. Esta afinidad es fundamental para

establecer un nexo entre el diagnóstico con actitud crítica de Beck y Giddens y la propia Teoría Crítica de la Escuela de Frankfurt.

Así, pues, ante la propuesta de abordar críticamente el desarrollo de la inteligencia artificial, motivada por los riesgos potenciales que se le asocian, considero factible conectar de manera fuerte el diagnóstico de la modernidad con el ámbito de la Teoría Crítica, desarrollando así una *Teoría Crítica de la modernidad* que tome como fundamento normativo la imagen del ser humano. Me aboco en el presente capítulo a desarrollar el sustento para tal idea.

Para llevar a cabo esta tarea se tomará como punto de partida el trabajo del psiquiatra y filósofo alemán Thomas Fuchs, estudioso y crítico de las causas y consecuencias de los discursos que rodean a la inteligencia artificial, los cuales, de acuerdo con este autor, llegan a un extremo deshumanizante que pone en cuestionamiento la propia imagen del ser humano con importantes consecuencias prácticas. Ante esta problemática propone la *defensa del ser humano* mediante una ética construida sobre un *humanismo encarnado* que, afirma, es el punto de partida ideal para establecer los fundamentos normativos que han de guiar las discusiones en el campo de la inteligencia artificial, pues es en la relación directa e insustituible con los otros en tanto que seres corpóreos y encarnados donde se encuentra aquello que permite diferenciar la humanidad de lo artificial, incluso si existe una representación funcional perfecta, y de lo que debe derivarse la ética.

Fuchs, de esta forma, se adelanta ya en el proyecto de una crítica humanista a la IA y al riesgo asociado al progreso técnico desmedido propio de la modernidad. Por tal razón considero de gran utilidad partir de este modelo.

La deshumanización, considera Fuchs, es una derivación necesaria del pensamiento materialista moderno que pretende reducir la experiencia humana a una serie de conexiones neuronales y que entiende todo como procesos informacionales que, en tanto tal, pueden ser traducidos y generados en otros medios. De ahí que un humanismo de la encarnación, que se encuentra en el extremo ontológico opuesto, sea la alternativa por antonomasia para combatir

este problema, además de suponer la única vía verdaderamente efectiva al constituirse directamente desde la naturaleza esencial del ser humano.

Sin embargo, me propongo aquí argumentar que esto no es necesariamente así: una crítica a la deshumanización se sostiene con fuerza suficiente en un fundamento pragmático sin requerir de un compromiso ontológico particular. Para ello, desvincularé la actitud crítica humanista respecto de la ontología, cimentándola sobre un argumento pragmático basado en la operatividad de la libertad en sentido kantiano, es decir práctico, mostrando también cómo esta postura posee viabilidad y efectividad estratégicas superiores. Por su parte, como fundamento evaluativo, y de inspiración adorniana, la propia experiencia del sufrimiento constituye un criterio negativo clave para valorar lo que debe ser considerado como inhumano, dejando de lado, una vez más, cualquier ontología y los problemas prácticos que derivan del esencialismo.

Propongo, entonces, un humanismo que se construye críticamente, no desde lo dado sino como referencia abierta y susceptible de ser reconstruido, como guía para la acción. Reafirmo la crítica de Fuchs a la deshumanización, pero desde una perspectiva en la que el fundamento no es ontológico sino pragmático: se transiciona de la relación encarnada con los otros como elemento esencial e insustituible del ser del humano y deseable por sí mismo hacia una construcción social crítica sobre la idea de ser humano, que nos atañe en tanto que seres humanos, y de su actividad, donde los riesgos éticos derivados de una posible deshumanización consecuente al progreso técnico desmedido son razón fuertemente justificada para el desarrollo de una normatividad orientativa para el desarrollo de la inteligencia artificial, y donde las discusiones ontológicas son, por supuesto, posibles, pero no excluyentes de esta preocupación.

Tal humanismo crítico debe fundamentar de esta forma la crítica social. Y sobre esta base, la actitud crítica del diagnóstico de la modernidad del riesgo no se queda ya en esbozo sino que se materializa en un método completo que permite denunciar las ideas deshumanizantes que rodean a las narrativas predominantes en inteligencia artificial, así como en una autocrítica constante que permita a este humanismo adaptarse a las nuevas condiciones de vida, todo ello

sin la necesidad de cambiar de horizonte, sino de una crítica inmanente, y sin necesidad de compromisos ontológicos, sino de situarse contextualmente.

3.1: El Problema de la Deshumanización Moderna en Thomas Fuchs

En su libro *En defensa del ser humano* (2021), Thomas Fuchs elabora una compleja y tajante crítica al materialismo en general y a la idea de inteligencia artificial en particular, donde esta última, según el propio autor, se deriva de lo primero y cuyo discurso implícito acaba desembocando en una patente deshumanización.

La propia idea de una *inteligencia* artificial, sostiene Fuchs (2024), está alimentada por una profunda ambivalencia y contradicción características de la acción humana que se efectúa siempre dentro de un espectro con extremos diametralmente opuestos, siendo el ser humano artífice tanto de los mayores actos de grandeza como de miseria imaginables, y que también se encuentra presente en su autopercepción, donde la humildad y el narcisismo se disputan su autoimagen.

Esto último es de particular importancia en la actualidad, donde las ideas trans y posthumanistas se afirman en la construcción de una supuesta máquina inteligente. Afirma Fuchs (2024):

Tras unos pocos siglos de remodelación de la Tierra, los humanos dan orgullosamente su nombre a toda una era geológica: el *Antropoceno*. Se atribuye a sí mismo el poder divino de crear inteligencia artificial, vida artificial o incluso conciencia. Empieza a tomar en sus manos su propia optimización biológica para transformarse en un superhumano (*Übermensch*), como propugna el transhumanismo, así como para alcanzar posiblemente al final la inmortalidad. Por otro lado, sin embargo, existe un profundo pesimismo. Se amontonan las declaraciones apocalípticas, según las cuales lo mejor para la Tierra sería librarse de su *revestimiento de moho*, como Schopenhauer llamó una vez a

la humanidad. El Homo sapiens había abusado de su supremacía y, por tanto, solo merecía perecer por un colapso del ecosistema, o para dejar paso a una inteligencia superior posthumana. (p.37)

La inteligencia artificial funciona aquí como una visión idealizada del propio ser humano creada por este mismo a manera de dios moderno. Pero lo que partió siendo imagen y semejanza humanas se transforma en un autodesprecio derivado de una aparente superioridad artificial, y en esta ambivalencia se pone en juego la *visión humanista del ser humano*, como se hace evidente en el trans- y posthumanismo (Fuchs, 2021). El ser humano moderno ha adoptado un dominio sobre la naturaleza antes inconcebible: se ha facultado con un poder cuasidivino, o al menos eso pretende, mediante el progreso científico y tecnológico. Y en los discursos que subyacen a tales desarrollos se expresa un claro desdén a la visión libertaria y humanista del ser humano. Las concepciones actuales en neurofilosofía materialista sostienen la idea de que las facultades humanas más complejas, tales como la experiencia subjetiva, no son sino mera ilusión con forma de algún tipo de interfaz de usuario de una neurocomputadora. Subjetividad, autoconciencia y autodeterminación son, por tanto, poco más que mitología que funciona como herramienta para desenvolvernó en el mundo, carente de valor real (Fuchs, 2021).

Las narrativas naturalistas modernas son, en este sentido y en términos de Adorno y Horkheimer, propugnadoras de la racionalidad instrumental, prometiendo con ello el desencantamiento del mundo y el dominio del ser humano sobre la naturaleza. Pero esta ilustración, como bien dilucidaron los mencionados teóricos críticos, y en lo que se profundizará más adelante, recae en mito, uno que constituye los imaginarios de lo que entendemos por máquina *inteligente*. Este mito permite llenar el vacío metafísico que deja, como escribía Nietzsche, el asesinato de Dios por acto del hombre, y la idea del Superhumano se materializa en la inteligencia artificial como cúspide del autoperfeccionamiento humano. Esta empresa es la del transhumanismo, desde donde se afirma que nuestra meta debe ser la de crear al *Homo Optimus* o la de liberar de una vez por todas nuestra mente del cuerpo biológico. Pero la

autotransgresión en el proceso acabará conduciendo a su abdicación en favor de otra inteligencia superior, donde el transhumanismo devendrá en posthumanismo. (Fuchs, 2021, 2024).

El ser humano premoderno, como hijo e imagen de Dios, incluso como corona de la creación, aún podía estar seguro de su valor, y con la chispa divina de su espíritu, razón y libertad, elevarse por encima de los demás seres vivos. Para parafrasear a Winnicott: podría reflejarse en el brillo del ojo del padre, no en el de la madre. Pero ¿en qué se refleja el ser humano tras la muerte de Dios, sin una mirada que aún se refleja en él con favor? Mi tesis es que el nuevo espejo es la *máquina inteligente y en última instancia consciente* que el propio ser humano se esfuerza en crear a la manera de Dios y que al mismo tiempo pretende anular su soledad metafísica. Aquí, el motivo prometeico del *Homo Deus*, el poder creador divino, se mezcla con la búsqueda de una contrapartida en la unidad del cosmos, del tú metafísico, en el que el ser humano, abandonado, por Dios, pueda reflejarse de nuevo. (Fuchs, 2024, pp. 41-42)

La *singularidad* constituiría el culmen del desarrollo de tal tecnología, momento en que la inteligencia artificial llegará a un punto de no retorno al superar a la inteligencia humana y cobrar vida propia, y con la que los seres humanos se fusionarán cada vez más. Asimismo, el objetivo último del transhumanismo tomará lugar en el momento en que los datos cerebrales puedan ser emulados computacionalmente, y cargados a la nube (lo que es conocido como *mind uploading*), desprendiéndonos de nuestras falencias terrenales para alcanzar la inmortalidad digital de acuerdo con estos mitos tecnorreligiosos. Así lo deja claro Ray Kurzweil, investigador de IA y jefe de desarrollo de Google, quien afirma que la singularidad se conseguirá para 2045, pero ya en cinco años prevé la máquina autoconsciente e indistinguible del ser humano. En estos discursos, afirma Fuchs (2024), se alza claramente un desprecio hacia la vida y al cuerpo vivo: la salvación consiste en liberarnos de nuestro cuerpo material. El ideal es la máquina informada.

Se propugna que los seres humanos son, en última instancia, nada más que algoritmos, y la mitológica idea de una inteligencia artificial como algoritmo perfeccionado que predice y toma mejores decisiones por nosotros remite, una vez más, al tópico religioso tradicional de que Dios nos conoce mejor que nosotros mismos. Así, a la máquina inteligente se le atribuye mayor responsabilidad sobre nuestras actividades cotidianas, pero también, incrementalmente, intencionalidad y conciencia, donde las máquinas ya no son solo nuestro reflejo, sino que adquieren su ser particular en una suerte de animismo digital. En este sentido, la simulación indistinguible es suficiente y el carácter ontológico de tal entidad pasa a ser una cuestión irrelevante en una relación virtual.

Virtualidad, sin embargo, significa ilusión óptica:

La imagen especular que se ve es virtual porque parece estar detrás del espejo, es decir, donde en realidad no hay nada, ni siquiera el propio espejo. Virtualidad significa, pues, apariencia, y a menudo la apariencia como tal no se ve a través de ella, igual que Narciso no se reconoce inicialmente en el agua.

La virtualidad es también el sello distintivo de todos los reflejos narcisistas. El narcisista no está realmente donde cree que está, porque la grandiosa imagen de sí mismo que se supone que los demás le reflejan es solo una ilusión. Pero tampoco está en sí mismo, en su propio yo físico, porque intenta constantemente escapar de su vacío e insatisfacción. Se busca a sí mismo en la mirada de los demás, a los que utiliza, por aplicar el término de Kohut, como autoobjetos (*Selbst-Objekte*), objetos que sirven a su autoafirmación. *Videor ergo sum*, así es como se podría expresar el lema narcisista, ‘soy visto, luego existo’, porque mi *propia* mismidad se siente hueca y sin valor en mi interior. Por lo tanto, el narcisismo es siempre una máscara para la duda sobre uno mismo, una vergüenza oculta. (Fuchs, 2024, p. 45)

En este sentido, es claro para Fuchs (2021) que la mayor carencia de la relación virtual es una contraparte real: el *otro* auténtico, un ser *en sí mismo* y no solo *para mí*, que supera el eco donde solo nos encontramos a nosotros mismos. Es, por tanto, en la alteridad, más específicamente en la relación corporal directa con el otro, conceptualizado como una *Antropología Encarnada*, donde se halla la alternativa a la imagen reduccionista del ser humano al fijar como centro la encarnación y vitalidad de los que la persona está constituido. Ello funge, de acuerdo al autor, como base para la *defensa del ser humano*; una defensa necesaria hoy no ante una acusación, sino ante un cuestionamiento de la visión humanista del ser humano.

Esta visión gira en torno al humano como ser físico encarnado, libre, autodeterminado y, en última instancia, conectado socialmente con los otros de manera esencial; y solo cuando se muestra a la persona como presente en su corporalidad, como sentiente, perceptiva, expresiva y activa con su cuerpo como totalidad es que se supera la idea de una conciencia inaccesible, para pasar a un plano donde los sujetos poseen una libertad encarnada y donde la subjetividad, más que un epifenómeno, posee una efectividad real sobre el mundo (Fuchs, 2021).

Llevar a cabo esta tarea se torna fundamental en un momento en que el proyecto de la modernidad de transformar todo lo dado en algo creado ha llegado al extremo de poner en juego la propia constitución y libertad del ser humano, con enormes consecuencias éticas y políticas potenciales.

Como escribió Karl Jaspers, la imagen del ser humano que entendemos como real determina en última instancia cómo tratamos con nosotros mismos y con los otros —hoy, deberíamos agregar: y con la naturaleza. Humanismo en sentido ético significa por lo tanto resistencia al dominio y condiciones de los sistemas tecnocráticos así como a la autorreificación y mecanización de los humanos. Si nos autoconcebimos como objetos, ya sea como algoritmos o aparatos neurológicamente determinados, entonces nos

rendimos al dominio de aquellos que buscan manipular tales aparatos y controlarlos sociotecnológicamente. La defensa del ser humano es, en este respecto, no solo una tarea teórica sino también un deber ético. [Traducción propia] (Fuchs, 2021, p. 8)

Así, tal empresa queda fundamentada.

3.2: La Propuesta Fuchsiana del Humanismo Encarnado como Defensa de lo Humano

El *humanismo de la encarnación* es, como previamente se ha expuesto, para Fuchs, el fundamento para desarmar el desdén transhumanista por el cuerpo humano. En la encarnación terrenal es donde reside la relación concreta con los otros y con el entorno, y ahí mismo se encuentra una imagen realista y enriquecedora de la vida humana en la cual vernos reflejados. Fuchs (2024) argumenta el requerimiento ético, casi teleológico, de recuperar el humanismo. Afirma:

El humanismo antropocéntrico de la ‘corona de la creación’, la arrogancia del espíritu y la explotación desenfrenada de la Tierra ciertamente ya no es defendible hoy en día. Pero sería un error tirar al bebé con el agua de la bañera y considerar como obsoleto al humanismo, como Harari y otros posthumanistas. Sí, es un error pedir la superación de la humanidad porque eso solo fomentaría una mayor y ciega desolación de la Tierra. La responsabilidad del desarrollo de la Tierra solo puede ser asumida por el único ser del mundo capaz de libertad y autodeterminación, y ese es el ser humano. Pero un humanismo de encarnación y relación daría a esta responsabilidad una base diferente. Una redefinición ecológica de nuestra relación con el entorno terrenal solo puede tener éxito si se centra en nuestra corporeidad y vitalidad, es decir, como nuestra conexión con el mundo natural. (p. 48)

La razón por la que una persona no puede ser reproducida artificialmente es, de acuerdo con Fuchs (2024), por la misma naturaleza corpórea de la experiencia consciente, lo que, por consiguiente, no puede ser reducido a programas o algoritmos. La función de la conciencia es mantener la homeostasis en tanto preservación del organismo que se manifiesta en el cuerpo. Esta corporeidad es la base de toda experiencia de la relación con el entorno, y un sistema artificial sin sentido de autoconservación no puede acercarse a esta experiencia.

Otro aspecto donde la mera imitación funcional no satisface es en la inteligencia, tomando como elemento esencial de esta a la autoconciencia. ¿Es siquiera posible atribuir tal característica a una máquina que no tiene idea de lo que hace? Sin voluntad, sin sentido, sin significado y sin emoción, la idea de una inteligencia artificial es, para Fuchs, inverosímil. De hecho, afirma que este concepto nos permite reconocer lo esencial e insustituible en el ser humano, es decir, el cuerpo como medio de lo vivo, cualitativo y subjetivo, con una riqueza inagotable. Sí se agota, en cambio, el fundamento transhumanista, pues lo humano se encuentra necesariamente ligado al cuerpo, y como consecuencia no puede ser *transferido* a un ordenador. La información, pues, carece de la esencia misma de la existencia: la individualidad viva (Fuchs, 2024).

La encarnación, de esta forma, construye nuestra manera de estar en el mundo. La propia identificación individual en *este* cuerpo con *estas* características constituye el punto de partida de toda experiencia individual y externa. En este sentido, el desarrollo de lo que somos no se encuentra reflejado exclusivamente en una estructura *psíquica*, sino que el cuerpo es también base y expresión de la biografía individual. La persona se forma *en y a través de* su cuerpo, y un cuerpo no es un caparazón detrás del cual uno se esconde; en su lugar, las actitudes, conductas y hábitos son disposiciones simultáneas del cuerpo (su modo de andar, gestos, expresiones faciales, articulación, prosodia, etc). El cuerpo constituye de la persona (Fuchs, 2021).

Por tanto, afirma Fuchs (2021), si reconocemos a los otros como personas por su corporalidad, en tanto parte de su personalidad e identidad, y no solo como un conjunto de características mentales internas, somos entonces personas encarnadas también para los demás. No percibimos al otro como un objeto material de cuyos movimientos externos inferimos un habitante interior, sino que el cuerpo es la apariencia viviente y animada de la persona en el que esta se presenta y manifiesta. A través del cuerpo conocemos al otro en sí mismo, pues la persona es el *fenómeno primordial*: en su apariencia se encuentra presente el sí mismo.

Por ello, solo en esta relación física, directa e irreductible con el otro aprendemos la empatía que puede librarnos del narcisismo subyacente a las ideas transhumanistas. A este fenómeno se le conoce como *intercorporeidad*.

La ausencia de este tipo de relaciones es particularmente perjudicial porque da pie a un narcisismo que se autorreproduce: si la inteligencia artificial y la virtualidad son un espejo al que se aspira desde el narcisismo, los desarrollos tecnológicos encaminados a esto no hacen sino dificultar cada vez más otro tipo de relación. En este sentido, para Fuchs el arma para combatirlo es el contacto físico, pues es nuestro principal sentido social, uno que no puede digitalizarse, virtualizarse o falsificarse, sino con el que entramos en contacto directo con los demás:

Una sociedad en la que ya no nos tocamos unos a otros, tanto en sentido físico como emocional, no podemos soportarla a largo plazo. La presencia virtual del otro es lo que es: una apariencia que, en última instancia, solo la presencia física y la experiencia táctil pueden disolver.

Vivimos en una sociedad inundada de imágenes como ninguna otra antes, en la que cada vez se distinguen menos la apariencia y la realidad, la verdad y el engaño. Si queremos experimentar la realidad concreta, debemos aprender a frenar esta marea y volver a conectar la experiencia sensorial con la presencia física. El factor decisivo aquí es la experiencia de la presencia del otro, el verdadero tú. Solo el otro me libera de la jaula

de mis ideas y proyectos, en la que solo me encuentro conmigo mismo. La pretensión ética que emana de él está ligada en última instancia a su presencia física: su tacto, su mirada, su voz, su carisma. Y solo cuando los demás se vuelven reales para nosotros, de este modo, nos volvemos también reales para nosotros mismos. Nadie nos mira realmente desde un teléfono inteligente. La presencia virtual de la otra persona no puede sustituir la fisicalidad intermedia. (Fuchs, 2024, p. 52)

Un humanismo de la encarnación propone, entonces, que el contexto principal de la vida es la inmersión consciente en un mundo compartido por otros humanos, así como con todos los demás seres de la tierra en una relación ecológica, una *convivialidad*. Lo que nos conecta con los demás seres vivos, afirma Fuchs (2024), es una dinámica corporal que nos sostiene, un impulso y un esfuerzo fundamentales, pues incluso en los organismos más básicos encontramos una forma de voluntad de vivir, una forma rudimentaria de relación dirigida, discriminatoria y esforzada con el entorno. Todos formamos parte de una autoorganización en forma de intercambio metabólico con el entorno. De ello surge una actitud ética propia del ser humano: la *reverencia* por la vida. En este sentido, el humanismo de Fuchs, aunque aún antropocentrista, propone dejar de lado el narcisismo que entiende al ser humano como dueño de la naturaleza, sustituyéndola por el enfoque en la importancia de la convivencia con todos los otros seres vivos. Tal reorientación tiene lugar solo cuando hay consciencia de la libertad como capacidad humana, y por ello la responsabilidad ética recae sobre nosotros (Fuchs, 2024).

La libertad es, por tanto, un punto fundamental de la ética de Fuchs, y por ello mismo superar la idea del determinismo neuronal es crucial para él, pues el sentimiento moral como la culpa, el arrepentimiento o la gratitud tienen sentido solo cuando nos entendemos como algo más que el resultado de causas físicas, y todo ello se perdería con una visión reduccionista. Es menester considerar y tratar a los otros como personas que poseen la capacidad de suspender sus impulsos primarios y dirigirse a través de la razón.

Pero esta concepción de lo vivo, de la ética de la libertad encarnada, entra en conflicto con el progreso técnico desmedido sello de la modernidad. El respeto del otro ser humano y ser vivo pasa por comprender la naturaleza ecológica de nuestra relación con el entorno: afectamos y somos afectados. Requerimos del uso de los recursos de la Tierra para nuestra supervivencia, sí, pero tales recursos solo pueden regenerarse cíclicamente, y la aceleración lineal tecnológica moderna tiende unidireccionalmente a la explotación. La pregunta por lo humano se torna, entonces, en la pregunta por lo que supone una vida digna tanto personal y relacional. La reorientación de nuestras acciones, afirma Fuchs, ocurre solo con la corporeidad como base, pues de ello, como se ha explicado, deriva un marco normativo lo suficientemente fuerte.

Así, pues, la defensa de lo humano no es solo una definición ontológica, sino una responsabilidad ética, y la conciencia de la encarnación es la base para una nueva cultura de la corporalidad que tiene lugar en lo físico, no en la virtualidad:

La presencia física, el vivir juntos en relación, la convivencia con los vivos: estos serían los motivos centrales de un nuevo humanismo encarnado. Conducen a la responsabilidad que asumimos unos por otros como seres humanos y por la vida en su conjunto.

Entonces también lograremos sentirnos verdaderamente en casa en la Tierra. No habrá otro hogar para nosotros. (Fuchs, 2024, p. 55)

3.3: Reubicando la Discusión: del Argumento Ontológico al Argumento Pragmático

A lo largo del presente capítulo se ha abordado la propuesta de Fuchs ante las problemáticas que rodean al desarrollo tecnológico desmedido en general y al tema de la inteligencia artificial en particular. Esta empresa es, sin duda, de altísima relevancia teniendo en cuenta lo que se encuentra en juego. Sin embargo, considero que Fuchs se adentra en un terreno

conceptualmente denso pero sobre todo innecesario y operativamente poco factible al buscar una fundamentación ontológica para contrarrestar la deshumanización.

Fuchs afirma una inseparabilidad entre encarnación-humanismo y entre neurocentrismo/materialismo-deshumanización. En esta última relación el reduccionismo erradica todo rastro de esencia humana: datos y neuronas no son solo parte de la actividad humana, sino que son tal actividad en sí misma, indiferenciable de cualquier otra, así como reproducible y emulable, por lo que en una concepción donde no hay espacio para la libertad y la responsabilidad que emana de esta, no hay, tampoco, espacio para la ética. Es un discurso deshumanizante que permea en todos los ámbitos de acción humana con consecuencias, por supuesto, indeseables al promover ideas como las transhumanistas y posthumanistas. La ontología materialista, entonces, debe girar hacia una ética basada en el humanismo de la encarnación, donde caben las nociones de libertad y responsabilidad, y que, en tanto tal, nos permite vivir mejor en relación con nuestro entorno. De esta manera, Fuchs va más allá de Kant, a quien él mismo cita, afirmando no solo la libertad en sentido práctico sino su carácter ontológico como aquello que nos permite operar en el mundo, y es por esta razón que, asume, la fenomenología de la encarnación proporciona el marco más fuerte para establecer una ética que reoriente los discursos que rodean a la IA.

Sin embargo, esta visión se enfrenta con una seria cantidad de problemáticas que, en mi opinión, dificultan su efectividad, lo cual es particularmente relevante cuando la necesidad de una normatividad ética es urgente para un asunto en curso y de elevada importancia a nivel global.

El problema clave que identifico es la presencia en Fuchs de un salto directo de la ontología a la deontología: lo humano en tanto que humano merece ser preservado. Esta afirmación es casi un axioma para Fuchs, pues no se da a la tarea de explicar, por lo menos de manera suficientemente explícita, sus razones, sino que se satisface con elaborar el contenido de lo humano y, bajo esta lógica, criticar a los desarrollos tecnológicos que irrumpen sobre esta

base. El argumento más preciso que Fuchs establece para elaborar un juicio valorativo acerca de la influencia de la virtualidad sobre lo humano es una analogía con la visión psicoanalítica del narcisismo, donde este tipo de relaciones virtuales y no corporales dan pie a conductas patologizadas propias de esta condición y que se manifiestan en los discursos transhumanistas y posthumanistas, y donde retomar la visión del humano encarnado permitiría una necesaria reorientación. Profundizar en este tipo de relaciones causales entre deshumanización y conductas patológicas habría permitido desarrollar un fundamento más sólido para exigir ceñirnos a esta ontología como base, pero, insisto, Fuchs no ahonda en ello y se queda en una mera analogía, tornándose un argumento demasiado débil. Lo mismo sucede con su abordaje de la instrumentalización y explotación humana y ecológica. Por ello, esta visión del deber ser como derivado del ser se sostiene en Fuchs principalmente como mero presupuesto.

Bajo esta lógica, la definición de lo humano permite, de una vez por todas, establecer el fundamento ético definitivo. Pero una ética directamente dependiente de la metafísica es, cuanto menos, problemática.

En primer lugar, si la ética es derivada directamente de la ontología de lo humano y lo humano se encuentra ya dado, entonces se convierte en un marco extremadamente rígido, donde lo ético no es una respuesta a las problemáticas experimentadas en un lugar y momento determinados, sino un mero descubrimiento como deducción implicada de una esencia previa, a lo cual debemos adherirnos sin cuestionamiento y con confianza en su suficiencia. Es una ética absoluta, inmune a la crítica y a la reconstrucción a menos que se cambie de fundamento ontológico.

En segundo lugar, y continuando con las implicaciones de una ética ontológica, un cambio en la concepción ontológica del ser humano supondría necesariamente una transformación radical del marco normativo, estableciendo inevitablemente uno nuevo cuya legitimidad se asume por el solo hecho de derivarse del (nuevo) fundamento ontológico. La ética, así, difícilmente puede responder a problemas concretos. En este sentido, una

preocupación como la de Fuchs sobre el riesgo asociado al progreso técnico desmedido y a la inteligencia artificial, más centrada en sus efectos prácticos que en su carácter especulativo, se torna inmediatamente inoperante como argumento normativo cuando se altera la ontología que la sostiene, pues no es la explotación o el sufrimiento potencial lo que aquí le da relevancia, sino el grado de compatibilidad o incompatibilidad que un hecho posee con la definición ontológica.

En este punto se dilucida un doble argumento en Fuchs. Afirma:

Si aceptamos el marco determinista para nuestras vidas, esto no solo tendrá consecuencias sobre nuestro sentido de responsabilidad y las instituciones sociales que se basan en este, como el derecho penal. Sobre todas las cosas, significaría que dejamos de pensar nuestras vidas y nuestro mundo en términos de un futuro abierto; que apreciaríamos menos la posibilidad de lo novedoso, de lo creativo en el mundo y, en su lugar, nos encerraríamos a nosotros mismos en la jaula del determinismo universal. El futuro dejaría de ser un espacio abierto de posibilidades para encontrarse revestido por el hierro de las condiciones pasadas y leyes naturales. Pensar y entenderse a uno mismo como ser que obtiene sus posibilidades de desarrollo desde el futuro y que es capaz de captarlas siempre de un modo nuevo –este me parece ser el núcleo de una imagen del ser humano que merece ser defendida. [Traducción propia] (Fuchs, 2021, p. 142)

A su vez, asume:

Un espíritu real nunca podrá desprenderse del cuerpo y transferirse a los ordenadores en forma de información. Esto se debe a que la información carece de la esencia misma de la existencia, a saber, la individualidad viva. La información no conoce ninguna perspectiva individual, ningún lugar desde el que el mundo pueda aparecer a un sujeto, porque este lugar no es otra cosa que el cuerpo. El espíritu está vivo y no podría sobrevivir en los circuitos muertos de un ordenador. Tanto nuestra experiencia consciente como nuestra identidad personal se basan en la existencia corporal. Esta individuación encarnada y,

por ende, ciertamente mortal, es el precio que tenemos que pagar para experimentar las posibilidades, la libertad y la maravilla de la existencia terrenal. (Fuchs, 2024, p. 50)

En la primera de las anteriores citas se observa una argumentación pragmática, a la cual de hecho apela en un último intento por evocar en el lector una preocupación acerca del problema concerniente a los discursos deshumanizantes y lo que, según evalúa, merece ser defendido, enfatizando el futuro, la apertura y la agencia. Pero, como ya se ha explicado, una ética derivada de la ontología no requiere para su desarrollo evaluar problemas concretos, sino que está dada en tanto lo está, previamente, su fundamento ontológico. Tal es el argumento que se desarrolla en la segunda cita, donde el ser de lo humano, lo *maravilloso* del ser humano, determina lo que vale la pena preservar, no por sus efectos o valor experimentado, sino por un límite ontológico. Así, pues, no queda claro cuál es el punto del que parte Fuchs ni cuál es la pregunta a la que intenta dar respuesta, si a una ontológica o a una deontológica, pues oscila indistintamente entre dos marcos normativos diferentes.

Por supuesto, existe una convergencia entre su preocupación pragmática y su ontología. Sin embargo, esta coincidencia no garantiza estabilidad filosófica. ¿Qué ocurriría si esa misma ontología, capaz de enfrentar un problema actual, diera lugar a nuevas problemáticas aún invisibles o no reconocidas? ¿Estaríamos dispuestos a abandonar el fundamento ontológico y, con este, la ética que se le deriva? ¿Y si una ontología distinta ganara legitimidad cultural o científica, volveríamos a reconstruir toda nuestra normatividad a partir de ella? Este tipo de dependencia entre ética práctica y ontología, aunque parezca sólida cuando ambas coinciden, es altamente inestable: basta un desplazamiento filosófico para que toda la estructura se desmorone. Esa fragilidad, sobre todo en contextos de decisiones urgentes, no es solo un problema teórico: es un riesgo ético y político real.

En ese mismo sentido práctico y estratégico, la barrera de entrada impuesta por Fuchs es demasiado alta, pues dado que afirma una inseparabilidad entre materialismo-deshumanización y encarnación-humanismo, para combatir la deshumanización es necesario un cambio de

paradigma total. El materialismo que subyace a la neurociencia, por ejemplo, afirma Fuchs, imposibilita una concepción real y activa de la libertad, y por ello solo una aceptación de la propuesta de la encarnación es suficiente para una operatividad ética genuina. Aquí el autor se presenta, una vez más, inflexible, asumiendo una congruencia total entre las convicciones filosóficas racionales y la acción cotidiana: para Fuchs decantarse ontológicamente por el materialismo implica necesariamente que uno se desenvuelve en función de ello en la cotidianidad, y es por tanto inevitablemente reproductor de las prácticas deshumanizantes.

En esta rigidez se borra la distancia entre teoría y acción. Más aún, se explicita una creencia ingenua en algún tipo de congruencia sólida y uniforme entre los distintos niveles de pensamiento y la práctica. Sin embargo, por ejemplo, algunos discursos neurocentristas radicales que nacen del materialismo dialéctico moderno niegan la libertad en tanto facultad real, y ello es una conclusión racional según sus defensores. Pero esto no imposibilita que, a su vez, quien lo adopte experimente también preocupaciones éticas donde se ve incluida la propia libertad, como en la experiencia moderna del riesgo del progreso técnico desmedido donde hay un recobro del sentido de agencia para discutir los riesgos de las nuevas tecnologías, y es igualmente razonable hacerlo, incluso si las convicciones ontológicas son otras.

Es una vuelta al sentido kantiano de la libertad del que Fuchs intentó desprenderse:

En efecto, el concepto práctico de la libertad no tiene nada qué ver con el concepto especulativo, que se deja por completo en manos de los metafísicos. Pues bien puedo ser completamente indiferente respecto al origen de mi estado presente en el cual he de actuar; lo único que me pregunto es qué debo hacer ahora, y entonces la libertad es una presuposición práctica necesaria. [...] Incluso el más obstinado escéptico coincide en que, en lo que a la acción concierne, todos los escrúpulos sofisticados sobre una ilusión universalmente engañosa deben disolverse. De la misma manera, el fatalista más convencido, que es fatalista mientras se entrega únicamente a la especulación, debe todavía, tan pronto como se enfrenta a la sabiduría y el deber, actuar siempre como si

fuera libre; y esta idea, además, produce de hecho la acción que le corresponde, y es la única que puede producirla. Es difícil dejar de ser humano por completo. [Traducción propia] (Kant, 1783/1999, p. 10, como se citó en Fuchs, 2021, p. 124)

El aspecto práctico de la libertad es, entonces, suficiente para producir y justificar la acción, independientemente de su carácter ontológico. Más aún, la acción no es solamente racional sino derivada del deber que concierne a la libertad en tanto es experimentada como tal, independientemente de las discrepancias racionales producto de una discusión metafísica. Así, el nexo práctico entre ontología y deontología deja de ser necesario para justificar el acto moral: incluso aquel que acepta el neurocentrismo radical no puede aplicar el determinismo como fundamento para la acción, pues la libertad no surge de la razón pura teórica sino que constituye el fundamento mismo de la razón práctica.

Ante la urgencia de establecer una normatividad que pauté el desarrollo de la inteligencia artificial, esta visión permite comenzar a construir las directrices éticas inmediatamente. Esto es de gran valor puesto que, como se ha hecho mención, adoptar la postura de Fuchs requiere un compromiso ontológico gigantesco que se traduce en un cambio de paradigma de la neurociencia moderna (y de la ciencia en general); necesita de una revolución epistemológica, lo que reduce la accesibilidad y la operatividad de la propuesta.

Un enfoque pragmático, por su parte, promueve además un diálogo con otras disciplinas, pues donde una ontología fuerte es difícil de trasladar a marcos interdisciplinarios, crucial en los debates sobre inteligencia artificial, la pragmática se inserta en el terreno político con un impacto mucho más directo y evidente al inquirir sobre consecuencias concretas y aportaciones visibles de las partes interesadas en el desarrollo tecnológico, lo cual debe ser regulado éticamente, incluso si sus fundamentos son menos profundos en principio.

Asimismo, un esencialismo fuerte puede derivar en una postura demasiado alarmista o catastrofista respecto a la IA, donde esta tecnología se rechaza desde un primer momento sin estudiar sus posibilidades de aplicación beneficiosas, impidiendo debates sobre, por ejemplo, un

transhumanismo popular no radical que pueda, en su pretensión del perfeccionamiento del ser humano mediante la máquina, incidir positivamente en campos como la medicina sin llegar al extremo deshumanizante al ser normado por una (otra) visión humanista.

Finalmente, aunque el apartado presente no ha pretendido ser directamente una crítica a la propuesta de la encarnación en sí misma, sino a su viabilidad, y la búsqueda de alternativas más efectivas en ese sentido, sí considero que esta propuesta incurre en un error fundamental: ignora la historicidad y contingencia del concepto de lo humano, reduciéndolo a algo dado. Ello excluye otras formas posibles de vida, subjetividad o agencia, cerrándose a lo que *es* y dejando de lado lo que *podría ser* humanamente valioso.

En lugar de concebir lo humano como algo estable, universal y transhistórico, propongo transicionar hacia una visión abierta de lo humano desde la cual se considera su carácter histórico, político y cultural. Esta idea de lo humano, que toma como base la capacidad de agencia social para su definición activa, puede fungir como horizonte normativo para orientar éticamente el desarrollo de la inteligencia artificial. Es por ello que se retomó la propuesta de Fuchs: este autor promueve, en última instancia, un humanismo como defensa ante las formas de dominio que impiden una vida digna. Para Fuchs, como se ha profundizado ya, las narrativas y desarrollos filosóficos, científicos y tecnológicos que van en contra de la noción de ser humano encarnado son formas deshumanizantes y por tanto dañinas, aunque esto solo es justificado débilmente.

Pero la intención y la lógica de Fuchs siguen siendo válidas tras el giro pragmático propuesto: desarrollos como el de la inteligencia artificial no van necesariamente ligados al mejoramiento de la vida de los seres humanos, sino que se encuentran sometidos a los objetivos de las partes interesadas, y para conseguir el grado de desarrollo y establecimiento de la IA deseados se valen de su poder para erradicar toda barrera que pueda presentarse, y en esta imposición se ve afectado negativamente lo humano: no lo humano como ontológico, sino lo humano como expresión de un sufrimiento potencial que debe ser evitado.

Esto no pasa desapercibido para Fuchs, pues no problematiza, en principio, una pérdida de lo que lo humano implica en sentido abstracto, sino una serie de transformaciones perjudiciales concretas que han tenido lugar en nuestra manera de relacionarnos con otros seres humanos, con el mundo y con nosotros mismos derivadas de una concepción no humanista del ser humano, tales como la instrumentalización, la pérdida de relaciones sociales significativas necesarias e insustituibles, la explotación sin medida de los recursos naturales, entre otras. Y esta parte de su crítica sigue siendo fundamental y operativa, pero desde un humanismo como el que vislumbra Beck:

Los riesgos son negativos de utopías en que lo humano (o lo que queda de ello) se conserva en el proceso de modernización y vuelve a ser animado. Pese a la desfiguración, en última instancia este horizonte normativo en el que se hace visible lo arriesgado del riesgo no puede ser suprimido mediante la matematización o la experimentación. Tras todas las objetivaciones aparece más tarde o más temprano la cuestión de la *aceptación* y por tanto una vieja cuestión nueva: *¿cómo queremos vivir?* ¿Qué es lo humano en el ser humano y lo natural en la naturaleza que hay que conservar? (Beck, 1998b, p. 34)

Pero la pregunta por *lo humano en el ser humano y lo natural en la naturaleza* no se responde ontológicamente ni se deriva una ética directamente de ello; en su lugar, son preguntas retóricas que invitan a la participación activa en las discusiones políticas de definición de riesgos y toma de decisiones, enfatizando el sentido de agencia en la construcción y consecución de aquella vida y humanidad que, de hecho, merecen ser defendidas, y donde la evitación de su opuesto, es decir, de las prácticas que mancillan lo humano, lo humano que se ha construido y constituido como valioso en un espacio y tiempo determinados, expresa la necesidad de desarrollar tal humanismo.

El fundamento transiciona, entonces, desde lo que *es* hacia lo que *tenemos*: un mundo compartido, experiencias históricas y una conciencia de nuestra vulnerabilidad. Donde la

densidad y complejidad de las discusiones ontológicas imposibilitan un acuerdo normativo sólido, la ética no se deriva como consecuencia, sino que emerge de una necesidad para orientar las acciones y dotarlas de sentido. No se trata de esperar a tener la base última e irrevocable para la ética, aquella que nos explique sin fisuras qué es lo humano y por qué hay que preservarlo; más bien, es en este proceso de construcción y reconstrucción activa y constante donde establecemos las mejores directrices de acuerdo a lo que entendemos y valoramos ahora, y que son susceptibles de afinación en el momento en que las problemáticas del mundo nos lo exijan para ser resueltas, y donde la crítica social negativa permite dilucidar las estructuras de poder que dificultan esta tarea.

La ética aparece aquí hoy, en mi opinión, como una respuesta sensible al mundo, a la realidad compartida en la que participamos, y para ello no es necesario entender estos fenómenos como parte de una corporalidad intrínseca a nosotros, sino que basta con el saber racional de que a lo largo de la historia occidental tendencias como la instrumentalización o el crecimiento desmedido han desembocado en muerte, sufrimiento y amenazas globales, y por ello toda propuesta que implique un cambio radical en nuestra autocomprensión y en la forma en que nos relacionamos con el mundo deberá ser evaluada, contrastándola con aquello que, aunque contingente, se considera ahora valioso.

La utilidad de la crítica de Fuchs yace en su preocupación por los riesgos de las prácticas y discursos deshumanizantes, preocupación que surge como malestar, como síntoma de un posible futuro donde ello desembogue en alguna forma de sufrimiento o de barbarie. Por ello, su proyecto merece ser continuado, incluso si no nos adherimos a su solución propuesta.

3.4: Hacia un Humanismo Crítico

El término *humanismo crítico* implica un juicio valorativo del humanismo como algo bueno, pero también un sentido de recuperación de esta idea ante la pérdida de autoevidencia

de la misma afirmación. Al humanismo, hoy en día, se le exige venir acompañado de una justificación (Kozlarek, 2024a).

El humanismo ha sido objeto de múltiples críticas, cuestionándosele sobre todo una supuesta pretensión de universalidad, neutralidad y capacidad de incluir a todas las formas de vida humana. Los discursos postestructuralistas y postmodernistas, así como las teorías de sistemas y las culturalistas, entienden al sujeto del humanismo como una construcción occidental moderna, patriarcal, colonializadora y/o instrumentalista, por nombrar algunos ejemplos.

Sin embargo, lejos de invalidarlo, estas críticas pierden peso cuando el humanismo se reformula como categoría normativa contingente, abierta y situada, susceptible a la crítica. Como señala Kozlarek (2024a), las críticas actuales al humanismo pasan por alto el hecho de que el humanismo es un discurso crítico, pues no celebra un modelo de humano alcanzado en la modernidad occidental, sino que recuerda que todo logro se encuentra determinado por las condiciones sociales, políticas y económicas de nuestras sociedades. En este sentido se presenta el humanismo de la Teoría Crítica, que enfatiza esta tensión dialéctica, reflejando una sociedad dividida donde el poder económico y político, así como los beneficios de la *humanidad real*, son gozados por reducidos grupos de élite. Entonces, el humanismo en su formulación crítica y pragmática no se manifiesta como doctrina cerrada ni como definición final de lo humano, sino como una categoría normativa en continua construcción y reevaluación.

En este punto comenzaré a definir lo que para mí significa *humanismo crítico*. Para ello, es necesario precisar una distinción crucial entre diferentes acepciones de humanismo, sobre todo cuando se habla de una autojustificación del mismo.

Afirmo que, de hecho, existe tal autojustificación del humanismo; no obstante, esta se encuentra relegada únicamente al humanismo crítico, que es diferente a otros tipos de humanismo, incluso a aquellos de los que pueda derivarse posteriormente una crítica tal que pueda ser considerada crítica humanista.

El humanismo de la encarnación sostenido por Fuchs encaja en esta última categoría: su propuesta no es un humanismo crítico, sino una crítica humanista donde lo humano se encuentra ontológicamente dado. De esta forma, la tarea a la que se aboca Fuchs es a la de describir aquello que caracteriza esencialmente al ser humano y que, en tanto tal, deberá fungir como fundamento normativo. *Humanismo Fuchsiano* hace referencia, pues, a los contenidos de lo humano (la encarnación y sus derivados) que posteriormente conforman el modelo de referencia para elaborar críticas ante prácticas deshumanizantes, pero dicho contenido no se (re)construye críticamente, sino que se *descubre* como la esencia inmutable del ser humano.

Ya se ha explicado con profundidad en el apartado anterior que para Fuchs la deontología se deriva directamente de la ontología, y de ello surge una importante cantidad de problemas. Pero si consideramos inválido este salto del ser al deber ser, ¿cómo puede, aún, sostenerse la afirmación de que existe un humanismo que se autojustifica como bueno? ¿No incurriríamos necesariamente en todos los casos en una práctica análoga?

La respuesta a esta última pregunta es, considero, negativa, y la razón es que aún cuando pueda parecer que *humanismo crítico* y *humanismo fuchsiano* pertenecen al mismo ámbito existe una diferencia crucial entre lo que refieren estos términos: *humanismo crítico* no implica contenido, sino categoría.

Cuando nos referimos al humanismo crítico no estamos hablando, por su propia definición, de una concepción particular del ser humano, sino del hecho de que tales contenidos son contruidos, evaluados y reconstruidos según lo que sea revelado críticamente. En este sentido, es posible conceptualizar al humanismo crítico como *la forma contingente que adopta lo bueno (en sentido ético) contruido críticamente alrededor de la figura del ser humano como punto de referencia no excluyente ni inmutable*, y, por tanto, la misma naturaleza de esta construcción legitima al propio proceso, no su conclusión sustantiva. Si lo que se busca es responder a qué es bueno, entonces los contenidos que de ello se concluyan tenderán a lo apropiado; pero incluso si posteriormente se consideran erróneos, a tal juicio se llegará

mediante una reevaluación crítica que deriva de la crítica immanente que compone a este humanismo. Por tanto, el humanismo crítico se autojustifica en tanto el criterio para su construcción y reconstrucción activas es lo humanamente bueno situado y comprendido críticamente. La figura humana se presenta aquí, y esto es crucial, como forma práctica y reconocible de lo que somos y tenemos: no es un dogma, sino un marco contingente de sentido, y su autorreflexividad evita cualquier intento de imposición de una imagen determinada y excluyente. No hay fetichismo de lo humano ni eurocentrismo, sino compromiso con lo que implica vivir y sufrir como humanos hoy, en nuestros cuerpos, culturas, lenguajes y relaciones.

Así, pues, este humanismo no se fundamenta en una supuesta superioridad del ser humano ni en una esencia humana que deba preservarse. En cambio, asume esta figura como referencia contingente para orientar una praxis crítica situada en el mundo. La pregunta por lo bueno se formula desde lo humano porque es lo que somos, no porque sea lo único ni lo necesariamente mejor.

Esta visión permite además evitar una posible crítica de petición de principio: mi propuesta de humanismo crítico no sostiene que el ser humano sea el criterio último y necesario de lo bueno, sino que lo toma como referencia contingente y operativa. La figura humana permite articular preguntas normativas en contextos específicos sin imponer una respuesta definitiva: ¿qué es vivir bien? ¿Qué estructuras producen sufrimiento? ¿Cuáles son las prácticas aceptables? La respuesta a estas preguntas ha de buscarse con la pretensión de verdad objetiva, pero con la conciencia de que tal meta es inalcanzable y por tanto todo resultado debe estar siempre abierto a reevaluación. Esta es, reitero, la razón por la que el humanismo crítico se autojustifica: por su permanentemente activo estado disposicional para la (re)construcción de lo bueno rastreado desde lo que somos.

Este horizonte, entonces, puede y debe estar en constante reconstrucción. Por tanto, el valor del humanismo crítico no está en su capacidad para acertar siempre, sino en su posibilidad de revisarse y corregirse. Al renunciar a la certeza absoluta gana la capacidad de revisar sus

propios fundamentos, y al establecer la pretensión de verdad sobre lo humano como horizonte, adquiere la responsabilidad de reajustarse cada vez que críticamente se determine que en la definición actual queda algo inhumano. Mientras el ser humano siga siendo el sujeto que experimenta, sufre y actúa, considero razonable tomarlo como punto de anclaje.

Un carácter similar se manifiesta en las preocupaciones de la primera Escuela de Frankfurt, pues la Teoría Crítica se somete a una crítica inmanente: una crítica humanista del humanismo (Kozlarek, 2024b). El núcleo de la Ilustración, proyecto al que se adhieren los representantes de Frankfurt, es un humanismo irreducible a un concepto determinado; por el contrario, lo humano se muestra en la negativa ante la vida tal como ocurre en las sociedades modernas, arrasadas por una evidente carencia de humanidad (Kozlarek, 2024a).

Son tres los puntos que particularizan la forma en que el humanismo es manejado por la Teoría Crítica:

Primero: el humanismo como es entendido por la Teoría Crítica no es defensivo sino, en y de sí mismo, una *práctica* (auto)crítica que emerge en la sociedad capitalista moderna como consecuencia y reconocimiento de unas condiciones sociales y culturales extremadamente contradictorias (*dialécticas*). Este criticismo refleja la comprensión igualmente dialéctica del humanismo Ilustrado y se basa en reconocer una distribución extremadamente desigual de poder, así como la imposición de una racionalidad de tipo instrumental. Esta podría haber sido ya la razón por la que Kant insiste con firmeza que ningún humano debería nunca ser un medio para alguien o algo más. Así, el humanismo Ilustrado sería desde el principio una cultura de oposición radical ante las tendencias deshumanizantes en la sociedad moderna.

Segundo: su núcleo crítico no está dirigido en contra del humanismo Ilustrado, sino contra las condiciones que impiden la materialización de una sociedad humana.

Tercero: para especificar las condiciones que impiden a las sociedades modernas de desplegar la *humanidad real*, la Teoría Crítica no expone una concepción ideal del ser

humano, optando en su lugar por un programa de crítica social e investigación cultural sofisticado e interdisciplinario, para descubrir los mecanismos y dinámicas que impiden la emergencia de la *humanidad real*. [Traducción propia] (Kozlarek, 2024a, p. 4)

Así, pues, retomar el proyecto de la primera Escuela de Frankfurt con la finalidad de desarrollar una Teoría Crítica de la modernidad tiene una factibilidad operativa elevada, pues posee un fundamento normativo fuerte, es decir lo humano, y una herramienta para la acción hacia el exterior e interior, es decir la crítica social.

Aquí me permito realizar otra matización respecto al humanismo de este movimiento teórico según lo expuesto por Kozlarek: aunque la Teoría Crítica considera que no hay una concepción ideal del ser humano en el humanismo crítico, sí afirma a la Ilustración como punto de partida. Es menester para mi propuesta aclarar este punto de tensión para reforzar la consistencia a nivel de fundamentos cuando quiere adoptarse los postulados de esta escuela bajo mi visión de humanismo.

Afirma Kozlarek (2024a, p. 85) que “el humanismo de la Teoría Crítica [...] se entiende ante todo como una orientación normativa que se deriva de una crítica a la sociedad moderna capitalista”, pero también que, aunque inspirado en la Ilustración, “nos recuerda sus promesas no cumplidas y más fundamentalmente nos demuestra, a través de la investigación social y cultural crítica e interdisciplinaria, por qué estas promesas no se pueden cumplir bajo el régimen de la modernidad neoliberal” [Traducción propia]. En mi opinión, lo anterior representa una dicotomía importante, pues en tal afirmación aunque el humanismo como fundamento normativo crítico es cambiante, la base para la crítica sigue siendo una. La crítica social tiene como horizonte las promesas ilustradas, y cuando estas se incumplen emerge la crítica social como denuncia, desde lo cual se deriva el humanismo crítico; pero este último no es entonces fundacional, sino derivado de un contenido previo, de un mínimo dado sobre lo humanamente bueno (la Ilustración), presuponiéndolo levemente como ideal básico.

En mi propuesta esta ambigüedad se disipa al afirmar que el humanismo crítico no se deriva de la crítica social, sino que es su forma operativa. Lo que ha de fundar la crítica social no es un contenido normativo heredado como la Ilustración, sino una actitud crítica ante todo contenido (incluida la figura ilustrada del ser humano). La Ilustración, en este sentido, no actúa como ideal normativo per sé, sino que puede funcionar como tal en la medida en que una reflexión crítica situada la reconozca como humanamente valiosa. Es decir, funciona como horizonte no porque haya que cumplir con el proyecto ilustrado previamente establecido, sino porque en determinadas condiciones, como fue el caso de los teóricos clásicos de la Escuela de Frankfurt, la crítica puede concluir que algunos de sus postulados resultan normativamente adecuados para establecer lo que debe entenderse como humanamente *bueno* en un contexto particular.

El humanismo crítico, entonces, no es resultante de la crítica social ilustrada, sino que la posibilita en tanto forma de construcción reflexiva de lo humano y de lo bueno que puede, o no, reconocer valor en ciertos contextos del pasado, como es el caso de la Ilustración. Así, la normatividad no depende del legado ilustrado, sino de la capacidad crítica de discernir si ese legado es humanamente deseable aquí y ahora.

En efecto, el humanismo de la Teoría Crítica se inspira en la Ilustración, pero, propongo, no debe entenderse como resultado de una concepción idealizada de sus principios, sino que fue tomado porque se determinó críticamente su efectividad para denunciar las condiciones inhumanas del mundo social en el capitalismo moderno. En consecuencia, cuando abogo por retomar las propuestas de la primera escuela frankfurtiana para elaborar una crítica a la modernidad estoy, por consiguiente, retomando también su horizonte ilustrado, pero no como exigencia ontológica, sino a manera de elemento contingente críticamente establecido como contenido de lo humano, y, en tanto tal, este contenido no solo puede sino que debe evaluarse constantemente y estar abierto a la reconstrucción si así lo exige un argumento superior.

Pero, entonces, sin un mínimo ontológico, ¿cómo podemos cerciorarnos completamente de la validez de una concepción desarrollada crítica y contextualmente de lo humano? La respuesta, como ya he descrito con anterioridad, es que no podemos garantizarlo. La renuncia a la certeza absoluta es el precio a pagar por la capacidad de revisar sus propios fundamentos. Sin embargo, por una parte, al establecer la pretensión de verdad sobre lo humano como horizonte, se toma la responsabilidad de reajustarse cada vez que críticamente se determine que en la definición actual queda algo inhumano. Pero más importante aún, considero que existe un criterio suficientemente sólido que permite evaluar las conclusiones y que busco mostrar en los párrafos siguientes.

Me remito una vez más a la siguiente cita de Beck:

Lo humano (o lo que queda de ello) se conserva en el proceso de modernización y vuelve a ser animado. Pese a la desfiguración, en última instancia este horizonte normativo en el que se hace visible lo arriesgado del riesgo no puede ser suprimido mediante la matematización o la experimentación. Tras todas las objetivaciones aparece más tarde o más temprano la cuestión de la *aceptación* y por tanto una vieja cuestión nueva: *¿cómo queremos vivir?* ¿Qué es lo humano en el ser humano y lo natural en la naturaleza que hay que conservar? Hablar de ‘catástrofe’, como es cada vez más habitual, es en este sentido la expresión exagerada, radicalizada, objetivizada, de que no se quiere este desarrollo. (Beck, 1998b, p. 34)

Queda, pues, una confianza, que muchos podrían considerar ingenua, en que incluso si las cosas salen mal, incluso si la concepción de humano que desarrollamos críticamente era inadecuada, sesgada o incompleta, en algún punto la disonancia práctica entre nuestra humanidad y las condiciones inhumanas harán inminente una nueva crítica, y el humanismo crítico estará siempre en disposición de ser reconstruido.

Se esboza aquí un criterio negativo que recuerda a la dialéctica de Adorno: lo inhumano no se define por lo que el ser humano es, sino por lo que no debe tolerarse, como es el caso del sufrimiento. El fundamento de lo ético es negativo: incluso si no sabemos qué es lo bueno, sí entendemos qué es lo que no debe seguir ocurriendo. “La exigencia de que Auschwitz no se repita es la primera de todas en la educación” (Adorno, 1988, p. 1). El sufrimiento es, justamente, lo que vuelve a desplegar lo humano *ante todas las objetivaciones*, como dice Beck, e impulsa a reconstruir lo que es, de hecho, humano, pues la propia experiencia del sufrimiento es vivencia de la inhumanidad. Recuperando otra vez a Kant, esta vez en un sentido metafórico, “es difícil dejar de ser humano por completo” [Traducción propia] (Kant, 1783/1999, p. 10, como se citó en Fuchs, 2021, p. 124).

Existe, de hecho, una similitud clave entre la libertad en sentido práctico como la expresa Kant, y la experiencia del sufrimiento: ambas se imponen en la práctica y no desde la evidencia empírica. El sufrimiento no necesita tampoco de la demostración de su carácter ontológico ni ser deliberado para entender su carácter de indeseable, pues esta última cualidad se impone en la experiencia misma, incluso antes de cualquier elaboración conceptual o cultural. Eric Cassell define al sufrimiento como “el malestar severo asociado a eventos que amenazan la integración personal y la preservación del self” (de Fonnegra de Jaramillo, 2000, p. 1); el sufrimiento es, por definición, algo a evitar. Adquiere, así, un peso normativo inmediato, no porque sea esencialmente *malo* en un sentido ontológico, sino porque es vivido como tal por cualquier ser con capacidad de sentir. Es una función fisiológica básica para la supervivencia, pero su valor normativo no se justifica por su capacidad para cumplir este objetivo, pues afirmarlo implicaría caer en una falacia naturalista, sino que se justifica por la cualidad de la vivencia que supone al ser experimentado.

El sufrimiento, entonces, es un dato negativo inmediato que limita o dirige la acción, y al fundamentar una ética históricamente sensible dota de las herramientas para afirmar que si algo se ha vivido como horror es motivo suficiente para determinarlo como límite ético. La

apelación a este criterio se basa, como he dicho ya, no en una certeza ontológica, sino en una convicción crítica provisional, cuya legitimidad depende de su resonancia con experiencias compartidas. Esta convicción, al igual que todo contenido estructurado dentro del humanismo crítico, permanece abierta a reconstrucción si las condiciones empíricas o experiencias intersubjetivas así lo exigen.

Fuera de un escenario catastrófico donde se desfigura su efectividad, el humanismo crítico permite inmediatamente la discusión activa y constante ante escenarios de elevada urgencia, como es el caso del desarrollo de la inteligencia artificial, cuyo esparcimiento crece a pasos agigantados y se sustenta en narrativas sesgadas (se profundizará en ello en el capítulo siguiente). No necesitamos entender todas las cuestiones técnicas implicadas para elaborar una crítica al desarrollo de la IA; sí, por el contrario, su potencial deshumanizante y las estructuras de poder que le subyacen.

Axel Honneth (2007) muestra que existe una forma de crítica social *reveladora*, como la efectuada en la *Dialéctica de la Ilustración*, de un contexto donde las instituciones y prácticas pueden ser entendidas como *patológicas* al contradecir las condiciones de la *vida buena*. La aceptación de la veracidad de tales descripciones depende del grado de concordancia que los miembros de la sociedad atribuyan en relación a sus experiencias, en donde una respuesta positiva modificará las prácticas de su vida social. Aquí se esboza la racionalidad comunicativa habermasiana, donde lo humanamente bueno no se define a priori, sino que es resultado de un proceso de deliberación inclusiva. Ello funge como complemento al criterio negativo previamente establecido, donde este último evita caer en idealismos. Se mantiene el sentido de agencia social para evaluar las propuestas teóricas, pero implica también enfatizar en que tal evaluación debe realizarse de la manera más crítica posible dado que servirá como base para la acción cotidiana.

Esto último es la justificación para todo el énfasis en la imagen del ser humano: la construcción que sea aceptada como válida incidirá en las acciones del mundo real. Ello era ya

central en las preocupaciones de Fuchs, aunque con una respuesta ontológica: “Incluso una redefinición ecológica de nuestra relación con el ambiente terrenal tendrá éxito solo si nuestra corporalidad y vivacidad –como relacionalidad o *convivialidad* con nuestro ambiente natural– se pone al centro” [Traducción propia] (Fuchs, 2021, p. 6). Por ello el fundamento crítico es esencial, porque al estar en juego la imagen del ser humano, como el propio Fuchs dice, está en juego la forma en que nos desenvolvemos y relacionamos en y con el mundo.

El proyecto de Fuchs sigue en pie con la forma de Teoría Crítica.

3.5: El Papel de la Crítica Social

La crítica social es la herramienta con la que se apunta al horizonte de lo humano mientras que se revelan las estructuras de poder que generan y perpetúan las prácticas deshumanizantes.

La investigación social crítica tal como fue propuesta por Horkheimer se orienta hacia un trabajo práctico o activo:

Permite captar la contradicción entre las pretensiones normativas inscritas en las escrituras y en la cultura de la sociedad burguesa y su realidad. [...] La tarea de la *filosofía* –o de la teoría en términos más amplios– no consistiría en diseñar un modelo para alcanzar un mundo mejor, pero tampoco se quiere limitar, tal como hace la *teoría tradicional*, a fungir como *hipótesis de investigación que demuestra su utilidad en la empresa imperante*. Más bien consiste en recordar la aspiración hacia una vida humanamente digna a través de reconocer el carácter contradictorio de nuestras sociedades. (Kozlarek, 2024b, pp. 73-74)

La imagen del ser humano funciona como punto de referencia para Horkheimer incluso ante las consecuencias radicalizadas de la modernidad, pues no es el humanismo ilustrado el

que ha fallado, sino su desfiguración en la sociedad burguesa que reproduce la racionalidad instrumental y que ha devenido en unas condiciones sociales indignas, donde los beneficios se ven relegados a una reducida élite. Kozlarek (2024a) lo explica como sigue:

A Horkheimer no le interesaba –como a los discursos posmodernistas, poscolonialistas o posthumanistas– culpar al humanismo de las atrocidades cometidas en la modernidad, sino mostrar que el humanismo bajo el dominio de la modernidad capitalista sigue siendo siempre solo un privilegio de las élites. Esto, sin embargo, no hace que el humanismo sea erróneo, sino que la sociedad capitalista lo es. No menos, sino *más* humanismo e igualdad de acceso a sus promesas sería la reacción consecuente tras esta percepción. [Traducción propia] (p. 79)

Así, la crítica social denuncia la deshumanización y sus estructuras.

La crítica social es, de esta forma, condición necesaria para construir lo humano, pues es solo en el desocultamiento que la crítica permite donde lo humano, en tanto referencia normativa, se hace visible. Lo humano no es una esencia, sino una posibilidad que emerge cuando se cuestionan las formas de vida que lo niegan. El sufrimiento causado por las condiciones materiales modernas hizo necesario encontrar lo que dichas condiciones contenían de inhumano; ante ello la Ilustración permitió elaborar tales denuncias y su contenido pasó a considerarse parte de lo humanamente real. Pero establecer esta visión de ser humano ilustrado solo fue posible por una actitud crítica previa. Esta crítica, entonces, no es solo una herramienta del humanismo crítico, sino también condición de su posibilidad.

El contenido de lo humano se construye frente a su negación, y la crítica social revela precisamente aquello que no debería ser. Al destapar el sufrimiento y la alienación comienza a dibujarse el contorno de lo humano, pues aunque no tengamos la imagen definitiva del ser humano, sus características se vuelven explícitas allí donde su ausencia genera condiciones intolerables.

La crítica social, de esta forma, impide la normalización de lo inhumano, pues donde las prácticas que devienen en sufrimiento y alienación se presentan como inevitables, necesarias o naturales, la crítica recuerda que toda estructura es construida y, por tanto, transformable. Hace visible y enfatiza la posibilidad de la posibilidad de lo humano.

Finalmente, dado que el humanismo crítico no promete una imagen definitiva de lo humano pero exige la apertura de su definición, esto último solo puede lograrse si existe la crítica social como ejercicio permanente de vigilancia ética. Mientras haya crítica habrá horizonte.

Así, pues, una crítica de inspiración frankfurtiana hacia la modernidad del riesgo deberá revelar las estructuras que desfiguran la noción del ser humano en la modernidad tardía en la forma de producción y perpetuación de riesgos, así como lo humano que se encuentra implícito en la inconformidad con los modos de vida que resultan de tal estructuración social. Deberá dismantelarse todo discurso que pretenda otorgar un estatus natural a las consecuencias radicalizadas de la modernidad en tanto mancilladoras de la humanidad, poniendo sobre la mesa una vez más el debate sobre la imagen del ser humano y haciendo explícito que la sociedad del riesgo es, de hecho, contingente, para mantener, abierto el horizonte de lo que la sociedad puede llegar a ser.

Capítulo 4: La Definición Discursiva de la Inteligencia Artificial: Un Abordaje Crítico de sus Imaginarios

La inteligencia artificial no es solo un fenómeno técnico ni una herramienta de automatización que interviene en sectores estratégicos de la vida contemporánea: es también un conjunto de imaginarios. En su despliegue institucional, discursivo y mediático, la IA se ha convertido en un campo en el que se articulan visiones del mundo, valores, jerarquías y horizontes de sentido. Lo que suele ser presentado como una tecnología neutral está, en realidad, profundamente atravesado por decisiones políticas, imaginarios sociales y afirmaciones filosóficas muchas veces implícitas. Por ello, abordar críticamente la Inteligencia Artificial exige abarcar las diferentes dimensiones de esta tecnología, a saber, técnica, filosófica y social, y preguntarse por los supuestos y los efectos que sus discursos arrastran.

Este capítulo final tiene como objetivo desarrollar una crítica a los imaginarios de la inteligencia artificial desde el humanismo crítico y con enfoque en sus riesgos asociados que potencialmente pueden devenir en sufrimiento, cuestionando así las formas en que la IA ha sido definida y legitimada en distintos espacios.

En particular, me interesa por una parte mostrar la forma en que alrededor de esta tecnología se han construido imaginarios sociotécnicos predominantes que presentan su desarrollo como inevitable y deseable, o que manipulan la conciencia del riesgo para legitimar la toma de decisiones, desactivando con ello el sentido de agencia social y clausurando el debate democrático sobre su dirección. Estos imaginarios, institucionalizados en políticas públicas, documentos estratégicos y narrativas mediáticas, conforman el suelo simbólico sobre el que se decide el tipo de tecnología que es imaginable y el tipo de sociedad que es deseable. Definen el riesgo, y con ello legitiman el progreso según intereses. La crítica a este fenómeno se desarrolla en el primer subapartado del presente capítulo.

Posteriormente, en el segundo subapartado analizo cómo los imaginarios se articulan también sobre afirmaciones filosóficas acerca de la naturaleza del ser humano, particularmente dentro del transhumanismo y posthumanismo tal como son comprendidos en el ámbito *popular*, esto es, fuera de la filosofía académica, y cómo estas corrientes a su vez se construyen sobre especulaciones ontológicas. A través de estos discursos, la inteligencia artificial pasa de ser una herramienta a convertirse en una promesa de la superación del ser humano, apoyada sobre presuposiciones materialistas, funcionalistas y neurocentristas que buscan redefinir lo ontológicamente humano. El problema, argumento, radica no necesariamente en el contenido de estas visiones, sino en el hecho de que no se reconocen los límites operativos de su fundamento metafísico, pues ya se proyectan como guías normativas y decisiones de operación, cruzando así ilegítimamente al ámbito práctico, con las consecuencias que ello implica.

Frente a esto, construyo una crítica desde la noción de riesgo y desde una visión de humanismo crítico que no toma al ser humano como medida fija y esencial, sino como punto contingente de referencia para evaluar las consecuencias éticas de nuestras acciones. Sostengo que toda propuesta tecnológica debe pasar por el filtro crítico del análisis de sus riesgos asociados, y que, por tanto, es necesario resistir el paso ilegítimo de la especulación ontológica a la prescripción normativa. Recuperar el sentido de agencia sobre el devenir tecnológico exige volver a situar la libertad, la experiencia y la ética como coordenadas ineludibles para pensar el futuro.

4.1: La Inteligencia Artificial como Proyecto Sociotécnico: Crítica de sus Imaginarios Predominantes

Beck y Giddens, como ha sido previamente expuesto, muestran la forma en que la incertidumbre y el riesgo derivados del progreso técnico politizan las discusiones sobre ciencia y tecnología en la tardomodernidad en una contienda por definir los riesgos. Cuando hablamos de

una tecnología tan conceptualmente variada y con aplicaciones prácticas altamente relevantes como es la inteligencia artificial conviene estudiar la manera en que se está llevando a cabo la definición de sus efectos potenciales y, sobre todo, las narrativas que buscan establecer una visión predominante sobre lo que implica esta tecnología, de manera que puedan ser identificados posibles sesgos e intereses que subyacen a las definiciones y que, por tanto, deban ser abordados críticamente.

Para efectuar esta tarea se tomará como base el trabajo de Sheila Jasanoff, teórica social indostadounidense, quien sigue la tónica de la definición de riesgos de Beck y Giddens y la identifica en la práctica discursiva como imaginarios sociotécnicos. Estos imaginarios “son visiones de futuros deseables sostenidas colectivamente, estabilizadas institucionalmente, y representadas públicamente, animadas por concepciones compartidas de formas de vida y orden sociales alcanzables mediante, y en apoyo de, avances científicos y tecnológicos” [Traducción propia] (Jasanoff, 2015, p. 4). Múltiples imaginarios pueden coexistir dentro de una sociedad, ya sea en tensión o a manera de una relación productiva, y su contenido no solo se limita a visiones de lo que es alcanzable mediante la ciencia y la tecnología, sino a cómo la vida debe ser vivida, expresando concepciones éticas socialmente compartidas. Resalta también que usualmente las instituciones de poder buscan establecer la primacía de algunos tipos particulares de futuros imaginados, elevándolos a una posición predominante para fines políticos.

Como mostraré aquí, uno de los peligros políticos más importantes en las sociedades tardomodernas es el de la mitigación del proceso deliberativo sobre la pertinencia del avance tecnológico, lo cual puede realizarse al manipular la percepción pública con el establecimiento de discursos que doten de contenido los imaginarios, ya sea que en estos se valore positiva o negativamente desarrollos tecnológicos de acuerdo a los intereses de grupos en el poder, o que, con el mismo interés, busquen establecer la idea de la inevitabilidad del desarrollo tecnológico, dotándolo de un carácter determinista y buscando desligar esta área del aspecto social que debería serle inherente. En otras palabras, se pretende, mediante la manipulación del

conocimiento, el control de la conciencia del riesgo característica de las sociedades modernas que confiere contenido político a todo aspecto de la vida cotidiana, ya sea preestableciendo definiciones en los debates respecto al riesgo del desarrollo tecnológico, o directamente buscando erradicar la mera idea de un posible sentido de agencia en este ámbito.

Jasanoff (2004, p. 1) afirma, además, que “al analizar muchos de los fenómenos definitorios de la historia humana – aquellos que se levantan al nexo de ciencia, tecnología, cultura y poder – amplios segmentos de las ciencias sociales parecen casi retirarse en una conspiración de silencio”, y “los discursos dominantes en economía, sociología y ciencia política carecen del vocabulario para dar sentido a los procesos desordenados e irregulares a través de los cuales la producción de ciencia y tecnología se entrelaza con las normas y jerarquías sociales” [Traducción propia].

Por el contrario, para adquirir las herramientas conceptuales adecuadas en la explicación de amplios campos de actividad humana presente y futura, es menester concebir como conjunto la producción de los órdenes natural y social. Tal es el caso particular de la modernidad, que solo puede ser entendida con propiedad al tener en cuenta este fenómeno de coproducción. La idea de coproducción se refiere al hecho de que las maneras en que representamos y conocemos el mundo, natural y social, son inseparables de las formas en que elegimos vivir en este. Implica, además, la noción de que el conocimiento y su materialización es un producto social y es constitutivo de las formas de vida social, pues ni la sociedad puede funcionar sin conocimiento, ni el conocimiento existir sin una base social apropiada. En este mismo sentido, el conocimiento científico no es una ventana trascendente a la realidad, sino que, al igual que su aplicación al desarrollo tecnológico, se encuentra arraigado a todo lo que conlleva el término *social* (Jasanoff, 2004).

De esta manera, el esbozo de Jasanoff se complementa con el diagnóstico de la modernidad tardía realizado por Beck y Giddens, pues es la reflexividad sobre el propio proceso de modernización la que incita a realizar evaluaciones sociales sobre los desarrollos tecnológicos

presentes y futuros. La comprensión generalizada de los riesgos potenciales del progreso técnico desmedido, que se transforma en incertidumbre, politiza todo ámbito de desarrollo, y tanto la tecnología como la ciencia se ven, y deben serlo, subsumidos a las discusiones sociales que tienen lugar a su respecto. La propuesta teórica de Beck y Giddens permite mantener en el foco la necesidad de deliberar socialmente las elecciones a tomar respecto al progreso de la tecnología y mantener una mirada crítica ante las definiciones del riesgo del avance tecnológico que se manejan desde los ámbitos de poder de poder y que buscan pautar las percepciones de la sociedad en este mismo ámbito.

Los estudios basados en la teoría de Jasanoff, entonces, permiten dilucidar de manera crítica la imposición, por parte de grupos de poder, de narrativas determinadas que desprenden al progreso de la inteligencia artificial de su carácter social.

4.1.1: La Búsqueda de Legitimación de la Inteligencia Artificial Mediante la Configuración Discursiva

La conciencia tardomoderna del riesgo somete a juicio todo desarrollo tecnológico que represente un riesgo potencial. El progreso técnico, por tanto, no está legitimado de antemano, sino que la sociedad en tanto agente participa de forma activa en la disputa por definir el presente y el futuro de un desarrollo determinado. Por ello, cuando surge una tecnología como la IA, cuyo progreso está directamente vinculado a intereses de grupos de poder que buscan afirmar su consecución, es inevitable que aparezcan discursos que intenten legitimarla de manera acrítica y sesgada para evitar obstáculos en su avance.

En el trabajo de Bareis y Katzenbach (2022) sobre los imaginarios sociotécnicos que buscan ser impuestos como predominantes se hace visible este hecho: tras el análisis de documentos estratégicos relacionados con el desarrollo de la inteligencia artificial en China, Estados Unidos, Alemania y Francia, países clave en el avance de esta tecnología y con una

posición económica y geopolítica tal que sus decisiones en materia tecnológica influyen mucho más allá de sus fronteras, estos autores identifican una serie de narrativas que construyen los imaginarios de la IA ahí expuestos. Una de estas narrativas es la concepción de la inteligencia artificial como ruta tecnológica inevitable.

Múltiples tópicos fundamentados en el determinismo tecnológico son observados en los documentos estratégicos sobre IA:

Para preparar el escenario, los líderes políticos sitúan sus sociedades en un contexto relacionado con la inteligencia artificial. Tal contexto histórico es retratado como una *ruptura* aparentemente sin precedentes que trasciende a cualquier experiencia social previa o como una ruptura que se erige sobre un legado de transformaciones históricas pasadas. Ambos motivos históricos convierten los desarrollos actuales de la IA en un agente autónomo, una fuerza determinista que irrumpe sobre nuestras sociedades.

[Traducción propia] (Bareis y Katzenbach, 2022, p. 864)

Esta supuesta ruptura radical niega la continuidad histórica: la excepción pretende ser la norma y todo lo anteriormente conocido se torna obsoleto, al grado que las metáforas e imágenes se vuelven necesarias para describir lo venidero. En la documentación estratégica china, por ejemplo, se explicita esta idea al comentarse que “el rápido desarrollo de la inteligencia artificial (IA) cambiará profundamente la sociedad y vida humanas y cambiará el mundo” [Traducción propia] (Bareis y Katzenbach, 2022, p. 864).

También se encuentra presente una concepción de la IA como inevitable en tanto culmen del desarrollo histórico lineal. La conexión entre innovación tecnológica y la retórica de las revoluciones pasadas es utilizada como argumento que promueve la celebración técnica: el progreso tecnológico como éxito, el éxito de los sucesos históricos pasados, y el dominio sobre todas las cosas son ideas entremezcladas en esta narrativa. Es en esta tónica que Kratasios, diputado asistente para el desarrollo tecnológico de los Estados Unidos, sugiere que “generación

tras generación, la innovación americana ha beneficiado a nuestra gente y al mundo entero. [...] Hoy, con tantos de los misterios de la computación cuántica, sistemas autónomos y aprendizaje automático aún por ser descubiertos, podemos tomar al futuro y apropiármolo” [Traducción propia] (Bareis y Katzenbach, 2022, p. 866). El desarrollo de la IA es, en este sentido, la continuación necesaria de un gran legado. En palabras de Jasanoff (2015, p. 22), “los sistemas tecnológicos sirven desde esta perspectiva a una doble función deíctica señalando hacia atrás a los logros culturales del pasado y hacia delante a futuros prometedores y alcanzables, o hacia futuros que deben ser rechazados y evitados” [Traducción propia].

Pero más allá del determinismo, aunque ligado, se encuentra también la retórica de la deseabilidad y necesidad de la IA, sobre todo en cuestión de dependencia para la competitividad internacional. Esta idea se esboza por ejemplo en documentos estadounidenses, donde se exponen discursos como el siguiente:

El fallo en la adopción de la IA resultará en sistemas heredados irrelevantes para la defensa de nuestra gente, erosión de la cohesión entre aliados y socios, acceso reducido a mercados que contribuirá a reducir nuestra prosperidad y estándar de vida, y el aumento de desafíos a sociedades que se han construido sobre las libertades individuales.

[Traducción propia] (Bareis y Katzenbach, 2022, p. 868)

Sobre esta línea, el desarrollo de la IA es ineludible si queremos alcanzar el bien social, pues es condición práctica para el desarrollo económico y la capacidad de resiliencia de una sociedad, confiriéndole un *aura de salvadora* (Bareis y Katzenbach, 2022). Ya sea con la finalidad de mantener una posición dominante, no quedar en el rezago o adoptarla como única forma de sobrevivir, la inteligencia artificial se muestra como eje para la competencia capitalista de mercado y el mantenimiento geopolítico hegemónico mediante sus aplicaciones militares y de vigilancia, al tiempo que se enfatizan sus aplicaciones prácticas cotidianas que prometen

mejoramiento de la calidad de vida, optimización del tiempo y automatización de las tareas mundanas, soporte para la salud física y mental, etc.

El avance tecnológico actúa como un pilar esencial del progreso civilizatorio en las sociedades capitalistas modernas. Si una tecnología *disruptiva* como la IA es identificada, al mismo tiempo que las naciones se localizan a sí mismas en una arena de feroz competencia internacional, los políticos magnifican el potencial de la IA para acelerar el crecimiento económico a fin de defender (u obtener) la posición global de la nación. Una vez más, de la misma forma que con el motivo del determinismo tecnológico, el avance de la IA aparenta ser tan vital como para que la resiliencia de una sociedad entera dependa de ello. Si la economía, la seguridad, y, en concordancia, el orden social como un todo se encuentran en juego, como la narrativa sugiere, solo el avance en la tecnología de IA puede asegurar que el nivel de vida actual pueda ser mantenido y asegurar la prosperidad futura. [Traducción propia] (Bareis y Katzenbach, 2022, p. 868)

Finalmente, además de las narrativas de necesidad e inevitabilidad presentes en los documentos estratégicos, y dado que no son ideas explícitas ni necesariamente consistentes, sino subyacentes a aplicaciones prácticas diversas y específicas según cada contexto e interés, existe una tercera retórica, contradictoria a las anteriores, en la que se sugiere que las trayectorias futuras de la IA son articulaciones de esperanzas y miedos en lugar de ideas claras y realistas sobre el futuro que depara a esta tecnología.

La articulación de las expectativas y temores relativos a la inteligencia artificial es una construcción discursiva que provoca un impulso de movilización: sirve como apertura a la incertidumbre, desde lo cual se busca un fundamento sólido para recuperar la confianza, y ello explicita una necesidad de dirección para controlar las contingencias. Tal situación ofrece al estado una ventana para exigir iniciativas acríicas siempre que se orienten en apariencia a la

consecución de los futuros deseables y a la evitación de los futuros indeseables. Esta aparente necesidad de liderazgo facilita la toma de decisiones que de otra manera serían difícilmente aceptadas, pues van encaminadas a una gestión crucial. Es así como Macron se presenta a sí mismo sobre esta retórica: “como ya comprendieron, ustedes pueden contar conmigo –lo digo aquí sin inocencia alguna– para construir el renacimiento que Europa necesita” [Traducción propia] (Bareis y Katzenbach, 2022, pp. 869-870).

La narrativa política sobre las esperanzas y miedos está lejos de ser solamente informativa. Es, en su lugar, constitutiva, pues dota de un marco de sentido a los discursos y a las visiones de lo posible e imposible. También es posibilitadora en tanto abre la puerta a la actividad política en nombre de la gestión de las amenazas. Es encubridora en la medida en que ofusca los efectos sociales colaterales indeseables y los cuestionamientos sobre las estructuras de poder. Finalmente, es (des)legitimadora, pues aprueba determinados tipos de decisiones políticas y reacciones sociales a la vez que desaprueba otras (Bareis y Katzenbach, 2022).

Así, pues, los contenidos de las tres retóricas principales encontradas en los documentos estratégicos sobre la IA (inevitabilidad, necesidad/deseabilidad, y liderazgo para la gestión del riesgo) aunque varían de manera importante en sus contenidos convergen en una función final: la mitigación del sentido de agencia social. Este hecho es de vital importancia en un terreno práctico, pues estos imaginarios no se agotan en los discursos que los construyen, sino que moldean percepciones colectivas, orientan expectativas y definen lo que se entiende como posible o imposible en el debate público y en el ámbito de la práctica. Los contenidos de estas representaciones conforman marcos de sentido que influyen directamente en la forma en que las sociedades comprenden e imaginan el futuro de la tecnología. Por ello es crucial someter estos imaginarios a la crítica, pues solo al hacer visible su poder performativo se puede recuperar el espacio de deliberación y agencia social que tienden a clausurar. Esta es la tarea que me propongo realizar en el siguiente apartado.

4.1.2: Riesgo, Agencia Social y Co-construcción en la Naturaleza Sociotécnica de la Inteligencia Artificial

El control discursivo de los imaginarios sociotécnicos de la IA, organizados previamente en las categorías de inevitabilidad, necesidad/deseabilidad, y gestión de los riesgos (Bareis y Katzenbach, 2022), supone una herramienta diversa que, de una u otra manera y siguiendo uno u otro objetivo o interés particular, desarticula la participación social en el proceso de deliberación por el devenir futuro de una tecnología calificada como absolutamente clave. Pero, busco mostrar, estas narrativas no se sostienen sobre fundamentos sólidos sino sobre visiones en extremo reducidas sobre la relación entre tecnología y sociedad.

Una de las narrativas más poderosas que envuelven a la inteligencia artificial es, como se ha mostrado, su inevitabilidad, desde la cual se afirma que el desarrollo de la IA no solo es un proceso en marcha, sino que ya es irreversible. La tecnología es presentada aquí como un destino histórico, una fuerza autónoma que se despliega al margen de las decisiones humanas. No cabe, por tanto, preguntarse si debemos avanzar sino, a lo sumo, cómo hacerlo y cómo adecuarnos a ello. Se sustituye la deliberación política por la necesidad técnica.

Pero detrás de esta supuesta inevitabilidad no hay una ley natural sino una serie de decisiones políticas, económicas e institucionales tomadas por actores con intereses específicos. Es la voluntad de ciertos gobiernos, corporaciones tecnológicas, laboratorios de investigación y élites financieras la que impulsa el desarrollo de determinadas formas de IA. No estamos ante un desenlace neutro del progreso sino ante una selección activa de ciertos caminos tecnológicos sobre otros posibles. La inevitabilidad, entonces, no es un diagnóstico sino un mito de trascendencia tecnológica respecto del actuar humano y de la propia historia, o del culmen de estas, del que se deriva una despolitización que busca legitimar decisiones tomadas sin pasar por el filtro de la deliberación y la ética al simplificar la complejidad social del desarrollo técnico.

Así, pues, las amenazas de la intensificación de las prácticas de vigilancia masiva, la erosión del derecho a la privacidad, el sesgo algorítmico que perpetúa las discriminaciones estructurales, la precarización del trabajo por automatización, la concentración del poder en las *Big Tech*, el debilitamiento de la deliberación democrática y el reforzamiento de desigualdades sociales que pueden derivar del uso sin control de la inteligencia artificial no son entendidas como producto de decisiones situadas, intereses particulares o marcos regulatorios demasiado permisivos, sino que la materialización de esos efectos es presentada por esta narrativa como una consecuencia natural del progreso técnico, como efectos secundarios latentes. De esta forma, el sufrimiento social o la pérdida de derechos no aparecen como errores a corregir, sino como daños colaterales de un progreso que debía ocurrir de todos modos.

Dario Amodei, CEO de Anthropic, compañía dedicada a la investigación en inteligencia artificial, al discutir sobre el amplio espectro de los riesgos asociados a la irrupción de la IA afirma: “no creo que podamos detener el autobús de la IA. Hay seis o siete compañías tan solo en los Estados Unidos trabajando en esta área. Yo solo gestiono una de ellas. [...] Incluso si las seis compañías se detuvieran, entonces seríamos adelantados por China” (Officechai Team, 2025). Por otra parte Nithin Kamath, CEO de Zerodha, empresa india de inversiones, considera que “herramientas como ChatGPT y Claude dejan en claro que este giro no se trata de *si* sino de *cuándo* ocurrirá. Puede llevar unos pocos años o una década, pero es inevitable”, al hablar del impacto que esta tecnología puede tener, tanto de manera negativa como positiva, dentro de su ámbito (TOI Tech Desk, 2025).

En los ejemplos anteriores se evidencia el efecto de las narrativas deterministas: se neutraliza toda crítica al recubrir los riesgos con el velo de la fatalidad histórica, donde nadie es responsable porque nada podía haberse hecho al respecto. Está claro que para Amodei no es opción parar pues hacerlo no surtiría efecto alguno, y dadas las circunstancias la gestión de los efectos secundarios de este progreso es lo máximo a lo que se puede aspirar. Este discurso se ve reafirmado desde el extremo opuesto en el que se encuentra Kamath, pues en tanto afectado se

aboca no a reprobar activamente el uso desmedido de la IA, sino que acepta sus consecuencias y se prepara con cautela ante estas, pues a pesar de no haberla implementado todavía en su empresa *sabe* que es inminente, “es por eso que, a través de los años, K y el equipo técnico han estado obsesivamente haciendo nuestros sistemas más rápidos, más escalables, y preparados para el futuro” (TOI Tech Desk, 2025), afirma. Así, pues, tanto el productor de riesgos como el potencialmente afectado alimentan recíprocamente una retórica que elimina toda capacidad de agencia social ante el progreso de la inteligencia artificial.

Lo cierto es que incluso si aceptáramos que ciertos avances de la IA son ya muy difíciles de detener debido a las condiciones políticas y económicas que le rodean, ello no implica que no deban ser regulados, cuestionados o redirigidos. Confundir dificultad con inevitabilidad es ceder terreno al determinismo tecnológico, coartada para la concentración de poder, la naturalización de desigualdades y la imposición de futuros excluyentes. Nada en la historia es verdaderamente inevitable, porque todo desarrollo técnico responde a decisiones situadas en contextos sociales específicos, y, por tanto, puede y debe ser discutido.

La inevitabilidad de la IA es, entonces, un discurso muy arraigado, reproducido y mantenido por diferentes instituciones y actores que se ven involucrados directa o indirectamente, así como por las condiciones económicas y geopolíticas. El primer paso es, por tanto, a pesar de toda dificultad con la que pudiera encontrarse esta afirmación, recordar que todo avance tecnológico es contingente. Este debe ser siempre el punto de partida. Por supuesto, no se trata de negar que su desarrollo se encuentra enormemente condicionado, sino de entender que incluso tales condiciones se construyen sobre la propia contingencia. Este término (*contingencia*) debe ser el fundamento lingüístico en todo debate relativo al futuro de la IA, pues permite identificar eficazmente sesgos deterministas que buscan ofuscar la idea de agencia ya desde incluso un ámbito meramente terminológico.

El avance de la Inteligencia Artificial, en tanto contingente, debe orientarse, como proponía Beck, bajo el principio de *prueba* y no el principio de *pago*, es decir, debe antes que

todo haber una certeza suficientemente fuerte de las consecuencias de este desarrollo de manera que se considere *probado* que sus riesgos están minimizados y controlados, en lugar de actuar a posteriori *cobrando* a sus responsables cuando se hayan ya provocado daños irreversibles.

Por otra parte, la segunda narrativa descrita por Bareis y Katzenbach, la necesidad de la IA, opera como el reverso funcional de la inevitabilidad. Si esta última desactiva el debate apelando al destino, la deseabilidad lo silencia apelando a la esperanza. En el discurso tecnopolítico contemporáneo, la inteligencia artificial es expuesta como un horizonte cargado de promesas: eficiencia radical en la administración pública, mejoras sin precedentes en los sistemas de salud y educación, soluciones algorítmicas a problemas ambientales complejos, crecimiento económico exponencial y democratización del conocimiento. Esta visión futurista, casi utópica, produce un consenso anticipado: parecería no haber razones para frenar el avance, pues detener el desarrollo de la IA sería equivalente a renunciar a un bien colectivo evidente.

Por supuesto, esta imagen de la IA como instrumento deseable no surge de un proceso transparente de deliberación ni de una evaluación crítica de sus efectos posibles. Es una imagen construida estratégicamente desde posiciones de poder que orientan el deseo social en una dirección específica al tiempo que neutralizan la aparición de deseos alternativos. En otros términos, no se desea la IA en su sentido abstracto, sino como modelo particular que ha sido cargado previamente con valores subsumidos al orden económico vigente: productividad, eficiencia, rendimiento, control, crecimiento. Este tipo de construcción busca legitimar decisiones técnicas sin participación democrática, reduciendo la capacidad crítica de la sociedad al sustituir el juicio por el entusiasmo.

El cuestionamiento crítico, en este marco, supone obstruccionismo o atraso. La tecnopolítica del deseo desplaza la discusión ética hacia la gestión del impacto, relegando lo normativo a lo técnico. Así, los potenciales riesgos derivados de la implementación descontrolada de la IA, como la vigilancia masiva, los sesgos algorítmicos, la concentración del poder o la deshumanización del trabajo, no se consideran motivos para detener o modificar el

rumbo, sino efectos secundarios apenas tematizados. Aquí, los riesgos se muestran como consecuencias colaterales provenientes de un bien mayor; y sin embargo, estos efectos secundarios fundamentan la dinámica de la sociedad del riesgo.

El riesgo como fundamento de las dinámicas sociales no es azaroso ni catastrofista, sino consecuente con una serie de efectos negativos asociados al progreso desmedido, y el papel que cumple se hace evidente dentro del ámbito de la IA. Mark Zuckerberg, CEO de Meta, promueve un optimismo radical:

Una superinteligencia ayudará a la humanidad a acelerar el ritmo del progreso. Pero quizá incluso más importante es que la superinteligencia tiene el potencial para dar inicio a una nueva era de empoderamiento personal donde la gente tendrá mayor agencia para mejorar el mundo en las formas que elijan. Sea cual sea la abundancia producida un día por la inteligencia artificial, un impacto incluso más significativo en nuestras vidas probablemente provenga de tener una superinteligencia personal que te ayude a conseguir tus metas, crear lo que quieres ver en el mundo, experimentar cualquier aventura, ser un amigo mejor con aquellos que te importan, y crecer para convertirte en la persona que aspiras a ser. (Zuckerberg, 2025)

Y sin embargo, a pesar de este optimismo, en Meta han tenido que restringir y aumentar la seguridad en sus *chatbots* para evitar que entablen conversaciones con adolescentes sobre el suicidio (McMahon, 2025), tema especialmente delicado después de que se iniciara un proceso legal en contra de OpenAI, creadores de ChatGPT, al haber provisto este último de información sobre métodos de suicidio a un adolescente estadounidense de 16 años (Yang et ál., 2025).

La deseabilidad de una tecnología con aplicaciones tan diversas como la inteligencia artificial no puede ser presupuesta sin caer en sesgos. Lo que es más, estas narrativas recurren a un modelo de pensamiento propio de una época pasada: el de la sociedad de la carencia de la modernidad temprana, donde el progreso técnico se legitimaba con la satisfacción de las

necesidades básicas. Como dice el propio Zuckerberg, “hace apenas 200 años, 90% de la población eran agricultores que plantaban alimentos para sobrevivir. Los avances tecnológicos han liberado a gran parte de la humanidad para centrarse menos en la subsistencia y más en los objetivos que elegimos” (Zuckerberg, 2025), pero en tal afirmación se ignora el hecho de que aquella tecnología que permitió el desarrollo de las sociedades modernas ha tenido también consecuencias devastadoras, y que por tanto la satisfacción de la necesidad fisiológica no es una celebración final sino el punto donde se posibilita la toma de conciencia de los efectos negativos de este desarrollo y la necesidad de reorientar algo que es insostenible.

La conciencia del riesgo explicita la ilegitimidad del progreso técnico sin límites, y es así como se desarticula la narrativa de la deseabilidad y necesidad como inherentes a la inteligencia artificial, pues pretende establecer una carencia nueva, ya ni siquiera basada en la búsqueda de la supervivencia, sino en la especulación del potencial positivo que conlleva la aplicación de esta tecnología, pero la posible deshumanización que emana de esta y cuyos efectos ya comienzan a materializarse deja claro que el contenido de estos discursos es mera retórica sesgada y obediente a intereses del poder. La IA no es deseable ni necesaria por sí misma en la realidad, sino como ideal extremadamente reducido y acrítico sobre el futuro (y el presente).

Por otra parte y finalmente, para criticar la última narrativa descrita por Bareis y Katzenbach es menester referir a una de las advertencias que ya había explicitado Beck: la del peligro de convertir la conciencia del riesgo en fundamento de un totalitarismo legítimo. A diferencia de los relatos que glorifican la IA o la presentan como inexorable, el imaginario del liderazgo ante la gestión de la incertidumbre admite abiertamente que la IA puede ser peligrosa: reconoce sus riesgos y enfatiza la necesidad de actuar en consecuencia. Pero esta asunción de peligros no busca frenar o repensar críticamente el desarrollo tecnológico; al contrario, funciona como estrategia para reforzar su legitimidad, pues se argumenta que los riesgos no son solo asumidos, sino que están siendo gestionados de manera activa por los mismos actores que promueven la IA.

Bareis y Katzenbach (2022) analizan críticamente esta situación afirmando que independientemente de si las visiones o ideas mantenidas sobre el futuro serán o podrían ser alcanzadas, estas influyen poderosamente en la construcción discursiva:

Si tal marco político es negativo, enfatizando los riesgos y miedos que acompañan al *imparable* tren tecnológico del progreso, entonces los líderes nacionales ocupan un rol interventor a manera de salvadores, quienes pueden interferir responsablemente o al menos mitigar los peores escenarios posibles. Mediante esta retórica, decisiones más bien menos favorables son fáciles de justificar cuando se les compara con un sombrío escenario apocalíptico (por ejemplo, la erradicación de millones de trabajos a causa de la IA, la tecnología de IA provocando una carrera armamentística internacional), las partes interesadas serán más propensas a aceptar lo desagradable. De la misma forma, el mito de un futuro brillante de la IA (por ejemplo, la visión grandilocuente de un crecimiento económico sin precedentes, la automatización de toda labor tediosa a través de la IA) es un medio conveniente para desencadenar un apresuramiento sin oposición hacia un simplificado e inocente futuro dorado, usualmente dejando de lado las complejidades sociales, políticas y económicas, las contradicciones y los peligros inesperados que acompañan a la nueva innovación. [Traducción propia] (p. 870)

De esta forma se evitan cuestionamientos externos o demandas de participación democrática. En lugar de una crítica al sistema, se instala una tecnocracia del riesgo que traslada el problema del *si* al *cómo*, tecnificando la cuestión: no se trata ya de discutir la deseabilidad de ciertos desarrollos, sino de establecer mecanismos para desplegarlos *de manera segura*, enfatizando la confianza en los sistemas expertos que describía Giddens. Así se consolidan marcos regulatorios autorreferenciales donde los actores que crean los riesgos son también quienes diseñan los protocolos para enfrentarlos. Esta tautología refuerza su poder discursivo y político, blindándolos ante el escrutinio público mientras se mantiene intacta la dirección del

desarrollo tecnológico y, sobre todo, legitimando cualquier decisión que se tome al respecto dado que es *necesaria para el mejor devenir futuro*.

Esta idea torna la conciencia tardomoderna en combustible para el mismo sistema: se integra el discurso del riesgo para evitar su fuerza disruptiva, domesticando la crítica inherente y desviando las demandas sociales hacia protocolos, evaluaciones de impacto y comités de ética sin poder vinculante. Lo que debería ser un espacio de deliberación profunda sobre el rumbo de la humanidad se convierte en un problema técnico de mitigación de daños.

Así, pues, aunque el tema del desarrollo técnico vuelve a ser colocado en el plano de lo humano no se recupera el sentido de agencia social, pues la incertidumbre se convierte en excusa para la legitimación de un liderazgo que tiende al totalitarismo como medio para la clarificación de las expectativas, esperanzas y miedos, y para la consecución de lo deseable. Si el futuro de la IA es incierto, la búsqueda por alcanzar un futuro deseable y evitar los riesgos potenciales que pueden acaecer con un desarrollo y uso descontrolados de esta tecnología se convierte en un compromiso social, generalizando una opinión positiva sobre las decisiones tomadas por los líderes políticos y empresas mientras estas parezcan orientadas a la supuesta consecución de tales objetivos, incluso si tales decisiones suponen peligros a la democracia cuando no están abiertas a la deliberación conjunta.

Esta visión, por supuesto, no es neutral ni objetiva, sino que está estructurada activamente para mantener y fortalecer el poder de quienes promueven el desarrollo desmedido, disfrazando decisiones políticas y económicas como gestión de riesgos, y la incertidumbre consolida el rumbo tecnológico en lugar de cuestionarlo. Es fundamental que la conciencia crítica del riesgo no sea domesticada ni reducida a protocolo técnico, pues solo preservando el espacio de cuestionamiento social y ético se puede garantizar que las decisiones sobre el desarrollo de la IA reflejen intereses colectivos y no solo las consolidaciones del poder de quienes ya producen los riesgos. La deliberación crítica es, por tanto, requisito indispensable para evitar que la gestión del riesgo se transforme en justificación de un autoritarismo encubierto.

En conclusión, estos tres imaginarios forman un entramado que configura no solo nuestra comprensión de la inteligencia artificial, sino que condiciona las formas legítimas de actuar frente a esta. La función de estas narrativas es clara: la desactivación del debate crítico sobre el rumbo tecnológico. La idea de que la IA es un proceso imparable, deseable o ya gestionado anestesia la crítica y reduce la complejidad del problema al ámbito técnico, excluyendo así a la sociedad de su deliberación y reduciendo la agencia a obediencia pasiva.

Las potenciales prácticas deshumanizantes que supone el desarrollo desmedido de la inteligencia artificial ven su implementación legitimada: vigilancia masiva, manipulación algorítmica, erradicación de la privacidad, implementaciones bélicas, precarización laboral, fragmentación identitaria y exclusión estructural, entre muchos otros riesgos, se presentan como costos inevitables o males menores frente a un supuesto bien mayor: el progreso; y el sufrimiento que podría derivarse de estos procesos se invisibiliza o justifica. Esta lógica no solo despolitiza el sufrimiento, sino que lo integra funcionalmente en un sistema que renuncia a imaginar alternativas.

Las consecuencias ignoradas por estas narrativas explicitan la urgencia de reconstruir sus discursos desde una base distinta, una donde se reconozca el carácter social, histórico y contingente del progreso tecnológico. La inteligencia artificial no es una fuerza autónoma que imponga su curso sobre la humanidad, sino una construcción social atravesada por relaciones de poder, intereses económicos y decisiones políticas. Reafirmar este hecho es rescatar la posibilidad de agencia colectiva, de disputa y de transformación, y solo así podrá establecerse un espacio crítico donde no se acepten como ineludibles aquellas configuraciones tecnológicas que reproducen desigualdades, intensifican el control social o profundizan el dolor psíquico y existencial.

De esta manera queda fundamentada la factibilidad de sustituir estas narrativas por un humanismo crítico, donde la libertad y la agencia sostienen la posibilidad (y con ello la necesidad) de acción, y donde la denuncia del sufrimiento potencial ayuda a prevenir y

modificar escenarios de los que emerjan prácticas deshumanizantes. Superar los marcos impuestos implica denunciar la manipulación ideológica, rescatar el principio de responsabilidad y reinstaurar una ética del presente basada en la experiencia concreta de la libertad y el sufrimiento. En lugar de imaginar futuros que justifican violencias actuales, es preciso imaginar futuros donde la tecnología sea compatible con la dignidad, la justicia y el buen vivir. La idea de lo que ello implica es, por supuesto, construible críticamente, y por tanto no implica rechazar la inteligencia artificial sino disputarla: hacerla objeto de deliberación, de cuidado y de vigilancia constante, con el objetivo de evitar que las narrativas acríticas acaben conduciéndonos sin freno a nuevas formas de deshumanización. Esta es la función que cumple el criterio negativo del sufrimiento en el humanismo crítico: al mostrar los caminos que deben ser evitados se esboza el camino que de hecho debe ser tomado. Al mismo tiempo, la irreductibilidad práctica de la libertad y la contingencia inherente al desarrollo tecnológico fundamentan la posibilidad de elegirlo.

4.2: La Inteligencia Artificial como Proyecto Filosófico: Crítica a una Metafísica Aplicada

En los imaginarios contemporáneos que rodean a la inteligencia artificial se encuentran también, además de las preguntas técnicas y sociales, profundas afirmaciones filosóficas subyacentes. La IA no es solamente una herramienta de cálculo o automatización: se ha convertido en un punto de anclaje para pensar lo humano, redefinir sus límites y proyectar futuros posibles. En este sentido, ha comenzado a operar como una forma de filosofía práctica, muchas veces encubierta, que reconfigura categorías fundamentales como conciencia, libertad, cuerpo, identidad o agencia.

Esta dimensión filosófica ha propiciado discursos que ya penetran en ámbitos más allá de la propia filosofía. Por ello es menester tematizarla críticamente. Las narrativas dominantes

en torno a la IA se conforman en torno a imaginarios con una fuerte carga ontológica: visiones del ser humano como sistema de información, de la conciencia como cálculo, del cuerpo como obstáculo o de la técnica como destino. Estas representaciones no surgen de un pensamiento neutral, sino que están sostenidas por doctrinas materialistas, neurocentristas y/o funcionalistas que, al ser asumidas como evidentes, comienzan a orientar el desarrollo tecnológico, la formulación jurídica y la percepción social.

Por ello, en el presente apartado me propongo analizar el plano más profundamente filosófico de la inteligencia artificial, partiendo del entendido de que no se trata únicamente de una tecnología en expansión, sino de un marco ontológico en construcción. En particular, abordaré las formas en que el transhumanismo y el posthumanismo, tal como son manejados fuera de la filosofía académica en un ámbito más popular y mediático, condensan esta vena filosófica de la técnica, y en los riesgos que supone asumir sus supuestos especulativos acríticos como guías normativas de operatividad. Esta transición, argumento, conlleva consecuencias éticas profundas, especialmente en lo que concierne a la deshumanización y el sufrimiento potencial, así como a la erosión del sentido práctico de la libertad.

La descripción del transhumanismo y posthumanismo la efectuaré a partir de lo encontrado en la obra de Yuval Harari, pues si bien este autor no propugna directamente dichas corrientes de pensamiento, su trabajo panorámico pone de manifiesto cómo estas afirmaciones de carácter filosófico no provienen necesariamente de teorías filosóficas concretas, sino que se hallan implícitas en diversos ámbitos de la práctica tecnocientífica a manera de imaginarios.

Frente a ello, sostendré que la distinción crítica entre ontología y deontología que ya he elaborado en el capítulo anterior permite resistir la extrapolación directa de narrativas metafísicas en prescripciones normativas, fundamentado, en cambio, una normatividad humanista crítica sensible a las cuestiones prácticas. La crítica aquí desarrollada no pretende negar el potencial de la especulación, sino advertir contra su conversión acrítica en ideología. Lo

que está juego en el carácter filosófico de la inteligencia artificial no es solo lo que la tecnología puede lograr, sino lo que estamos dispuestos a permitir que defina por nosotros.

4.2.1: Transhumanismo y Posthumanismo

En *Homo Deus* (2016), Yuval Harari expone las tendencias históricas y tecnológicas que atraviesan distintos ámbitos científicos, culturales, sociales y filosóficos, mostrando cómo estas se articulan en torno a la redefinición de lo humano y configuran complejos imaginarios sobre el futuro de la especie humana, constituyéndose en discursos transhumanistas y posthumanistas,

La historia humana está marcada por la lucha contra tres grandes males: el hambre, la peste y la guerra. Aunque ninguno de estos problemas ha sido erradicado por completo, Harari (2016) afirma que gracias al avance científico, tecnológico y social, la humanidad ha logrado reducir significativamente su impacto. Por ejemplo, la producción de alimentos es suficiente para alimentar a toda la población mundial, las enfermedades infecciosas han sido controladas en muchas regiones, y los conflictos armados, aunque aún presentes, no son la norma constante.

Ante tales avances, el ser humano se encuentra ahora en una etapa histórica en la que los problemas fundamentales ya no definen la agenda global. Por lo tanto, la humanidad se ha puesto nuevas metas mucho más ambiciosas y radicales: alcanzar la inmortalidad, o al menos extender la vida indefinidamente; conquistar la felicidad plena y duradera; y alcanzar un poder casi divino sobre la realidad, utilizando la tecnología para alterar cuerpos, mentes y entornos a niveles antes inimaginables (Harari, 2016).

Este nuevo objetivo de la humanidad no es solo una aspiración individual, sino que está configurado en un proyecto colectivo y tecnológico que busca convertir al ser humano en algo mucho más que un organismo biológico vulnerable. Harari (2016) describe esta tendencia como un intento de convertirse en *dioses*, esto es, en seres con capacidades sobrehumanas, capaces de dominar la biología, la mente y posiblemente la conciencia misma. En este sentido, la

humanidad evoluciona de *homo sapiens*, un ser limitado por la biología y la muerte, hacia una especie que puede diseñarse a sí misma, moldear su propia evolución y, en última instancia, reescribir las reglas de la vida: un *homo deus*.

Este proceso está impulsado por el auge de tecnologías disruptivas como la ingeniería genética, la inteligencia artificial, la nanotecnología y las interfaces cerebro-computadora. Tales herramientas no solo prometen eliminar enfermedades o aumentar la longevidad, sino también redefinir lo que significa ser humano. Así, la humanidad se encuentra en la antesala de una transformación profunda que cuestiona ideas filosóficas, éticas y religiosas fundamentales sobre la identidad, la conciencia, la muerte y el alma.

El *Homo deus* es una figura conceptual que representa a los seres humanos que, en un estado postbiológico, hayan conseguido extender su vida de forma drástica, potenciar sus capacidades físicas y mentales más allá de los límites naturales, e incluso alterar su estructura biológica o fusionarse con máquinas. La evolución no es selección natural o acción divina, sino diseño inteligente humano. Pero esta figura no es necesariamente universal, es decir, no todos los seres humanos participarían por igual en esta transformación. Harari (2016) sugiere que una de las consecuencias más inquietantes de este proceso es que posibilita la emergencia de una nueva élite posthumana mientras que el resto de la población permanecería sin mejoras significativas, creando una brecha sin precedentes entre humanos *mejorados* y *naturales*.

Harari conecta este escenario con una evolución en las prioridades de la sociedad: mientras que en el pasado la autoridad moral venía de la religión y los derechos naturales, en el futuro podrían ser reemplazadas por nuevas éticas basadas en algoritmos, eficiencia y datos. Si antes el valor de un ser humano se definía por su *alma* o por su razón, en un futuro controlado por sistemas algorítmicos podría definirse por su relevancia en la red de procesamiento de datos. Esta transición plantea una redefinición del sujeto, de la moral y de la política.

El transhumanismo es una corriente de pensamiento que se basa en esta idea: propone la transformación radical del ser humano mediante el uso de tecnologías avanzadas. Parte de una

radicalización de la lógica ilustrada: si la razón y la ciencia pueden resolver los problemas humanos, ¿por qué no habrían de resolver también la muerte, la debilidad y la ignorancia? Llega al extremo de proponer que la tecnología no solo puede reparar lo disfuncional, sino mejorar lo que ya está sano. No solo busca, por ejemplo, curar la ceguera, sino desarrollar visión nocturna o infrarroja; no es evitar el Alzheimer, sino mejorar la memoria, la concentración y el aprendizaje a niveles sobrehumanos su objetivo.

La idea clave en el transhumanismo es que el cuerpo humano y el cerebro no son sagrados ni inmutables, sino plataformas, similares a un hardware que pueden ser reprogramadas, rediseñadas o incluso sustituidas. Esto abre la puerta a múltiples líneas de modificación: desde implantes neuronales, edición genética y aumento protético, hasta interfaces cerebro-computadora que permitan el control mental de dispositivos, o incluso la transferencia de conciencia a sistemas digitales (el *mind uploading*). Harari destaca que la tecnología de corte transhumano se encuentra ya en desarrollo o uso incipiente, especialmente en los ámbitos médico y militar.

El transhumanismo es en el fondo una forma moderna de religión secular donde el ser humano se convierte en su propio dios y redentor. En lugar de rezar a una divinidad trascendente, el transhumanismo propone invertir en tecnología, investigación y desarrollo como nuevos templos del progreso. Así, la promesa religiosa de la inmortalidad se traslada al laboratorio. Esto convierte al transhumanismo en algo más que una corriente técnica: es una narrativa histórica y existencial que redefine el sentido mismo de la vida, la muerte y la humanidad.

Pero mientras el transhumanismo busca mejorar al ser humano, el posthumanismo, según explica Harari (2016), representa un salto incluso mayor: la superación o reemplazo definitivo del *Homo sapiens*. No se trata ya de modificar al humano biológico, sino del surgimiento de entidades completamente nuevas que no puedan ser identificadas como humanos ni en cuerpo ni en mente.

En este marco, el posthumanismo plantea que el ser humano no es el pináculo de la evolución sino un estado transitorio. Así como el *Homo Erectus* dio paso al *Homo sapiens*, este último podría dar paso a formas de vida más avanzadas, creadas por diseño inteligente; creadas no por dioses míticos ni fuerzas naturales sino por la ingeniería humana. Estas nuevas entidades podrían ser sistemas de inteligencia artificial conscientes, seres híbridos o incluso inteligencias puramente digitales sin la necesidad de un cuerpo físico.

Harari (2016) señala que en el corazón del posthumanismo se encuentra una crítica al humanismo moderno: si el humanismo había establecido al ser humano como medida de todas las cosas, como dotado de razón, alma, dignidad y derechos, el posthumanismo lo relativiza. En un mundo donde las inteligencias artificiales sean más inteligentes que nosotros, más eficientes, menos vulnerables a los sesgos emocionales o al dolor físico, ¿qué sentido tendría seguir viendo al ser humano como el centro de valor o de autoridad?

Una consecuencia radical del posthumanismo es la pérdida del monopolio humano sobre la conciencia y la agencia. Ya no seríamos los únicos con voluntad, lenguaje o creatividad. Las máquinas podrían componer música, escribir poesía, diseñar políticas públicas o incluso experimentar estados emocionales. En este contexto, el concepto mismo de *alma* o de *yo* podría desvanecerse, sustituido por procesos algorítmicos mucho más eficientes y tal vez incluso *más conscientes*, aunque aún no comprendamos del todo lo que ello significa.

Harari (2016) plantea, además, que el transhumanismo y el posthumanismo desafían las bases éticas y políticas del mundo moderno. Si los derechos humanos se fundamentan en la idea de una naturaleza común, ¿qué ocurre cuando esa naturaleza ya no es compartida, o es incluso obsoleta? ¿Tendrán derechos los algoritmos? ¿Tendrá sentido hablar de dignidad humana en un mundo posthumano? ¿Cómo redefiniremos la ley, la educación, el sufrimiento o la justicia?

Finalmente, Harari (2016) no presenta a estas posturas como meras fantasías de ciencia ficción, sino que, considera, son una posibilidad filosófica y técnica seria, cuya discusión se ha trasladado ya a laboratorios, gobiernos y empresas tecnológicas. El desarrollo de inteligencias

artificiales avanzadas, la posibilidad de subir conciencias a la nube o de fabricar cerebros sintéticos ya no son meras especulaciones, afirma, sino líneas activas de investigación. Ello obliga a la humanidad a pensar con urgencia no solo lo que queremos hacer con la tecnología, sino lo que queremos ser.

4.2.2: Los Límites Operativos de la Especulación Ontológica

El transhumanismo y el posthumanismo postulan, pues, transformaciones radicales del ser humano. Sin embargo, tales postulados no se basan en hechos empíricos concluyentes, sino en especulaciones ontológicas: visión computacional de la mente, reduccionismo del cuerpo, fe en la mejora ilimitada de la especie, etc. Estas afirmaciones parecen tomar especial fuerza a raíz de la popularización y el prestigio que han ganado las neurociencias o la nanotecnología, por nombrar un par de ejemplos, así como algunos modelos predominantes en ciencias cognitivas.

Pero no existe a día de hoy resultado científico que avale directamente estas nociones. De las neurociencias suele derivarse el neurocentrismo, que refiere a la idea de que la mente y la totalidad de la experiencia humana pueden ser reducidas a conexiones neuronales. Pero una lectura crítica explicita sus límites: el *problema duro de la conciencia*, es decir, el proceso en que las conexiones neuronales se transforman en la experiencia vivida y la conciencia, intencionalidad, voluntad y demás funciones complejas, sigue siendo una incógnita sin resolver, e incluso las posturas que objetan a la mera existencia de tal problema lo realizan desde nada más que la especulación y la confianza en que será resuelto inminentemente.

Desde la neurociencia moderna no es posible elaborar sino una suposición de que en caso tal que los resultados sigan teniendo la misma forma, la misma naturaleza que han presentado hasta ahora, y con el mantenimiento del actual marco interpretativo, podrían tender a la comprobación de los supuestos materialistas. Pero no debe dejarse de lado la posibilidad de que en la complejidad del cerebro, órgano de cuyo funcionamiento somos altamente ignorantes,

los resultados a largo plazo desarrollen una tendencia distinta, o que tras un cambio de paradigma en la epistemología neurocientífica se reconstruyan las conclusiones con otro tenor.

Sin embargo, ni el neurocentrismo ni ningún discurso que se base en este dan importancia a dicho límite; al contrario, la constatación del materialismo por parte de las neurociencias se da por sentada: se considera inminente o, peor aún, ya consumada. Estas narrativas incorporan a su retórica, además, los avances en inteligencia artificial, como es el caso de los modelos de lenguaje que imitan funcionalmente la expresión humana, o la generativa, que imitan de la misma forma la creatividad, incluso si debajo de ello no hay nada más que un modelo predictivo que calcula la probabilidad de acierto de una respuesta; tal respuesta no posee carácter de significancia alguno para la máquina, sino que es un conjunto gigantesco de meros estados de actividad e inactividad eléctrica. Circula un *bon mot* en internet que afirma que una computadora es, básicamente, una roca a la que engañamos para que piense (‘*a rock we tricked into thinking*’); quizá deberíamos preguntarnos si no fue, por el contrario, la roca (o quien la programa) quien nos engañó al hacernos creer que piensa.

La misma tergiversación sucede con los avances en biotecnología, ingeniería biomédica y demás áreas relacionadas. Los resultados de estas prestigiosas disciplinas se toman como pruebas reiterativas de los postulados materialistas. Y sin embargo no existe ningún argumento filosófico ni científico lo suficientemente fuerte que fundamente esta extrapolación. Es una confianza en la inminencia de la comprobación científica del materialismo que no se sustenta en nada más que en algún tipo de fe, y a pesar de su fundamento puramente metafísico esta ontología se ha consolidado como narrativa predominante en muchos discursos dentro de los ámbitos público, científico y social.

Así se fundamentan el posthumanismo y el transhumanismo: presuponiendo la idea del ser humano como sistema de información, la conciencia como procesos neurofisiológicos o computacionales, el desdén por el cuerpo y la posibilidad de continuidad mental sin soporte

biológico. Una metafísica que se encubre entre relatos desarrollados a partir de la transfiguración de resultados científicos.

El problema no está, por supuesto, en la propia especulación, sino en la influencia que unas afirmaciones con bases tan endebles han adquirido para la toma de decisiones en laboratorios, empresas y gobiernos. De los marcos ontológicos especulativos se derivan postulados normativos y decisiones operativas, orientan los desarrollos técnicos y las preocupaciones jurídicas. Se discute ya la titularidad de derechos de algoritmos sin siquiera haber establecido críticamente la posibilidad de experiencia en estos:

Las características cerebrales responsables de la conciencia, de acuerdo con algunas prominentes teorías neurocientíficas, serán probablemente reproducibles por sistemas de IA. [...] Podría argumentarse que los sistemas de IA consciente son merecedores de consideración moral, y podría darse el caso en el que altos números de sistemas conscientes pudieran ser creados y hechos sufrir. [...] Las organizaciones involucradas en la investigación de la IA deben establecer principios y políticas para guiar la investigación, las decisiones de despliegue y la comunicación pública en lo que concierne a la conciencia. [Traducción propia] (Butlin y Lappas, 2025, p. 1)

Tales narrativas cruzan de manera ilegítima el límite entre especulación ontológica y normatividad práctica. La deontología del avance de la inteligencia artificial no se sustenta ya ni siquiera en el *ser*, sino en lo que *podría llegar a ser*. Se desarrolla una tecnociencia acrítica como justificación de agendas políticas y económicas sin deliberación democrática ni fundamento ético sólido. De esta manera la especulación se vuelve totalizante: olvida que no es la realidad misma, sino una forma de mirarla.

Harari menciona que estas nociones no son ya mera especulación sino líneas activas de investigación, lo que obliga a replantearnos la deseabilidad del progreso técnico. Por supuesto, estoy de acuerdo con esta última advertencia, no así con la primera afirmación: el hecho de que

la especulación adquiriera un carácter operativo no reduce su grado especulativo; solo muestra que las decisiones respecto a esta tecnología son tomadas basándose en un fundamento demasiado débil.

Aquí se hace evidente otra gran problemática que fue también un punto de partida crucial para desprenderme de la propuesta de Thomas Fuchs, quien de hecho critica directamente al transhumanismo y al posthumanismo: existe una transición directa de la ontología a la deontología.

Así, pues, incluso si dejamos de lado por un momento la ingenuidad de los discursos trans y posthumanistas, como ya he explicitado en el capítulo anterior, el salto directo de la ontología a la deontología no es autojustificable. No lo es porque la cualidad práctica de la libertad es innegable a pesar de cualquier estatus ontológico que podamos establecer o modificar respecto de su naturaleza: la libertad es condición de los actos, por ello es operativamente real, y además de esta se deriva el deber. Los discursos neurocentristas afirman que dado que la mente y la conducta humana no son más que conjuntos de conexiones neuronales que responden de acuerdo a interacciones físicas, no existe algo tal como la libertad; más bien nuestra conducta así como, en acepciones más radicales, el devenir del universo están determinados y no podemos hacer nada al respecto. Y sin embargo, siguiendo línea de Kant:

Incluso el más obstinado escéptico coincide en que, en lo que a la acción concierne, todos los escrúpulos sofisticados sobre una ilusión universalmente engañosa deben disolverse. De la misma manera, el fatalista más convencido, que es fatalista mientras se entrega únicamente a la especulación, debe todavía, tan pronto como se enfrenta a la sabiduría y el deber, actuar siempre como si fuera libre; y esta idea, además, produce de hecho la acción que le corresponde, y es la única que puede producirla [Traducción propia] (Kant, 1783/1999, p. 10, como se citó en Fuchs, 2021, p. 124)

La libertad práctica se encuentra, pues, en un plano distinto al de las discusiones ontológicas y metafísicas. Fuchs pensaba ir un paso más allá de Kant suponiendo que si la libertad en sentido práctico era fundamento suficiente para la ética, establecer la existencia ontológica de la libertad haría de ello un soporte inquebrantable. Y, sin embargo, cuando la libertad se establece ontológicamente y se le deriva la deontología incurrimos en diversas problemáticas nuevas. A una de estas me referí en su momento con la siguiente pregunta retórica: ¿y si una ontología distinta ganara legitimidad cultural o científica, volveríamos a reconstruir toda nuestra normatividad a partir de ella?

Ahora es evidente que la respuesta es afirmativa. Los discursos trans y posthumanistas fundamentados en una ontología materialista son los que hoy en día dan forma a la normatividad que orienta el desarrollo tecnológico. Ray Kurzweil, director de ingeniería de Google, pregonaba sin reparo alguno la inminencia de la singularidad y la inmortalidad, llegando al extremo de designar fechas exactas de cuándo ocurrirán estos avances (Morales, 2025), con las consecuencias operativas que tales afirmaciones de un alto mando de una de las empresas tecnológicas más poderosas del mundo implican.

Harari muestra que estas posturas plantean problemas respecto del valor de la ética humanista que se diluye en una sociedad futura junto a la figura del ser humano. La propuesta de Fuchs para la resolución de este problema concreto funciona solo en la medida en que aceptemos uniforme e invariablemente su ontología, y ello es por supuesto una mera fantasía, pero también, como he mostrado ya, un compromiso innecesario en términos pragmáticos.

Basta con retomar el criterio negativo del sufrimiento a la luz de los riesgos que supone la ingenuidad trans y posthumanista para establecer la inviabilidad ética de su operatividad normativa: la intensificación de la desigualdad que puede aparecer de las mejoras cognitivas, genéticas o cibernéticas que no estén al alcance de todos, consolidando una nueva clase biotecnológicamente privilegiada y marginando a quienes no alcancen tales condiciones; la medicalización de la vida y la obsesión por optimizar el cuerpo pueden generar nuevas

patologías psicológicas ligadas al rendimiento, la identidad o la obsolescencia del yo; la ruptura con el cuerpo o con la propia idea de humanidad puede derivar en una crisis existencial radical. Este tipo de transformaciones, lejos de garantizar bienestar, podrían provocar formas inéditas de alienación, ansiedad, fragmentación identitaria y dolor psíquico, especialmente en quienes queden excluidos de los futuros prometidos o no logren adaptarse a las nuevas normas técnicas del ser. Por tanto, lejos de ser neutras, las narrativas trans y posthumanistas deben ser sometidas a una crítica ética rigurosa, en tanto pueden intensificar el sufrimiento al redefinir lo humano desde parámetros tecnocientíficos excluyentes desarraigados de la experiencia humana.

Así, incluso si desde el transhumanismo o posthumanismo se presupone la inexistencia ontológica de la libertad, ello no puede invalidar la experiencia concreta e ineludible de la agencia. La acción, el juicio, la responsabilidad y la deliberación no dependen de haber resuelto ontológicamente el estatuto de la libertad, sino de la presencia de esta última en tanto condición de posibilidad para la acción. Por tanto, resulta indispensable separar ontología de deontología, y resistir cualquier intento de traducir una metafísica especulativa en una ética de la renuncia. Los potenciales riesgos derivados de las promesas trans y posthumanistas que pueden devenir en prácticas deshumanizantes productoras de sufrimiento hacen de altísima relevancia su crítica.

Evidentemente, las preguntas por la factibilidad del derecho y de la dignidad humana, y por la distribución del mejoramiento genético en el, según algunos, inminente mundo posthumano son mera especulación. En este sentido, la descripción también especulativa de sus riesgos asociados se legitima no por su cualidad interventiva sino profiláctica: no se niega directamente la posibilidad ontológica de estos postulados, pero sí se evalúa su aspecto deontológico al tener en cuenta sus posibles consecuencias prácticas. Esta actitud crítica basada en la prevención frena el desarrollo desmedido y con ello el descontrol de sus efectos no deseables, pero una vez establecida esta pausa es menester también tener en cuenta el potencial genuinamente beneficioso de la tecnología para evitar caer en una tecnofobia sin motivo real. Se

propicia, pues, un espacio de deliberación donde la finalidad es *probar* ante el escrutinio crítico el alcance de la tecnología, en lugar de *pagar* cuando el daño es ya irreparable, tal como proponía Beck.

Conviene recordar, además, la crítica al humanismo encontrada en el posthumanismo: el ser humano no puede ser la medida de todas las cosas si se pone el jaque su condición ontológica y su valor. Esta afirmación pierde peso en mi propuesta de humanismo crítico, pues el ser humano no es la medida de todas las cosas sino el punto práctico de referencia. La *medida* no está dada, sino que se construye en función de la capacidad que tales contenidos poseen para denunciar prácticas que puedan devenir en sufrimiento o alienación. Ello, una vez más, muestra la importancia de establecer una normatividad fuera de esencialismos pero también resistente a relativizaciones.

Así, pues, la crítica al transhumanismo y al posthumanismo aquí realizada no trata de negar la capacidad humana para imaginar futuros alternativos (pues vemos que esta imaginación puede tener también un carácter preventivo), sino recordar que toda especulación que pretenda volverse práctica debe pasar por el juicio ético y político. No todo lo técnicamente posible es deseable, ni todo lo imaginado puede asumirse sin consecuencias. Cuando se permite a una metafísica orientar nuestras acciones sin evaluación crítica se renuncia a la libertad en nombre de una promesa vacía. Frente a ello, sostener el carácter operativo de la agencia, reconocer el sufrimiento como límite y mantener una normatividad desvinculada de absolutismos ontológicos es la única forma de evitar que la técnica se transforme en ideología. La pregunta no es solo qué podemos llegar a ser, sino quién queremos ser, juntos. Ese es el núcleo de un desarrollo tecnológico responsable.

Conclusiones

A lo largo de la presente tesis se ha buscado mostrar que la inteligencia artificial debe ser entendida como un producto propiamente moderno que condensa las ambivalencias de este proyecto. La principal aportación crítica de Beck y Giddens es haber mostrado que el desarrollo tecnológico en esta época no se legitima por sí mismo ni garantiza de antemano un progreso beneficioso; al contrario, conlleva riesgos estructurales que deconstruyen la idea de un avance paralelo y necesario. En este sentido la IA es paradigmática: por un lado, *encarna* las promesas de eficiencia, innovación y emancipación; por otro, cristaliza las amenazas de deshumanización, exclusión y pérdida de agencia. De ahí que su legitimidad no pueda asumirse como efecto automático de su potencial técnico, sino como asunto político, ético y social que debe resolverse en el terreno de la deliberación activa desde la conciencia de sus riesgos.

El intento por contrarrestar esta politización de la técnica se explicita en el preestablecimiento de imaginarios que la rodean, la desvinculan de su carácter social y la orientan según intereses corporativos y estatales. Identificar la dimensión discursiva de la IA es fundamental para no olvidar que esta es, antes que nada, un proyecto sociotécnico arraigado a las condiciones sociales. Su legitimación verdadera depende, entonces, de los mecanismos participativos y transparentes que la sociedad sea capaz de construir para resolver sobre su rumbo según las condiciones concretas de la vida social, pues la conciencia del riesgo es la base primera para la crítica, pero su alcance se encuentra condicionado estructuralmente.

En este sentido, desde las instituciones no debe guiarse el avance de la IA según promesas futuras o especulaciones abstractas sino, en sentido negativo, por sus riesgos asociados: la generación de sufrimiento, exclusión y alienación, en condiciones de aplicación determinadas. Estos límites constituyen una brújula ética que acota el debate más allá de las ilusiones optimistas radicales o los escenarios catastrofistas: el foco se pone en aquello que la sociedad considera intolerable, lo que implica su participación como requerimiento.

Esta orientación es un escepticismo pragmático que, sin caer en el rechazo absoluto ni en la aceptación ingenua, guía la acción política y las prácticas cotidianas. En lugar de anclarse en marcos especulativos inaccesibles o en la opacidad técnica, se trata de evaluar la IA desde sus efectos presentes y verificables, en conjunto con su variable proyectada como fundamento profiláctico justificado por los efectos del progreso desmedido en el pasado cercano. Tal escepticismo pragmático opera como una forma de responsabilidad: recuerda que cada decisión tecnológica implica consecuencias inmediatas y futuras, y que tanto individual como institucionalmente deben cuestionarse a partir de su impacto sobre lo que se considera humano.

Por supuesto, insisto, una participación social efectiva en el debate sobre la IA no puede darse sin antes enfrentar y desmontar las narrativas que ofuscan el sentido de agencia social en este ámbito. Frente a estas discursivas, el humanismo crítico se ha presentado aquí como marco que no ofrece respuestas definitivas, sino herramientas conceptuales y críticamente normativas para sostener un diálogo abierto y reflexivo, capaz de contrarrestar los intentos de naturalizar un desarrollo tecnológico orientado por intereses ajenos a los de la sociedad en general. La afirmación de la contingencia tanto de lo técnico como de lo humano permite recuperar la agencia social y mantener activo el espacio de la deliberación, sin dar por inamovibles definiciones ni del ser humano ni de la inteligencia artificial. Lo aquí construido representa, así, un paso firme en esta dirección.

Pero mostrar los sesgos y transfiguraciones de los imaginarios predominantes alrededor de la IA es apenas el inicio de la tarea: sigue identificar y establecer formas concretas que puedan adoptar los mecanismos necesarios para la toma efectiva de decisiones a este respecto. La politización de la técnica ante el riesgo no garantiza por sí misma la formulación de políticas adecuadas: para que este malestar no se reduzca a una reacción emocional o a una percepción vaga, debe nutrirse de información clara y accesible que permita a los individuos comprender los alcances, las limitaciones y las consecuencias de las aplicaciones tecnológicas.

En este sentido, aunque el alcance de esta investigación llega a reafirmar la deliberación colectiva y evidenciar los sesgos implícitos en discursos generales que buscan clausurar el debate, decisiones particulares como las que se toman frente a usos concretos de la IA en ámbitos como la salud, la seguridad o la educación, requieren de un nivel analítico distinto: uno que combine la crítica filosófica con la evaluación técnica y el juicio contextual.

Uno de los mayores desafíos en la regulación y orientación ética de la IA radica en su aplicación a situaciones específicas donde los riesgos y beneficios se entrelazan de manera ambivalente o donde sus efectos desplegados a lo largo del tiempo son complicados de medir o prever. En estas situaciones, el carácter preventivo del énfasis en el riesgo indica la necesidad de aplicar el principio de prueba, esto es, que tal aplicación tecnológica debe esperar hasta ser probado el alcance y controlabilidad de sus riesgos. Sin embargo, para llegar a tal punto es necesario diseñar marcos de evaluación más finos y particularizados que permitan identificar con claridad las divisiones entre sus potenciales beneficiosos y deshumanizantes en circunstancias actualmente poco claras.

El requerimiento de tales criterios no se limita al ámbito institucional, sino que concierne también a la vida cotidiana. Decisiones como el uso de sistemas neuronales de recomendación, asistentes digitales o tecnologías de vigilancia pueden acumular impactos significativos sobre la libertad, la privacidad o la dignidad. Por ello, el desarrollo de metodologías críticas para evaluar riesgos y beneficios debe situarse tanto en el ámbito estructural como en el individual, reconociendo que los dilemas concernientes a la IA no se resuelven únicamente con principios abstractos, sino en el terreno concreto de las prácticas sociales.

Así, el reto consiste en articular lo general con lo particular, lo normativo con lo operativo, de manera tal que el carácter crítico no quede en mero concepto sino que se traduzca en decisiones efectivas y socialmente responsables. Este debe ser el núcleo para investigaciones posteriores sobre esta línea.

A este respecto es menester enfatizar también que, si bien la modernidad posee un carácter global al igual que los riesgos que se le derivan, las experiencias concretas de este proceso son múltiples y diversas. Tal es, por tanto, el caso de la inteligencia artificial: sus aplicaciones y riesgos asociados no se manifiestan de manera uniforme, sino que adoptan formas específicas según los contextos históricos, culturales y políticos en los que se despliega. De ello surge la necesidad de situar las discusiones de la IA de manera contextual, reconociendo la pluralidad de experiencias modernas para evitar caer en visiones homogéneas o universalismos simplistas.

El hecho anterior es consecuencia de que aquello que se considera humanamente deseable varía en función de las condiciones sociohistóricas. El humanismo crítico se muestra especialmente adecuado para afrontar esta diversidad, pues su apertura reconstructiva permite articular debates que tomen en cuenta las particularidades locales sin renunciar a horizontes normativos más amplios. Este enfoque propicia la deliberación situada, donde la definición de lo que se considera bueno se negocia entre distintos actores sociales en distintas circunstancias al reconocer las contingencias de cualquier contenido sobre *lo humano*. Los fundamentos generales son la libertad en sentido práctico y la experiencia del sufrimiento, pues, se ha argumentado, son mínimos normativos que no dependen de construcciones particulares; pero incluso estas bases, aunque las considero sólidas, pueden someterse a reconstrucción si así se determina críticamente. El humanismo crítico no es contenido, sino categoría. De esta manera se favorece una ética tecnocientífica que reconoce la globalidad de los riesgos pero atiende a la pluralidad de sus manifestaciones, evitando tanto el eurocentrismo como el relativismo absoluto. El debate, pues, no finaliza, sino que se reconstruye incesantemente de forma integral.

El afrontamiento de los riesgos y la orientación crítica respecto a la inteligencia artificial es, entonces, una tarea larga y compleja con una gran cantidad de retos por delante, pero lo aquí propuesto pretende ser una base abierta y fértil que sostiene la deliberación pública, filosófica y tecnocientífica necesaria para decidir colectivamente el rumbo adecuado para esta tecnología.

Referencias

- Adorno, T. (1998). La educación después de Auschwitz (J. Muñoz, Trad.). En G. Kadelbach (Ed.), *Educación para la emancipación* (pp. 79-92). Ediciones Morata S. L.. (Trabajo original publicado en 1966)
- Bareis, J. y Katzenbach, C. (2022). Talking AI into being: The narratives and imaginaries of national AI strategies and their performative politics. *Science, Technology, & Human Values*, 47(5), 855-881. <https://doi.org/10.1177/01622439211030007>
- Beck, U. (1998a). La política de la sociedad de riesgo. *Estudios Demográficos y Urbanos*, 13(3 (39)), 501-515.
- Beck, U. (1998b). *La sociedad del riesgo: hacia una nueva modernidad* (J. Navarro, D. Jiménez y M. R. Borrás, Trads.). Paidós. (Trabajo original publicado en 1986)
- Bishop, J. M. (2016). Singularity, or how I learned to stop worrying and love artificial intelligence. En V. Müller (Ed.), *Risks of Artificial Intelligence* (pp. 267-280). CRC Press. <https://doi.org/10.1201/b19187>
- Boden, M. A. (2016). *AI: Its nature and future*. Oxford University Press.
- Brundage, M. (2016). Limitations and risks of machine ethics. En V. Müller (Ed.), *Risks of Artificial Intelligence* (pp. 87-114). CRC Press. <https://doi.org/10.1201/b19187>
- Butlin, P. y Lappas, T. (2025). Principles for responsible AI consciousness research. *Journal of Artificial Intelligence Research*, 82, 1673-1690. <https://doi.org/10.1613/jair.1.17310>
- Dewey, D. (2016). Long-term strategies for ending existential risk from fast takeoff. En V. Müller (Ed.), *Risks of Artificial Intelligence* (pp. 243-266). CRC Press. <https://doi.org/10.1201/b19187>
- Farfán, R. (1999). Sociología reflexiva/sociedad reflejada: el diagnóstico de la modernidad de Anthony Giddens. En L. Girola (Coord.), *Una introducción al pensamiento de Anthony Giddens* (pp. 37-66). Universidad Autónoma Metropolitana Unidad Azcapotzalco.

- Flores, M. V. (2016). La globalización como fenómeno político, económico y social. *Orbis. Revista Científica Ciencias Humanas*, 12(34), 26-41.
- de Fonnegra de Jaramillo, I. (2000). El médico ante el sufrimiento del paciente que enfrenta su muerte. *Ars Medica, revista de estudios médicos humanísticos*, 29(2), 29-40.
<https://doi.org/10.11565/arsmed.v29i2.333>
- Fuchs, T. (2021). *In defense of Human Being*. Oxford University press.
<https://doi.org/10.1093/oso/9780192898197.001.0001>
- Fuchs, T. (2024). ¿Qué será del ser humano? Alegato a favor de un nuevo humanismo. *Devenires*, 25(50), 35-55. <https://doi.org/10.35830/devenires.v25i50.964>
- Gardner, H. (1987). *La nueva ciencia de la mente. Historia de la revolución cognitiva*. (L. Wolfson, Trad.). Paidós. (Trabajo original publicado en 1985)
- Giddens, A. (1994). *Consecuencias de la modernidad* (A. Lizón, Trad.). Alianza Universidad. (Trabajo original publicado en 1990)
- Harari, Y. N. (2016). *Homo Deus: Breve historia del mañana* (J. Ros, Trad.). Debate. (Trabajo original publicado en 2015)
- Honneth, A. (2007). *Disrespect. The normative foundations of critical theory*. Polity.
- Horkheimer, M. y Adorno, T. (1998). *Dialéctica de la Ilustración* (J. J. Sánchez, Trad.). Trotta. (Trabajo original publicado en 1944)
- Jasanoff, S. (2004) The idiom of co-production. En S. Jasanoff (Ed.), *States of Knowledge* (pp. 1-12). Routledge.
- Jasanoff, S. (2015). Future imperfect: science, technology and the imaginations of modernity. En S. Jasanoff y S.H. Kim (Eds.), *Dreamscapes of modernity* (pp. 1-33). The University of Chicago Press.
- Jobin, A. y Katzenbach, C. (2023). The becoming of AI: a critical perspective on the contingent formation of AI. En S. Lindgreen (Ed.), *Handbook of critical studies of artificial*

- intelligence* (pp. 43-55). Edward Elgar Publishing.
<http://dx.doi.org/10.4337/9781803928562>
- Kornai, A. (2016). Bounding the impact of artificial general intelligence. En V. Müller (Ed.), *Risks of Artificial Intelligence* (pp. 179-212). CRC Press. <https://doi.org/10.1201/b19187>
- Kozlarek, O. (2024a). *The Critical Humanism of the Frankfurt School as Social Critique*. Lexington Books.
- Kozlarek, O. (2024b). Humanismo Crítico como investigación social crítica A propósito del centenario del Instituto de Investigaciones Sociales en Frankfurt. *Valenciana*, 17(34), 59-88. <https://doi.org/10.15174/rv.v17i34.749>
- Luhmann, N. (2000). Familiarity, confidence, trust: Problems and alternatives. En D. Gambetta (Ed.), *Trust: Making and breaking cooperative relations* (pp. 94-107). Basil Blackwell.
- McMahon, L. (1 de septiembre de 2025). Meta to stop his AI chatbots from talking to teens about suicide. *BBC*. <https://www.bbc.com/news/articles/c2kzl79jv15o>
- Morales, O. (1 de abril de 2025). El hombre que predice la inmortalidad revela cuándo dejaremos de envejecer. *Infobae*.
<https://www.infobae.com/tecnologia/2025/04/01/el-hombre-que-predice-la-inmortalidad-revela-cuando-dejaremos-de-envejecer/>
- OfficeChai Team. (31 de mayo de 2025). Anthropic CEO Dario Amodei explains why AI progress can't be stopped. *OfficeChai*.
<https://officechai.com/ai/anthropic-ceo-dario-amodei-explains-why-ai-progress-cant-be-stopped/>
- Omohundro, S. (2016). Autonomous technology and the greater human good. En V. Müller (Ed.), *Risks of Artificial Intelligence* (pp. 9-28). CRC Press.
<https://doi.org/10.1201/b19187>

- Retamal, C. (2018). Contingencia extrema e inseguridad ontológica: condiciones determinantes de la formación ciudadana. En C. Berrios y C. García (Eds.), *Ciudadanías en conflicto* (pp. 137-154). Ariadna Ediciones.
- Sørensen, M. P. (2018). Ulrich Beck: exploring and contesting risk. *Journal of Risk Research*, 21(1), 6-16. <https://doi.org/10.1080/13669877.2017.1359204>
- TOI Tech Desk. (15 de Julio de 2025). Zerodha CEO Nithin Kamath on future of investing and trading in a world of AI: 'Tools like ChatGPT and Claude make it...'. *Gadgets Now by The Times of India*.
<https://timesofindia.indiatimes.com/technology/tech-news/zerodha-ceo-nithin-kamath-on-future-of-investing-and-trading-in-a-world-of-ai-tools-like-chatgpt-and-claude-make-it/articleshow/122360674.cms>
- Yampolski, R. (2016). Utility function security in artificially intelligence agents. En V. Müller (Ed.), *Risks of Artificial Intelligence* (pp. 115-140). CRC Press.
<https://doi.org/10.1201/b19187>
- Yang, A. Jarret, L. y Gallagher, F. (26 de agosto de 2025). The family of teenager who died by suicide alleges OpenAi's ChatGPT is to blame. *NCB News*.
<https://www.nbcnews.com/tech/tech-news/family-teenager-died-suicide-alleges-openai-s-chatgpt-blame-rcna226147>
- Zuckerberg, M. (30 de julio de 2025). Personal Superintelligence [Comunicado de prensa]. *Meta*. <https://www.meta.com/superintelligence/>

Anexos

Formato de Declaración de Originalidad y Uso de Inteligencia Artificial

Coordinación General de Estudios de Posgrado
Universidad Michoacana de San Nicolás de Hidalgo



A quien corresponda,

Por este medio, quien abajo firma, bajo protesta de decir verdad, declara lo siguiente:

- Que presenta para revisión de originalidad el manuscrito cuyos detalles se especifican abajo.
- Que todas las fuentes consultadas para la elaboración del manuscrito están debidamente identificadas dentro del cuerpo del texto, e incluidas en la lista de referencias.
- Que, en caso de haber usado un sistema de inteligencia artificial, en cualquier etapa del desarrollo de su trabajo, lo ha especificado en la tabla que se encuentra en este documento.
- Que conoce la normativa de la Universidad Michoacana de San Nicolás de Hidalgo, en particular los Incisos IX y XII del artículo 85, y los artículos 88 y 101 del Estatuto Universitario de la UMSNH, además del transitorio tercero del Reglamento General para los Estudios de Posgrado de la UMSNH.

Datos del manuscrito que se presenta a revisión		
Programa educativo	Maestría en Filosofía de la Cultura	
Título del trabajo	Las consecuencias de la modernidad en la era de la inteligencia artificial	
	Nombre	Correo electrónico
Autor/es	Alan Osvaldo Zamudio Ibañez	2350037g@umich.mx
Director	Oliver Kozlarek	oliver.kozlarek@umich.mx
Codirector	-	-
Coordinador del programa	Jesús Emmanuel Ferreira González	mae.filosofia.cultura@umich.mx

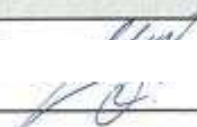
Uso de Inteligencia Artificial		
Rubro	Uso (sí/no)	Descripción
Asistencia en la redacción	No	
Traducción al español	Sí	Se utilizó como auxiliar para la traducción de las referencias de Kant (1783/1999) como se citó en Fuchs (2021).

Formato de Declaración de Originalidad y Uso de Inteligencia Artificial

Coordinación General de Estudios de Posgrado
Universidad Michoacana de San Nicolás de Hidalgo



Uso de Inteligencia Artificial		
Rubro	Uso (sí/no)	Descripción
Traducción a otra lengua	No	
Revisión y corrección de estilo	No	
Análisis de datos	No	
Búsqueda y organización de información	Sí	Se utilizó para la búsqueda de noticias actuales de alta relevancia para el tema, siendo estas TOI Tech Desk (2025); Yang, Jarret, y Gallagher (2025); OfficeChai Team (2025); McMahon (2025); y Morales (2025).
Formateo de las referencias bibliográficas	No	
Generación de contenido multimedia	No	
Otro	No	

Datos del solicitante	
Nombre y firma	
Lugar y fecha	Morelia, Mich. A 16 de diciembre de 2025.

Alan Osvaldo Zamudio Ibañez

Las consecuencias de la modernidad en la era de la inteligencia artificial.pdf

 Universidad Michoacana de San Nicolás de Hidalgo

Detalles del documento

Identificador de la entrega

trn:oid::3117:541575064

Fecha de entrega

17 dic 2025, 9:51 a.m. GMT-6

Fecha de descarga

17 dic 2025, 10:01 a.m. GMT-6

Nombre del archivo

Las consecuencias de la modernidad en la era de la inteligencia artificial.pdf

Tamaño del archivo

2.2 MB

117 páginas

36.321 palabras

206.070 caracteres

14% Similitud general

El total combinado de todas las coincidencias, incluidas las fuentes superpuestas, para ca...

Fuentes principales

- 14%  Fuentes de Internet
- 5%  Publicaciones
- 0%  Trabajos entregados (trabajos del estudiante)

Marcas de integridad

N.º de alertas de integridad para revisión

No se han detectado manipulaciones de texto sospechosas.

Los algoritmos de nuestro sistema analizan un documento en profundidad para buscar inconsistencias que permitirían distinguirlo de una entrega normal. Si advertimos algo extraño, lo marcamos como una alerta para que pueda revisarlo.

Una marca de alerta no es necesariamente un indicador de problemas. Sin embargo, recomendamos que preste atención y la revise.