

**Los dilemas de la inteligencia artificial.
De los sistemas sociotécnicos a la ética del diseño**

**Os dilemas da inteligência artificial.
Dos sistemas sociotécnicos à ética do design**

***The Dilemmas of Artificial Intelligence.
From Socio-Technical Systems to Design Ethics***

**Santiago José Roca Petitjean 
y Alejandro Elías Ochoa Arias **

Este artículo explora el debate acerca de la ética en el desarrollo de inteligencia artificial (IA), con el propósito de contribuir con la comprensión de sistemas sociotécnicos que integren la deliberación ética y la codificación de criterios de decisión en el desarrollo de sistemas inteligentes. Se inicia con un examen de los temas de ética y responsabilidad en el contexto de los sistemas inteligentes y los agentes autónomos. A continuación, se consideran los marcos interpretativos de bienes públicos (ciencia abierta) y bienes privados (ciencia mercantil), abordando los discursos de dos entes multilaterales y dos empresas, respectivamente. En este artículo se considera la coconstrucción de sistemas sociotécnicos donde se reconozca la pertinencia de la ética y la responsabilidad, se conciba el desarrollo informático como práctica social y las demandas éticas sean transferidas operativamente a los sistemas inteligentes a través de métodos de diseño orientado a valores. Finalmente, se discute la relación entre la ética y el desarrollo tecnológico con respecto a la cocreación de sistemas sociotécnicos y sistemas inteligentes que cumplan con criterios de responsabilidad.

191

Palabras clave: ética; responsabilidad; sistema sociotécnico; inteligencia artificial; diseño orientado a valores

* *Santiago José Roca Petitjean*: investigador del Centro Nacional de Desarrollo e Investigación en Tecnologías Libres (CENDITEL), Venezuela. Correo electrónico: sroca@cenditel.gob.ve. ORCID: <https://orcid.org/0000-0002-3701-3409>. *Alejandro Elías Ochoa Arias*: profesor de la Universidad Austral de Chile (UACH), Chile. Correo electrónico: alejandro.ochoa@uach.cl. ORCID: <https://orcid.org/0000-0001-8464-7108>.

Este artigo explora o debate sobre a ética no desenvolvimento da inteligência artificial (IA), com o objetivo de contribuir para a compreensão de sistemas sociotécnicos que integram a deliberação ética e a codificação de critérios de decisão no desenvolvimento de sistemas inteligentes. Começa com um exame das questões de ética e responsabilidade no contexto de sistemas inteligentes e agentes autônomos. A seguir, são considerados os marcos interpretativos dos bens públicos (ciência aberta) e dos bens privados (ciência comercial), abordando os discursos de duas entidades multilaterais e de duas empresas, respectivamente. Este artigo considera a coconstrução de sistemas sociotécnicos onde a relevância da ética e da responsabilidade é reconhecida, o desenvolvimento computacional é concebido como uma prática social e as demandas éticas são transferidas operacionalmente para sistemas inteligentes através de métodos de design orientado para valor. Por fim, discute-se a relação entre ética e desenvolvimento tecnológico no que diz respeito à cocriação de sistemas sociotécnicos e sistemas inteligentes que atendam a critérios de responsabilidade.

Palavras-chave: ética; responsabilidade; sistema sociotécnico; inteligência artificial; design orientado para valor

192

This article explores the debate about ethics in the development of artificial intelligence (AI), with the purpose of contributing to the understanding of sociotechnical systems that integrate ethical deliberation and the codification of decision criteria in the development of intelligent systems. It begins with an examination of the issues of ethics and responsibility in the context of intelligent systems and autonomous agents. Next, the interpretative frameworks of public goods (open science) and private goods (commercial science) are considered, addressing the discourses of two multilateral entities and two companies, respectively. This article considers the co-construction of sociotechnical systems where the relevance of ethics and responsibility is recognized, computer development is conceived as a social practice and ethical demands are operationally transferred to intelligent systems through value-oriented design methods. Finally, the relationship between ethics and technological development is discussed with respect to the co-creation of sociotechnical systems and intelligent systems that meet responsibility criteria.

Keywords: ethics; responsibility; socio-technical system; artificial intelligence; value-oriented design

Introducción

La inteligencia artificial (IA) se ha convertido en uno de los exponentes más significativos de la revolución digital. Esta premisa equivale a afirmar su importancia como parte de una revolución tecnológica, orientada a transformar los sistemas productivos y organizacionales, y por tanto a fortalecer el paradigma tecnoeconómico propio de las sociedades digitalizadas (Pérez, 2020). En otras palabras, la IA es, al mismo tiempo, una innovación parcial, parte de un conjunto de sistemas tecnológicos consolidados y motor de cambio socioeconómico y cultural. Éste se expresa en infinidad de propuestas académicas, empresariales, legislativas y gubernamentales que giran en torno a la necesidad de adecuar los procesos sociales y las organizaciones para la implementación de IA, explotar sus beneficios y asumir sus costos. Y en paralelo, se prepara la conformación de sistemas sociotécnicos mediados por la intervención de sistemas inteligentes y de agentes autónomos, en áreas tan estratégicas como las comunicaciones, la educación y el empleo, sin olvidar los sectores militar y médico que poseen características críticas relativas a la responsabilidad.

Las innovaciones tecnológicas tienden a transformar las relaciones socioproductivas, generando el abandono de las formas organizacionales precedentes y presionando para la conformación de nuevas relaciones (Pérez, 2020). Pero, al mismo tiempo, tales cambios preceden a las experiencias de aprendizaje cultural y de adaptación institucional a las nuevas condiciones productivas. Esto significa que la revolución digital contrae una mora con respecto a la generación de propuestas éticas, legislativas y políticas que permitan reducir los aspectos nocivos y aprovechar los beneficios de la innovación. Y también implica que, en general, se carece de un marco sociocrítico para interpretar y responder a las transformaciones tecnológicas. En tal sentido, es posible indagar en la tradición cultural en busca de propuestas para afrontar el impacto de las revoluciones tecnoeconómicas.

193

La ética puede considerarse la piedra angular de la filosofía política, el derecho y las políticas públicas, dado que en sus bases subyacen cuestiones sobre la naturaleza del bien, la importancia de las normas sociales o el significado de la vida plena. No obstante, hoy día la racionalidad instrumental y la racionalidad estratégica parecen dar forma a la interacción entre las personas y su entorno social, obstaculizando la deliberación propia de una racionalidad comunicativa, que debería apuntar a fomentar un consenso pluralista sobre esas cuestiones. En otras palabras, la consolidación de la separación entre el “sistema” y “el mundo de vida”, de acuerdo con Habermas (1999), se ha materializado en una colonización profunda del primero sobre el segundo. Por lo tanto, tras las narrativas tecnooptimistas y tecnopesimistas, se encuentra una suerte de determinismo tecnológico que induce a pensar que los problemas sociales merecen una respuesta mediada por la tecnología (o incluso, que los problemas socioeconómicos son consecuencia de la aplicación deficiente de soluciones instrumentales), y las posibilidades de generar dinámicas de deliberación comunitaria que tengan influencia en el espacio público (por encima de las lógicas de mercado) parecen extrañas y más aún, con escasa legitimidad.

Como consecuencia, la humanidad se encuentra frente a los retos que le impone la presente revolución tecnológica, con pocas herramientas para la conformación de un marco sociocrítico que permita comprender el lugar de la IA ante las exigencias que plantean materias como la ética, el derecho, la filosofía, la economía política y la gestión pública. De esa manera, es necesario abordar dos temas: cómo se construyen las alternativas éticas para convertirse en criterios de decisión, y de qué manera pueden ser transferidos éstos a los efectos de la producción tecnológico-cultural, como en el caso del desarrollo responsable de IA. Con el fin de abordar estas cuestiones, este trabajo plantea que la tecnología debe concebirse como un sistema sociotécnico, donde tienen relevancia los valores, intereses y relaciones de los agentes tecnológicos y aquellos que quedan al margen. De ese modo, la IA forma parte de “un sistema sociotécnico donde las nuevas tecnologías interactúan con elementos sociales” (Buijsman, Klenk & van den Hoven, 2025, p. 60). Tal enfoque permite no solo observar las relaciones sociotécnicas desde una perspectiva crítica, sino también ser creativos en la formulación de alternativas de respuesta a los problemas sociales desde la preocupación por la innovación abierta y responsable (Terrones, 2020).

194

Como parte de los sistemas sociotécnicos, la ética del desarrollo tecnológico se presenta como una exigencia que se realiza a los entes de regulación y de mercado en escenarios de conflicto. Si bien puede partirse de la tradición histórica de la ética, también es necesario reconocer las particularidades de la situación actual, con miras a señalar las condiciones que facilitan el reclamo de manejo ético de la tecnología y de aproximarse a las alternativas de prácticas de desarrollo éticas y responsables. De este modo, es necesario explorar cómo se construyen las opciones éticas en el caso de la IA, para entonces encontrar opciones teórico-metodológicas que permitan “abordar los desafíos de la ética de la IA combinando teorías éticas [...] con un enfoque de diseño” (Buijsman, Klenk & van den Hoven, 2025, pp. 76-77). Sin embargo, la preocupación por los valores y principios que deben guiar el desarrollo responsable de IA también debe contribuir con la fundamentación de políticas globales, iniciativas jurídicas, esquemas de diseño tecnológico y prácticas de mercado congruentes entre sí, que favorezcan una tendencia hacia la democratización de las dinámicas de apropiación social de las tecnologías.

En definitiva, este artículo se propone explorar la importancia de la ética en el desarrollo de la IA, con miras a ofrecer insumos para el diseño de sistemas sociotécnicos donde tengan cabida tanto la deliberación en torno a problemas fundamentales como la cocreación de bienes tecnológicos. En primer lugar, se parte de una exploración de temas como la ética y la responsabilidad frente al desarrollo de sistemas inteligentes y agentes autónomos. Posteriormente, se realiza una revisión de algunos discursos en torno a la ética y la responsabilidad en la implementación de IA, particularmente desde los contextos interpretativos de los bienes públicos (ciencia abierta) y privados (ciencia mercantil). Finalmente, se ofrece una discusión que se plantea interpretar la vinculación entre ética y producción tecnológica a la luz de la construcción de alternativas éticas, institucionales y de diseño pertinentes para la conformación de sistemas sociotécnicos de carácter más ético y responsable.

1. Ética y responsabilidad en la IA

Para abordar el tema es necesario tomar en cuenta diferentes enfoques éticos y su aplicación en el desarrollo e implementación de sistemas inteligentes. Puede plantearse que existen tres enfoques éticos conocidos: ética de la virtud, que se preocupa de los rasgos de carácter que deben cultivarse para tener una buena vida; consecuencialismo, interesado por las formas de comportamiento que contribuyen con el bienestar; y deontología, que se ocupa de normas universales de comportamiento (Buijsman, Klenk & van den Hoven, 2025). Respectivamente, la ética se origina en la concepción del individuo, en la percepción de los resultados o en la aceptación de normas consideradas universales. Estos enfoques pueden servir como guía para plantearse cuestiones clásicas de la ética, cómo por ejemplo qué es el bien, pero también para inferir distintos criterios de decisión, como si para alcanzar el bien deben respetarse ciertas normas o si deben buscarse resultados que contribuyan a crear condiciones de vida más deseables.

En ese sentido, tales enfoques pueden ofrecer información sobre qué es ético, cómo comportarse de forma ética, y en qué sentido los efectos de las acciones deben ser congruentes con lo anterior. En otras palabras, forman parte de los fundamentos de una ética aplicada, que debe contribuir a esclarecer las cualidades del desarrollo tecnológico al mismo tiempo que establecer un anclaje con contextos más amplios. Así, con respecto al diseño de sistemas, la cuestión podría enunciarse de la siguiente manera:

“¿Qué virtudes se deberían inculcar a quienes desarrollan y utilizan la IA? ¿Cómo puede contribuir la educación de los ingenieros a esto, para inculcar virtudes fundamentales como la conciencia del contexto social de la tecnología y el compromiso con el bien público y la sensibilidad hacia las necesidades de los demás?” (Buijsman, Klenk & van den Hoven, 2025, p. 71).

Estas preocupaciones han tomado vigencia recientemente con los avances de la IA, que posee mayores capacidades de actuación que otras tecnologías, pero que, del mismo modo, no posee agencia moral ni puede hacerse responsable por los resultados de su desempeño, lo que plantea la necesidad de establecer de forma sistemática principios éticos y de responsabilidad adecuados para su implementación (Buijsman, Klenk & van den Hoven, 2025). Por ejemplo, en el caso de los agentes autónomos, se ha planteado que éstos pueden encarar tres tipos de moralidad: operacional, propia de los diseñadores y usuarios; funcional, de los sistemas que pueden realizar cálculos morales; y genuina, en el caso de máquinas que puedan crear sus propios sistemas morales de decisión (Cortina, 2024). Por lo tanto, resulta necesario partir de los aspectos de ética y responsabilidad de los seres humanos, en cuanto que la reflexión sobre las implicaciones éticas de la tecnología tiene como eje al ser humano, sus interpretaciones e interacciones, incorporadas en sistemas sociotécnicos de carácter histórico y complejo (Coeckelbergh, 2021).

Resulta claro que cualquier marco ético que se intente insertar en un sistema inteligente es el resultado de la deliberación entre personas. En otras palabras, los conceptos éticos susceptibles de ser programados en las máquinas se derivan de la tradición histórica de reflexión sobre los valores éticos. Por lo tanto, no es sorprendente que la ética aplicable a los sistemas inteligentes pueda ser interpretada como una proyección de enfoques éticos en los procesos de producción de sistemas tecnológicos. Floridi (2023) realiza un planteamiento representativo del estado de la cuestión. Por una parte, reconoce los aportes de diferentes organizaciones en la elaboración de principios éticos para la IA, los cuales reúnen unos 47 axiomas. Posteriormente, los pasa a través del tamiz de la bioética para luego ofrecer una propuesta que incluye a los conjuntos mencionados:

- *Beneficencia*: mejorar la calidad de vida, respetar la dignidad humana y cuidar el medioambiente.
- *No maleficencia*: evitar causar daño, proteger la privacidad y la seguridad, vigilar las capacidades de la tecnología.
- *Autonomía*: empoderar a las personas para que tomen sus propias decisiones de manera libre y consciente.
- *Justicia*: fomentar la igualdad de oportunidades, mantener la cohesión social y prevenir el trato injusto.
- *Explicabilidad*: asegurar que los sistemas y sus decisiones sean comprensibles y que los responsables rindan cuentas por ellos.

196

El procedimiento de basarse en principios ha resultado fructífero para que diferentes organizaciones, públicas y privadas, ofrezcan sus propias perspectivas sobre los valores rectores en la implementación de IA, con diferente grado de abstracción (de las normas generales a las prácticas de desarrollo) y distinta orientación (de las normas abstractas a los estándares de mercado). Claro está, cualquier ente del ecosistema digital puede presentar sus propios juegos de principios. Por ejemplo, OpenAI (s/f) cuenta con una carta fundacional que plantea que sus objetivos son: beneficios ampliamente distribuidos; seguridad a largo plazo; liderazgo técnico y orientación cooperativa. En otro documento, se responde a la cuestión del desarrollo de IA responsable de la siguiente manera:

“El desarrollo responsable de la IA implica tomar medidas para garantizar que los sistemas de IA tengan un riesgo aceptablemente bajo de dañar a sus usuarios o a la sociedad e, idealmente, aumentar su probabilidad de ser socialmente beneficiosos. [...] Diremos que un sistema de IA se ha desarrollado de manera responsable si los riesgos de que cause daños están en niveles que la mayoría de la gente consideraría tolerables, teniendo en cuenta su gravedad, y que la cantidad de evidencia que fundamenta estas estimaciones de riesgo también se consideraría aceptable” (Askell, Brundage & Hadfield, 2019, p. 2).

Desde esta perspectiva consecuencialista, el desarrollo responsable de IA consiste en promover sistemas beneficiosos para la sociedad, mientras que resulten “aceptablemente” no dañinos. Un sistema inteligente ha sido desarrollado de forma responsable si el riesgo de que cause daño resulta tolerable para la sociedad. En este sentido, la ética del desarrollo es evaluada por la opinión pública (que puede expresarse de innumerables formas), y colocará en una balanza sus riesgos y sus beneficios. Sin embargo, esta aproximación plantea la cuestión de si un comportamiento puede considerarse ético solo por ser considerado beneficioso (o inocuo) por la opinión general o, en otras palabras, si existen diferencias entre la ética, la opinión de la mayoría y lo que se considera “verdadero” desde una posición de poder. Claro está, es necesario desgarnar diferentes posiciones para aproximarse a una comprensión de lo que puede entenderse como “responsabilidad” en cada contexto, y entonces procurar un diálogo entre las diferentes propuestas.

La responsabilidad puede abordarse como un caso especial de la ética. La facultad de rendir cuenta de las motivaciones, realización y efecto de las acciones debe entenderse en tres sentidos: como responsabilidad moral (énfasis en las intenciones de los agentes), responsabilidad causal (énfasis en las acciones) y responsabilidad de rol (énfasis en el lugar organizacional) (Lauwaert & Oimann, 2025). Esos tipos de responsabilidad pueden estar relacionados. Por ejemplo, una persona que causa un accidente de forma no intencional tiene una responsabilidad causal, pero no moral. En cambio, alguien que comete una acción maliciosa de forma intencional tiene responsabilidad moral y causal. La combinación de los tres tipos de responsabilidad puede dar lugar a diferentes formas de juicio en distintos contextos de acción, y se vuelve más compleja cuando se analizan escenarios donde agentes artificiales pueden ser causantes de accidentes o acciones maliciosas.

197

En este sentido, algunos autores opinan que la agencia de los sistemas inteligentes crea una “brecha de responsabilidad”, en cuanto que la responsabilidad es un atributo de los sujetos conscientes y no de los artefactos (Lauwaert & Oimann, 2025). En ese caso, si una máquina ocasiona un accidente, no solo es imposible imputar culpa, sino que tampoco se pueden señalar a otros responsables. Desde otra perspectiva, que no se pueda atribuir responsabilidad moral a una máquina no significa que no existan responsables. Entre otras razones, en los sistemas sociotécnicos participan agentes como reguladores, financistas, diseñadores, desarrolladores y usuarios, por lo que pueden considerarse diferentes modos de responsabilidad, distribuida entre los roles vinculados con la implementación de un artefacto. En ese sentido, el argumento de la brecha de responsabilidad está limitado como pretexto para liberar sistemas inteligentes sin supervisión, como se explica a continuación:

“La brecha de responsabilidad no significa que los desarrolladores y usuarios de sistemas de IA no tengan obligaciones especiales. Por el contrario, dicha tecnología afirma precisamente la importancia de los deberes morales. Debido a que el poder de toma de decisiones se está transfiriendo a esa tecnología, y que a menudo es imposible

predecir exactamente qué decisión tomará, los desarrolladores de sistemas de IA deben pensar incluso más cuidadosamente que otros diseñadores de tecnología sobre los posibles efectos indeseables que pueden resultar de las decisiones de los algoritmos” (Lauwaert & Oimann, 2025, pp. 104-105).

En este contexto, la falta de control directo sobre las acciones de los sistemas inteligentes no puede constituirse en favor de la ausencia de responsabilidad. En cambio, la declaración de responsabilidad refuerza la necesidad de implementar estándares de desarrollo que permitan mejorar aspectos como las normas de regulación y de calidad, entre muchos otros. En tal sentido, la implementación de sistemas inteligentes se encuentra subordinada al respeto de la dignidad humana y de los derechos fundamentales, aspectos susceptibles de ser reconocidos por las instituciones, pero también como parte de las dinámicas de innovación (Terrones, 2020).

Se considera entonces que los problemas de ética y responsabilidad tienen su origen en los contextos sociales (Coeckelbergh, 2021), dado que, en sentido estricto, es el único dominio donde un conflicto se presenta como tal. Esto permitirá observar dichos problemas como relaciones entre valores y sistemas sociotécnicos y, de esa forma, reconocer que hacen parte de marcos de diseño y evaluación de aplicaciones. En cuanto que la IA se considera imbricada en un sistema sociotécnico, la admisión de la ética y la responsabilidad en su desarrollo también “implica cambios en el sistema sociotécnico en el que está integrada” (Buijsman, Klenk & van den Hoven, 2025, p. 67). Por lo tanto, es necesario examinar la relación recursiva entre los dispositivos y sus contextos. La consideración sobre la ética de la IA se amplía, entonces, para abarcar la caracterización ética de sus contextos de diseño e implementación.

Los temas de la ética y la responsabilidad no escapan al interés de gobiernos y organizaciones multilaterales. Por ejemplo, el *Global Index on Responsible AI* (Adams *et al.*, 2024), considera 19 áreas temáticas agrupadas en tres dimensiones: derechos humanos e IA, gobernanza responsable de IA y capacidades de IA responsable. Cada área temática abarca tres pilares diferentes del ecosistema de IA: marcos gubernamentales, acciones gubernamentales e iniciativas de actores no estatales. Entre sus principales conclusiones, señala que las estrategias nacionales carecen de principios éticos y no son vinculantes; que las medidas para proteger los derechos humanos en el contexto de IA aún son limitadas; y que existen aspectos preocupantes en torno a temas como la igualdad de género, la inclusión y la equidad. También señala que existen brechas en materia de seguridad y confiabilidad de la IA, así como en protección de los trabajadores. Finalmente, recomienda incrementar la participación de las universidades y la sociedad civil, así como la cooperación internacional, en el cumplimiento de estándares de IA responsable.

El *Artificial Intelligence Index Report 2024* (AI Index Steering Committee, 2024) identifica varias categorías relevantes en materia de inteligencia artificial responsable:

privacidad, gestión de datos, transparencia, capacidad de explicación, justicia, seguridad y protección. Este informe ofrece información técnica sobre numerosos incidentes en materia de responsabilidad de IA. Por ejemplo, señala que desde 2013 los incidentes se han multiplicado por 20 (aunque pueden estar subreportados), y que existen muchos aspectos a mejorar en esta materia, como el manejo de datos privados de los usuarios (que en algunos casos pueden ser extraídos de las aplicaciones de IA generativa) y la utilización de material protegido con derechos de autor en el entrenamiento de IA. Los dos reportes mencionados son avalados por diferentes entidades multilaterales, gobiernos, organizaciones no gubernamentales y empresas, lo que permite afirmar que existe un nivel de conciencia significativo sobre los retos de la ética y la responsabilidad de la IA.

Ahora bien, la reflexión acerca de las relaciones entre la ética y tecnologías como la IA posee varias aristas. Puede tratarse sobre la ética de las superinteligencias al estilo del poshumanismo; del marco ético que adoptarán los seres humanos para interactuar con sistemas inteligentes o de la programación de máquinas que actúen de acuerdo con principios éticos programados (Cortina, 2024). Tales enfoques pueden tener afinidades y diferencias. Además, existen obstáculos comunes, como por ejemplo la dificultad de instrumentar los principios éticos en forma de criterios de decisión en los sistemas de información (Canca, 2020). Por lo tanto, la ética puede hacerse presente en los sistemas inteligentes en forma de los esquemas de valores de sus desarrolladores, en las características funcionales de los programas y en los efectos de su implementación en diferentes escenarios.

199

Se han vislumbrado tres enfoques disponibles para esta discusión, y por tanto susceptibles de ser tomados en cuenta en la conformación de sistemas sociotécnicos y el diseño de sistemas inteligentes (Cortina, 2025; Misselhorn, 2022):

1. *Arriba a abajo*: con insumos de la tradición deontológica (énfasis en las normas) y utilitarista (énfasis en las consecuencias). Se enfoca en la formulación e inserción de principios morales en forma de capacidades funcionales de las máquinas.
2. *Abajo a arriba*: se propone establecer marcos interpretativos basados en los contextos y en los sujetos, como la ética de la virtud. Se basa en una ética orientada por valores relevantes para cada situación, por lo que se propone encontrar patrones y adaptarlos al aprendizaje de las máquinas.
3. *Híbrido*: como combinación entre ambos enfoques, se parte de un marco de principios que podrían ser adaptados inductivamente a contextos específicos. Se trataría de transferir valores éticos definidos colaborativamente a las características funcionales de los sistemas.

En este contexto, resulta necesario establecer referentes éticos que permitan que los seres humanos actúen de forma creativa y consensuada de acuerdo con determinadas reglas, para que, dado el caso, las mismas se transfieran al diseño y programación

de la IA. Esta es la importancia del enfoque híbrido, el cual “opera sobre la base de un marco predefinido de valores morales que luego se adapta a contextos específicos mediante procesos de aprendizaje” (Misselhorn, 2022, p. 41). Por ejemplo, se necesitaría establecer, a través de un proceso fundado en una ética comunicativa, un conjunto de normas válidas en cada tipo de actividad (basadas, por ejemplo, en su bien interno) para entonces derivar reglas operativas y programarlas en los sistemas inteligentes. Tal sería la premisa de un enfoque dialógico basado en la hermenéutica crítica (Cortina, 2024), que puede favorecer la participación cívica en el desarrollo e implementación de IA.

Uno de los referentes de este enfoque es la ética de la virtud, cuya comprensión se puede hacer con referencia a obras de autores como MacIntyre (2013). Este autor plantea que las virtudes son importantes para la realización de prácticas sociales, lo que implica comprender la naturaleza de sus bienes, identificar las virtudes más importantes y realizar acciones pertinentes para alcanzarlas. El enfoque puede sintetizarse en el concepto de “práctica virtuosa”, comprendida como una forma de actividad humana establecida socialmente, que permite desarrollar sus bienes inherentes al mismo tiempo que alcanzar mayores niveles de perfección. Por lo tanto, una práctica es virtuosa si se plantea la búsqueda de un bien interno y tiene en miras el cultivo de la excelencia. Por ejemplo, la producción de *software*, como ejercicio instrumental, no es necesariamente virtuosa en sí misma, pero el desarrollo responsable requiere la incorporación de virtudes que conduzcan a la obtención de mejores resultados, así como a una mayor comprensión de su carácter como práctica virtuosa. Esta es una manera de concebirlo como una actividad orientada a fines éticos, y por lo tanto a la generación de bienes sociales. En otras palabras, la búsqueda virtuosa del bien interno de una práctica debe facilitar el logro de sus bienes externos. Una práctica de desarrollo ética debería ofrecer respaldo a sistemas sociotécnicos más éticos.

Si se toma como referente la teoría de las prácticas virtuosas, sería necesario que los agentes sociales definan los bienes internos de las diferentes prácticas, para que los especialistas logren “incorporar en las programaciones las normas que deben seguir las máquinas para ayudar a alcanzar el bien interno” (Cortina, 2025, p. 120). En el primer momento reside el carácter dialógico de la propuesta. La deliberación en torno a valores clave puede contribuir a establecer un marco de referencia en la esfera pública, que brinde el contexto de interpretación a entes reguladores, empresas y usuarios para orientar el desarrollo e implementación de IA. El segundo momento, la codificación de decisiones éticas, no tiene carácter meramente instrumental, sino que, basados en el perfil constructivo de los sistemas sociotécnicos, se puede aspirar a consolidar una idea de tecnología que implique una forma distinta de plantearse los medios, fines, recursos y agentes que toman parte en el desarrollo tecnológico. Ambos momentos pueden contribuir a fomentar la democratización de los procesos de innovación desde una perspectiva cívica.

Puesto que la deliberación en torno a valores éticos debe tener como resultado efectivo el surgimiento de una modalidad de desarrollo responsable, el tema del diseño

orientado a valores juega un papel fundamental (Van den Hoven, Vermaas & Van de Poel, 2015). Desde la perspectiva que se ha venido discutiendo, se podría “diseñar un marco, contando con teorías éticas acreditadas y, dentro de él, articular las normas y las reglas adecuadas para el razonamiento, la decisión y la actuación, obtenidas a partir de las particularidades y las experiencias concretas de los afectados en ese ámbito” (Cortina, 2025, p. 106). En este sentido, la conjunción de los diferentes enfoques éticos en un proceso de desarrollo debe tener como resultado que “los valores de las partes interesadas de una tecnología (de abajo hacia arriba) se sopesan con los valores derivados de marcos teóricos y normativos (de arriba hacia abajo)” (Buijsman, Klenk & van den Hoven, 2025, p. 73). En otras palabras, se trata de identificar los valores éticos deseados para convertirlos en insumo de los procesos de desarrollo y evaluación de los sistemas inteligentes, pero en un sentido constructivo y crítico, no solo instrumental. Esto supone, por ejemplo, una ética de la virtud que permite la incorporación de los marginados de los procesos de construcción de consenso, basados convencionalmente en el estado de las relaciones de poder.

Este aspecto es congruente con un enfoque de coconstrucción de sistemas sociotécnicos. La teoría crítica de la tecnología explica que ésta debe ser analizada a través de los niveles de instrumentalización primaria y secundaria: “El nivel primario simplifica los objetos para su incorporación en un mecanismo, mientras que el secundario integra los objetos simplificados en un entorno natural y social” (Giuliano, 2013, p. 66). No obstante, en este proceso los valores dominantes son agregados al diseño, de manera que se entretajan con los sistemas en forma de “códigos técnicos” y obtienen aceptación social como parte de sus propiedades. Esto significa que el diseño tecnológico no es neutral, y que se pueden encontrar distintas trayectorias de cambio tecnológico de acuerdo con la pugna entre los intereses y valores de diferentes grupos sociales. Precisamente, la idea de la doble instrumentalización de la tecnología abre espacio a la tesis de la racionalización democrática (Feenberg, 2005), que implica que una mayor participación cívica en los procesos de deliberación debe incidir en el desarrollo de tecnologías más responsables.

201

La combinación de un enfoque crítico de la tecnología y el reconocimiento de la importancia del diseño cívico (Feng & Feenberg, 2008) contribuye a fundamentar las ideas de que es posible incorporar intencionalmente determinados valores éticos en los sistemas, en forma de códigos técnicos y características funcionales de los mismos, de manera que tengan representatividad efectiva los valores y las normas previamente acordadas. Esto contribuiría a fomentar aspectos como “el diseño ético de los procedimientos de toma de decisiones, como una oportunidad para crear sistemas (sociotécnicos) aún más responsables” (Buijsman, Klenk & van den Hoven, 2025, p. 77), y por tanto aplicaciones con mejor desempeño en materia de responsabilidad. Así, parece pertinente concebir el diseño y desarrollo de IA como un proceso de innovación colaborativa (Terrones, 2020), más cercano a las dinámicas de la ciencia y tecnología abiertas (Fecher & Friesike, 2014) que a las propias de la ciencia y tecnología mercantil.

La deliberación pública en torno a problemas de ética y responsabilidad es necesaria para sentar las bases de la selección de valores y criterios de decisión más deseables, pero asimismo es importante la construcción social de dinámicas que permitan su integración en sistemas sociotécnicos que ya encuentran presiones significativas en áreas como políticas de regulación y productividad. En este sentido, resulta pertinente revisar algunos discursos acerca de la ética y la responsabilidad en el desarrollo e implementación de IA, con el fin de adentrarse en el marco del debate sobre este tema.

2. Dos discursos sobre la ética de la IA

En el campo del desarrollo tecnológico intervienen muchos factores que resulta necesario ordenar. En este artículo se plantea que la inversión de recursos destinados a la producción de bienes culturales-tecnológicos puede comprenderse desde dos marcos de interpretación: el interés público y el interés privado (y sus correlatos en una ciencia y tecnología de interés cívico y de interés mercantil, respectivamente). Si se reconoce la presencia de ambos marcos de interpretación como conjuntos de valores que orientan los sistemas de acción, entonces se debe admitir que la tensión entre el interés privado y el interés público coloca a la ética en un escenario de conflicto. En tal sentido, se plantea que el desarrollo de IA responsable se encuentra en un dilema: si los valores que promueven la innovación persiguen un fin mercantil, entonces los bienes tecnológicos no pueden tener carácter ético, pero –si se acepta esto– entonces el desarrollo de tecnología no es una actividad que merezca reconocimiento social. En otras palabras, el desarrollo técnico orientado al mercado es un ejercicio instrumental pero no una práctica virtuosa. Debatir acerca de la ética del *software* depende de que se visibilicen los alcances y limitaciones de cada contexto interpretativo.

Un contexto interpretativo es un marco conceptual que le brinda sentido a una situación, y que influye en la manera en que se entienden aspectos como la información y las decisiones (Fuenmayor, 1991). La importancia de los contextos interpretativos reside en que no pretenden ser neutrales, sino que intentan retratar de forma sistemática los valores, intereses y estilos organizacionales que se entremezclan en el campo sociocultural. En ese sentido, también tienen importancia como parte de la lógica de la acción de individuos, grupos e instituciones. Los contextos interpretativos aludidos en este trabajo pueden describirse sucintamente de la siguiente manera:

- Como bien privado, la tecnología es concebida como un activo económico sujeto a las dinámicas de mercado. Esto significa que los derechos de propiedad intelectual (autoría y explotación económica) son exclusivos de sus propietarios, y que los bienes se encuentran sujetos a transacciones comerciales. Los agentes económicos deben maximizar sus beneficios y compiten por dominar el mercado, lo que convierte a la innovación en una

ventaja competitiva. Asimismo, se tiende a minimizar los costos y externalizar las consecuencias negativas de la actividad económica, lo que crea fricciones con entes de regulación y organizaciones civiles. En este contexto, la tecnología se concibe como un bien mercantil.

- Como bien público, la tecnología se considera como un recurso compartido cuyos beneficios deben estar garantizados a toda la sociedad. Los derechos de propiedad intelectual conceden acceso a los bienes cognitivos para favorecer la inclusión. El desarrollo tecnológico está orientado al interés común y se plantea contribuir con el bienestar humano. Asimismo, se fomenta la colaboración y el compartir con miras a facilitar experiencias de apropiación del conocimiento y de innovación. Las políticas tienen como referente la participación ciudadana y la deliberación, lo que garantiza un papel a las instituciones públicas. El concepto de ciencia abierta como bien común recoge algunos aspectos del carácter de la tecnología en este contexto.

En situaciones reales, no se encuentran los casos “puros” que representan los contextos interpretativos. En las relaciones sociales, se despliegan dinámicas de poder que se entretajan con la coconstrucción de sistemas sociotécnicos que implementan tecnologías como la IA desde diferentes aristas (Roca, 2025). Desde una perspectiva convencional, el interés privado tiene el potencial de propiciar la innovación, pero contribuye a crear nuevos problemas sociales; mientras que el interés público favorece la inclusión, pero encuentra dificultades para gestionar eficientemente la inventiva. Los contextos interpretativos que son mutuamente excluyentes permiten delimitar críticamente las maneras en las cuales se practica el desarrollo responsable de IA. La definición de los contextos interpretativos mencionados tiene influencia directa en la selección de los casos de estudio que se ofrecen en este trabajo, en los cuales se intenta exhibir las características de los discursos sobre la ética en el desarrollo y la implementación de IA.

203

Como se explicó en el apartado anterior, una aproximación convencional a los temas de ética en la IA ha sido enunciar un conjunto de principios y valores que sirvan de orientación a las organizaciones relacionadas con su implementación. En tal sentido, se encuentran numerosas guías institucionales para el reconocimiento de la ética y la responsabilidad de la IA, que coinciden en aspectos como el respeto a los valores de transparencia, justicia y equidad, no maleficencia, responsabilidad y privacidad; pero que también difieren en materia de interpretación e implementación (Buijsman, Klenk & van den Hoven, 2025). En concordancia, el investigador puede proceder críticamente en el sentido de examinar los fundamentos y la pertinencia de los valores y principios enunciados en cada caso.

En el ámbito de las recomendaciones emanadas por entes multilaterales, se tomaron en cuenta las propuestas de la Organización de la Naciones Unidas para la Educación, la Ciencia y la Cultura (UNESCO) y de la Unión Europea (UE). Se puede asignar a tales propuestas un lugar en el contexto interpretativo del conocimiento y la

tecnología como bienes públicos, de manera que su estudio favorece el registro y su comprensión. Asimismo, la lectura de los respectivos documentos permite observar la manera en que sus discursos dialogan con los agentes económicos en la delimitación de los problemas éticos de la IA.

La UNESCO ofrece un marco de referencia para pensar en la ética del desarrollo de IA. Por una parte, comprende la ética como “una reflexión normativa sistemática, basada en un marco integral, global, multicultural y evolutivo de valores, principios y acciones interdependientes” orientada a fomentar “la dignidad humana, el bienestar y la prevención de daños” (2022, p. 9). En cuanto a los sistemas inteligentes, incluye entre las preocupaciones éticas las relacionadas con los ciclos de vida, abarcando aspectos como financiación, investigación, diseño, desarrollo, comercialización, implementación y desmontaje. Asimismo, comprende la IA en un contexto de relaciones entre agentes que contribuyen con su utilización: “los actores de la IA pueden definirse como todo actor que participe en al menos una etapa del ciclo de vida del sistema de IA y pueden ser tanto personas físicas como jurídicas” (2022, p. 9). En tal sentido, la UNESCO comprende los sistemas técnicos como fenómenos arraigados en un contexto social.

204

La UNESCO (2023) plantea la importancia de fortalecer un enfoque orientado al reconocimiento de los derechos humanos. Los valores y principios desempeñan un papel importante en la recomendación, como orientación para las acciones políticas y jurídicas de los agentes de IA (UNESCO, 2022). Entre los valores que menciona se encuentran: respeto, protección y promoción de los derechos humanos, las libertades fundamentales y la dignidad humana; prosperidad del medioambiente y los ecosistemas; garantía de diversidad e inclusión; vivir en sociedades pacíficas, justas e interconectadas. Asimismo, entre los principios aplicables a la IA se mencionan: proporcionalidad e inocuidad; seguridad y protección; equidad y no discriminación; sostenibilidad; derecho a la intimidad y protección de datos; supervisión y decisión humanas; transparencia y explicabilidad; responsabilidad y rendición de cuentas; sensibilización y educación; gobernanza y colaboración de múltiples partes interesadas.

La propuesta de la Comisión Europea (2019) sobre la implementación de IA confiable también es ampliamente reconocida en diversas fuentes. En esta propuesta, la fiabilidad de la IA se apoya en tres componentes: debe ser lícita (respeto a las leyes), ética (congruente con valores éticos) y robusta (funcional y socialmente estable). Algunos principios éticos que se consideran implicados en el desarrollo e implementación de IA incluyen: respeto a la autonomía humana, prevención del daño, equidad y explicabilidad; atención a las relaciones entre grupos sociales y especialmente a los más vulnerables; y consideración de los posibles riesgos y efectos negativos. A partir de estos aspectos, el documento plantea algunas medidas para avanzar hacia una IA fiable, entre las cuales aparece como principal:

“Garantizar que el desarrollo, despliegue y utilización de los sistemas de IA cumpla los requisitos para una IA fiable: 1) acción y supervisión

humanas, 2) solidez técnica y seguridad, 3) gestión de la privacidad y de los datos, 4) transparencia, 5) diversidad, no discriminación y equidad, 6) bienestar ambiental y social, y 7) rendición de cuentas” (Comisión Europea, 2019, p. 2).

En este sentido, la propuesta plantea que se deben impulsar medidas como la investigación abierta y la formación de especialistas, informar sobre las capacidades y las limitaciones de los sistemas inteligentes, facilitar la auditabilidad de los sistemas, trabajar sobre todos los agentes y etapas del ciclo de vida de los sistemas, y resolver las tensiones entre los diferentes principios e intereses (Comisión Europea, 2019). En esta propuesta se da un paso más allá de los principios y valores, y se plantean acciones e indicadores para avanzar en la dirección de entornos de implementación de IA más responsables.

La aproximación de entes multilaterales como la UNESCO y la UE retrata a los agentes de IA como parte de ecosistemas sociotécnicos y no como innovadores aislados, dado que se reconoce la presencia de diversos agentes humanos, intereses y relaciones en contextos reales de implementación. Asimismo, se establecen diferentes valores y principios para orientar las actividades de los agentes que forman parte de los sistemas sociotécnicos, pero especialmente de los productores de la tecnología. En cuanto a su aproximación ética, los entes utilizan una orientación basada en la formulación de normas generales, las cuales se suponen aplicables a todos los casos. No obstante, algunas de las recomendaciones específicas de la UE están más orientadas a las características técnicas de los sistemas, así como al conjunto de relaciones que se entretienen en los contextos de implementación.

205

En el ámbito de los discursos de los entes privados, se tomaron como referencia las posiciones de Google y Microsoft. Los documentos examinados se incluyen en el contexto interpretativo del conocimiento y la tecnología como bienes privados, de manera que refieren la posición de los agentes económicos frente a las cuestiones éticas que plantean otros organismos. Asimismo, se puede observar cómo aquellos dialogan con otros actores para ofrecer un sentido ético a la implementación de sistemas inteligentes en contextos de mercado.

Google (2024) menciona entre los objetivos del uso de la IA: ser beneficiosa a nivel social; evitar crear o reforzar un sesgo parcial; estar creada y probada para ofrecer seguridad; ser responsable frente a las personas; incorporar principios de diseño de privacidad; mantener estándares altos de excelencia científica; y estar disponible para usos que concuerden con estos principios. Asimismo, se incluye un apartado sobre los usos que no se darán a su IA: tecnologías que puedan causar daño general; armas o cualquier otra tecnología cuyo propósito sea causar o permitir daños a las personas; tecnologías que recopilen o usen información para vigilancia y que infrinjan las normas aceptadas a nivel internacional; tecnologías cuyo propósito desobedezca los principios ampliamente aceptados de leyes internacionales y derechos humanos.

En este caso, la empresa plantea, de forma amplia, los usos que se pueden y no se pueden dar a sus productos de innovación.

Por otra parte, Microsoft (2025) plantea los siguientes principios de desarrollo y uso de IA: equidad, fiabilidad y protección; privacidad y seguridad; inclusión, transparencia y responsabilidad de las personas. En otro documento, la empresa plantea objetivos de desempeño en áreas tales como rendición de cuentas, transparencia, equidad, confiabilidad, privacidad y seguridad, e inclusión (Microsoft, 2022). En ese sentido, la empresa establece un conjunto de principios rectores, pero también incorpora la preocupación por la ética entre los estándares de evaluación de sus productos. Por lo tanto, comunica su interés en convertir los principios en características funcionales de los sistemas.

Los enfoques de empresas como Google y Microsoft identifican la posición pública de las mismas en cuanto que agentes económicos, y responden directamente a preocupaciones éticas sobre el uso de la IA en entornos de implementación. Desde su perspectiva, también se reconoce la IA como parte de un conjunto de relaciones sociales, particularmente en su dimensión institucional. En cuanto a la dirección ética, se puede afirmar que los principios de Google y de Microsoft tienen una expresión orientada a las normas, aunque los objetivos de desempeño del último reflejan un enfoque en las características y los efectos de los sistemas tecnológicos.

206

En síntesis, en el discurso de los entes multilaterales se incorporan normas que pueden ser relevantes más allá de la implementación de IA, como el respeto a los derechos humanos; o bien, como en el caso de la UE, se hace referencia a principios y medidas de acción para la implementación de sistemas tecnológicos, como la supervisión humana y la rendición de cuentas. El discurso de los entes comerciales también incluye normas generales, más orientadas hacia los sistemas de IA; y como en el caso de Microsoft, se agrega una preocupación por las características y los efectos de los productos que ofrecen. De esta manera, los discursos examinados reflejan los valores e intereses propios de cada organización y, por lo tanto, son coherentes con su perfil y su lugar en el ecosistema digital. En tal sentido, también permiten registrar la manera en que diferentes agentes sociales dan cuenta de su posición ante las preocupaciones que despierta la implementación de sistemas inteligentes en espacios reales, tomando en cuenta la responsabilidad que deben tener en el contexto del debate público.

Es válido cuestionar si todos los principios y valores enunciados son representativos de las preocupaciones éticas de la esfera pública contemporánea. Por ejemplo, algunos movimientos sociales podrían reclamar mayor visibilización de grupos vulnerables en las recomendaciones de la UNESCO (2021), que también se ha pronunciado en favor de una política de ciencia abierta, pero no la menciona en los documentos examinados. Si bien es cierto que los enunciados de una “sociedad justa” e “interconectada” pueden abarcar estos aspectos, no lo hacen explícitamente. En el caso de los entes privados, los valores que proponen son aplicables a todos sus productos. Sin embargo, es conocido que los sistemas tecnológicos no son neutrales, sino que son desarrollados

con intencionalidad y reproducen los sesgos que intervienen en su programación. Así, tomando por caso las críticas recurrentes al manejo de datos de los usuarios por parte de las empresas tecnológicas, debe entenderse que su compromiso con la IA abarca valores distintos a los que incorporan otros productos enmarcados en sus prácticas comerciales. Éstas son solo dos críticas que pueden dirigirse a los discursos basados en principios y valores de la ética en los sistemas de IA.

Un aspecto importante es la manera en que las recomendaciones éticas pueden adaptarse al diseño de *software* como actividad concreta. No parece suficiente que una organización haga declaraciones públicas afirmando su apoyo al desarrollo responsable, sino también debe haber un proceso de adaptación organizativa a los fines de implementar sistemas éticos. Por ejemplo, se pueden crear departamentos sobre la materia, indicadores de calidad de *software*, requerimientos y casos de uso específicamente orientados a cumplir ciertos criterios, consultas a usuarios, etc., medidas cuyos resultados pueden publicarse periódicamente en informes de rendición de cuentas de los agentes involucrados en la IA. Así, es necesario investigar cómo operacionalizar los principios éticos en forma de modelos organizacionales, criterios de decisión y proyectos de desarrollo. Por ejemplo, existen recomendaciones para la implementación responsable de IA en las organizaciones (Wade & Yokoi, 2024), así como propuestas para el desarrollo de IA ética desde el diseño (Gordaliza, 2021).

En síntesis, los discursos acerca de la ética en el desarrollo de sistemas inteligentes parecen reflejar el perfil de las organizaciones que los emiten, y en cierto modo reproducen las lógicas orientadas al bien público y al bien mercantil. Es necesario encontrar discursos compatibles entre sí, que contribuyan a integrar los ecosistemas de desarrollo e implementación a través de iniciativas como, por ejemplo, el diseño participativo (Van der Velden & Mörtberg, 2015), y de ese modo facilitar la incorporación de valores y principios consensuados entre las características funcionales de los sistemas. Esto significa que la ética debe considerarse más que una lista de exigencias o de indicadores de evaluación, y que se debe avanzar hacia la conformación de sistemas sociotécnicos donde la ética y la responsabilidad se consideren relevantes desde las etapas de diseño, de forma que se pueda evaluar su efectividad a lo largo de los ciclos de investigación, desarrollo y aplicación de los sistemas tecnológicos. Precisamente, el enfoque de “ética desde el diseño” (Comisión Europea, 2021; Brey & Dainow, 2023) apunta en la dirección de ofrecer metodologías que permitan evaluar el carácter ético de los sistemas desarrollados a partir de sus etapas iniciales, en lugar de esperar el momento de implementación para observar sus efectos.

207

3. Hacia una ética del diseño de sistemas inteligentes

El desarrollo de IA incorpora la ética desde varias aristas. Aspectos como la limitada eficiencia de la regulación, la opacidad de las empresas en el manejo de los productos informáticos y el impacto de la implementación de IA generan preocupación en la esfera pública (AI Index Steering Committee, 2024). La ética de las tecnologías surge

como una preocupación compartida, en torno a la cual se pueden tejer acuerdos que favorezcan a todas las partes interesadas. Sin embargo, no hay más que mirar el devenir del desarrollo tecnológico para caer en cuenta de la prevalencia que tienen las prácticas de mercado en la adopción social de tecnologías digitales. El caso de la IA generativa resulta elocuente en este sentido: la puja entre diferentes desarrolladores de modelos de lenguaje retrata la competencia que existe entre las empresas por convertirse en los proveedores de facto de una tecnología que está cambiando la manera en que se utilizan las interfaces digitales. Esta constante en los procesos de incorporación de innovaciones tecnológicas, sin la revisión crítica de todas las variables, ha hecho que se hayan dado lugar a daños irreversibles a terceros, como en el caso de la degradación ambiental.

En lo que respecta a la introducción de valores éticos en el desarrollo tecnológico, se ha considerado la importancia de “descubrir orientaciones éticas interculturales para utilizar con bien los medios tecnocientíficos, pero también diseñar guías éticas para incorporarlas en los sistemas inteligentes de forma que actúen éticamente” (Cortina, 2024, p. 29). En ese sentido, puede plantearse que las lecciones en materia de normas universales y beneficios de los bienes tecnológicos contribuyan a fomentar una suerte de “práctica virtuosa” del desarrollo tecnológico, donde valores como la justicia y la responsabilidad sean reconocidos como parte de las virtudes de los desarrolladores y de los bienes internos de los productos de innovación. Así, podría transferirse el reconocimiento de valores compartidos a métodos de desarrollo que contribuyan a recrear un modo no instrumentalista de gestión tecnológica, con miras a transformar los procesos, las relaciones y los resultados de investigación. De esa manera, resulta necesario plantear cuáles son los valores, principios y términos de actuación que permitirán confirmar que un determinado producto es el resultado de un conjunto de acuerdos de contenido ético.

En términos de bienes públicos y bienes privados (ciencia abierta y ciencia mercantil), esta situación dibuja un escenario de conflicto, donde se hace patente la ausencia de marcos de interpretación de la tecnología coherentes con las necesidades de sociedades pluralistas. Por una parte, se puede recurrir a la tradición ética para encontrar discursos en torno a las virtudes de los agentes tecnológicos, las normas universales o las consecuencias de la acción ética. Pero, asimismo, los discursos gubernamentales y los intereses de mercado se combinan y se contrarrestan en el crisol de las relaciones políticas y económicas. Además, las prácticas de negocio de las empresas logran convencer a los usuarios, quienes, por medio del consentimiento explícito, parecen aceptar los términos de uso de las aplicaciones a cambio de costos percibidos como males menores. El resultado es un escenario complejo, en el cual las tecnologías digitales se han convertido en dispositivos no neutrales con capacidad para impactar en las relaciones sociales y culturales sin vocación ética ni responsabilidad (Suárez, 2024).

En este contexto, puede criticarse que los problemas que generan el desarrollo e implementación de IA parecen menos dilemas éticos que contradicciones de la

planificación del mercado. En un escenario de negociación, la ética puede representar la posición de entes reguladores y agentes de mercado con miras a definir acuerdos que permitan articular la intervención de las tecnologías digitales en los procesos sociales, como suele ocurrir con el ascenso de nuevos paradigmas organizacionales (Pérez, 2020). En ese sentido, también constituye una alternativa para participar en la definición de políticas públicas y pautas de gobernanza de los ecosistemas digitales. Los entes multilaterales y las empresas que se tomaron como referencia hacen su parte al expresar posiciones acordes con los derechos fundamentales o con la recepción de sus bienes de mercado, respectivamente. En contraste, las diferencias pueden comenzar a aparecer en los niveles de gobernanza de los ecosistemas digitales y en las prácticas de diseño de las aplicaciones. Es decir, en los ecosistemas sociotécnicos de la digitalidad.

Tanto la UNESCO como la UE siguen una orientación normativa que se propone guiar políticas y acciones, basada en temas como los derechos humanos y la sostenibilidad. En consecuencia, se enfocan en la definición de principios éticos universales como equidad, privacidad, responsabilidad y no discriminación en el desarrollo de IA, y se encuentran en sintonía con marcos de principios establecidos previamente (Floridi, 2023). La UNESCO tiene una perspectiva global y multicultural, centrada en los derechos humanos, la diversidad y la sostenibilidad, mientras que la UE ofrece un enfoque regional que prioriza aspectos técnicos como la legalidad, ética y robustez de los sistemas técnicos, incluyendo sus condiciones de fiabilidad. Por otra parte, Google y Microsoft también comparten principios éticos como equidad, privacidad, transparencia y responsabilidad en el diseño de IA, así como su preocupación por la no maleficencia y la seguridad de los sistemas. Google prioriza ciertos beneficios sociales, como la excelencia científica, a la vez que restringe algunos usos de la IA que pueden perjudicar al ser humano, mientras que Microsoft enfatiza aspectos técnicos como la seguridad e incorpora estándares de medición de desempeño.

209

En este sentido, todos los agentes de IA comparten principios éticos como transparencia, equidad, privacidad y responsabilidad, además de que reconocen el impacto social de las tecnologías a través de riesgos como la discriminación y la seguridad. La UNESCO y la UE son entidades cuya razón de ser las obliga a preocuparse por el bien público, mientras que Google y Microsoft son entidades privadas que deben encontrar un balance entre sus opciones éticas, iniciativas de innovación y oportunidades de mercado. Por lo tanto, las propuestas de la UNESCO y la UE pretenden ser aplicables en diferentes contextos, mientras que las de Google y Microsoft están orientadas a la recepción de sus prácticas comerciales. En perspectiva, los marcos de la UNESCO y la UE adoptan una ética deontológica (basada en la idea de bienes universales), mientras que los marcos de Google y Microsoft combinan deontología con consecuencialismo (se proponen maximizar los beneficios y reducir los perjuicios). El problema de la virtud de los agentes tecnológicos queda implícito, y su lugar lo ocupan los atributos de los bienes tecnológicos que deberían desarrollarse con criterios éticos. En otras palabras, se presume que un desarrollador responsable es aquel que produce sistemas de IA bajo criterios de responsabilidad.

De este modo, ambos enfoques pueden considerarse tanto conflictivos como complementarios. Contemplados como parte de un cuadro más amplio, las diferentes perspectivas de ética en el desarrollo de IA abarcan aspectos comunes. Se parte de que la IA debe contribuir con el bienestar humano, evitar daños y fomentar valores compartidos. Además, se reconocen temas clave como el respeto a la dignidad humana; la equidad y la no discriminación; la transparencia, la responsabilidad y la sostenibilidad. Estos aspectos son representativos de una ética mínima, que puede traducirse en políticas, marcos jurídicos y guías para el diseño de productos técnicos en entornos complejos. En concordancia, se estaría integrando los aportes de los niveles normativos (global), regulatorios (regional) y operativos (privado) en la definición de bienes cognitivos, avanzando hacia la determinación de su concepción ética, su alcance jurídico, sus características técnicas y sus restricciones de adopción en el mercado. Sin embargo, en este escenario subyace una suerte de sesgo de selección: los diferentes discursos no son complementarios *per se*, sino como consecuencia de la manera en que fueron caracterizados para este estudio. En realidad, se trata de discursos en conflicto que solo pueden reconciliarse a través de iniciativas de gobernanza global.

210

Otro escenario puede basarse en lo que cabe esperar de la innovación como proceso de interés público. La ciencia y tecnología abiertas ofrecen un conjunto de técnicas de gestión de los bienes cognitivos que permiten que sean administrados colectivamente de acuerdo con ciertas normas (Hess & Ostrom, 2016). Por ejemplo, un repositorio de publicaciones de acceso abierto puede ser manejado por una fundación, y los artículos almacenados cuentan con licencias que permiten su reutilización. En el caso de la IA, se ha recomendado tomar en cuenta una definición de código abierto que incorpore los valores de “autonomía, transparencia, reutilización sin fricciones y mejora colaborativa” en el desarrollo de modelos inteligentes (Open Source Initiative, 2025), tal como se establece para otros productos de *software*. De este modo, los modelos de lenguaje, por ejemplo, desarrollados a través de la colaboración entre diferentes actores y entrenados con bases de datos de dominio público, pueden ofrecerse con licencias abiertas como una manera de retribuir a la sociedad. Así, la ciencia abierta resulta ser más que una práctica que persigue la excelencia de los bienes internos, y representa una contribución a la transformación de marcos institucionales, procesos técnicos y productos de conocimiento, ciencia y tecnología, en el mismo sentido en que una ética aplicada a la IA tiene su anclaje en un conjunto de principios éticos generales.

Se deduce entonces que, para superar la dicotomía entre bienes públicos y privados, resulta necesario democratizar los espacios de gestión de los bienes tecnológicos, considerando, entre otros aspectos, que sería una alternativa para someter a escrutinio los intereses, valores y principios éticos implícitos en los procesos de diseño, desarrollo, implementación y apropiación del conocimiento. En otras palabras, los aportes normativos, regulatorios y operativos en la definición de bienes cognitivos deben tener una base cívica, que puede ser cultivada en espacios de ciencia abierta y ciencia ciudadana, pero que también cuenta con todos los

mecanismos de formación de la opinión pública. La democratización de la gestión de bienes tecnológicos representaría una alternativa para el establecimiento de criterios de ética y responsabilidad en su diseño y liberación, no solo por gracia de aspectos como la auditabilidad de las iniciativas privadas y públicas, sino también por su efecto en la consolidación de una ética del diseño participativo que contribuya a modelar las prácticas de desarrollo de sistemas informáticos. En tal sentido, la racionalización democrática puede dar lugar no tanto a la implementación de otros sistemas técnicos, sino sobre todo a la institucionalización de otros sistemas éticos.

Superar la prevalencia de los discursos público y privado, y propiciar la adopción de un discurso comunitario basado en la ciencia abierta, requiere mantener una dialéctica entre la reflexión sobre los deberes (deontología) y los efectos de la adopción de los sistemas tecnológicos (consecuencialismo), pero también exige la incorporación de virtudes éticas en el carácter de los agentes tecnológicos. Por lo tanto, la deliberación acerca de la ética de los agentes y las narrativas que la sostienen no puede postergarse indefinidamente. En la medida en que los sistemas sociotécnicos son complejos, la posición ética debe estar comprometida a responder por aspectos como los conflictos de intereses, los efectos inesperados de las tecnologías y la realidad de los diferentes contextos de aplicación. Entre los agentes de IA, para utilizar la terminología de la UNESCO, se deben incluir participantes ocupados de las normas, las regulaciones, las prácticas de negocio y el deber cívico en contextos regionales y globales, con los fines de que la crítica a las relaciones mediadas por la tecnología sirva de fundamento para la aprobación de iniciativas acordes con valores consensuados como el respeto a la dignidad humana, así como con el diseño de modelos organizacionales y métodos de desarrollo congruentes con los mismos.

211

En tal sentido, las herramientas conceptuales y metodológicas de la ciencia y tecnología abiertas pueden contribuir a orientar el manejo de bienes cognitivos mercantiles en la dirección de los bienes comunes. Por ejemplo, puede reivindicarse el carácter ético de las prácticas de desarrollo de IA (ética de la virtud) y crear acuerdos que permitan la incorporación de valores responsables en procesos de diseño tecnológico orientados a los usuarios, definidos a su vez como sujetos éticos (racionalización democrática). Asimismo, se pueden crear prácticas de participación cívica que ofrezcan espacio a la democratización de los procesos de implementación de tecnologías, salvaguardando la perspectiva ética frente a la aceptación del mal menor o las imposiciones de la autoridad. En tal sentido, el desarrollo ético y responsable de sistemas inteligentes pasa necesariamente por la cocreación de sistemas sociotécnicos donde tengan cabida la ética y la responsabilidad como parte del diseño y la implementación de tecnologías.

Conclusiones

Es claro que las falencias éticas de la IA no son imputables a la propia IA, al menos como producto terminado. El diseño y desarrollo de programas informáticos es un

proceso reflexivo y orientado subjetivamente donde toman parte diferentes intereses. La arquitectura y los términos de uso de las aplicaciones responden a esquemas de diseño que favorecen el modelo de negocio de las empresas. Así, pueden introducir especificaciones funcionales que no responden solamente a necesidades de eficacia y eficiencia. Además, una vez que los usuarios comienzan a utilizar una aplicación, los desarrolladores obtienen más información sobre sus gustos y pueden realizar cambios en el programa. En el caso de la IA generativa, se agrega el hecho de que los sistemas inteligentes son entrenados con textos abiertos, por lo que tienden a incorporar sesgos humanos, como por ejemplo con respecto a prejuicios sociales. La cuestión de los valores de los desarrolladores de IA confronta solo una parte del problema. Como ecosistemas sociotécnicos, los entornos de innovación deben fundamentarse en principios éticos que contribuyan con la deliberación sobre los verdaderos intereses y valores implícitos en el desarrollo de bienes tecnológicos.

212

Como señala una perspectiva crítica, los valores e intereses implícitos en las aplicaciones digitales tienden a normalizarse en contextos de uso. Los métodos de diseño convencionales se combinan con requerimientos de desarrollo derivados de los intereses de los agentes de mercado en el marco de la lógica propia del contexto interpretativo de la tecnología mercantil. Sin embargo, los mismos procesos de diseño, desarrollo e implementación pueden utilizarse para liberar programas donde tengan cabida valores éticos como la responsabilidad, intereses cívicos como la representación de grupos vulnerables y adaptaciones que respondan al bien común antes que al monopolio del mercado. La racionalización democrática de la tecnología digital es posible si los intereses cívicos ganan lugar en los espacios de debate sobre el diseño tecnológico, y esto se traduce en la adopción de esquemas de diseño participativo. Por ejemplo, prácticas de la ciencia y tecnología abierta, como los laboratorios ciudadanos, la documentación abierta y el *software* de código abierto pueden convertirse en mecanismos para reunir intereses comunitarios alrededor del diseño de *software* en entornos éticos. Pero, asimismo, la democratización implica otros aspectos como el escrutinio público de las iniciativas de innovación y la participación cívica en los procesos de regulación pública.

La deliberación abierta sobre los principios éticos que deben fundamentar las actividades de desarrollo tendría el papel de favorecer el consenso alrededor de los valores, intereses, prácticas y relaciones acordes con los mismos. La ética, distinta de la opinión general y de las decisiones de la autoridad, contribuye a establecer criterios de evaluación de la acción y ayuda a fundamentar acuerdos sobre los cuales pueden hacerse reclamos de responsabilidad. La introducción de la ética en el diseño debe responder qué valores y principios serán transferidos en forma de características funcionales a las aplicaciones tecnológicas, sean sistemas inteligentes o agentes autónomos. Pero, asimismo, implica un marco de conducta para los agentes de la IA, quienes tienen oportunidad de comprometerse en la coconstrucción de sistemas sociotécnicos más éticos y responsables. En ese sentido, el diseño ético apunta a fomentar un conjunto de prácticas virtuosas ocupadas de los bienes internos y la excelencia en el desarrollo de aplicaciones, pero puede contribuir a reforzar sistemas

sociotécnicos donde tengan lugar relaciones de responsabilidad mutua y apoyar un conjunto de valores que tengan arraigo más allá de los contextos restringidos de la ética aplicada, con el fin de propiciar otros modos de comprender la tecnología.

Es claro que este horizonte no implica una trayectoria lineal, sino que se requiere del diálogo entre todas las partes. Los valores éticos que pudieran proponer los académicos deben contrastarse con los valores culturales en cada contexto social. Las diferencias entre agentes que responden a la lógica del bien público y a la lógica mercantil pueden revertirse desde la base de la complementariedad que cabe en los espacios participativos de la ciencia abierta. Existen numerosas carencias desde el punto de vista material, como las dificultades para reunir capacidades de financiamiento y escalamiento para la innovación fuera de los contextos de mercado. Sin embargo, al delimitar una situación ideal es posible contribuir con la conformación de un marco de diseño que facilite que el problema de la ética en la IA sea una cuestión de base, y no solo un pretexto para establecer marcos de negociación con las empresas una vez que se perciben los efectos indeseados de la tecnología en la sociedad. El dilema de la ética en la IA puede resolverse *ab initio* desde la perspectiva de un diseño responsable y abierto a la participación cívica.

Bibliografía

213

Adams, R., Adeleke, F., Florido, A., de Magalhães Santos, L. G., Grossman, N., Junck, L. & Stone, K. (2024). Global Index on Responsible AI 2024. Recuperado de: https://agatadata.com/wp-content/uploads/2024/11/Global-Index-on-Responsible-AI-2024-Corrected-Edition-25_10_24-Gobierno-de-Canada-y-USAID.pdf.

AI Index Steering Committee (2024). The AI Index 2024 Annual Report. Stanford: Stanford University. Recuperado de: <https://aiindex.stanford.edu/report/>.

Askill, A., Brundage, M. & Hadfield, G. (2019). The Role of Cooperation in Responsible AI Development. Recuperado de: <https://arxiv.org/pdf/1907.04534>.

Brey, P. & Dainow, B. (2024). Ethics by design for artificial intelligence. *AI Ethics*, 4, 1265–1277. DOI: <https://doi.org/10.1007/s43681-023-00330-4>.

Buijsman, S., Klenk, M. & van den Hoven, J. (2025). Ethics of AI. Toward a “Design for Values” Approach. En N. Smuha (Ed.), *The Cambridge Handbook of the Law, Ethics and Policy of Artificial Intelligence* (59-78). Cambridge: Cambridge University Press.

Canca, C. (2020). Operationalizing AI Ethics Principles. *Communications of the ACM*, 63(12). DOI: <https://doi.org/10.1145/3430368>.

Coeckelbergh, M. (2021). *Ética de la inteligencia artificial*. Madrid: Cátedra.

Comisión Europea (2019). Directrices éticas para una IA fiable. Recuperado de: <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>.

Cortina, A. (2024). ¿Ética o ideología de la inteligencia artificial? El eclipse de la razón comunicativa en una sociedad tecnologizada. Madrid: Paidós.

European Commission (2021). Ethics By Design and Ethics of Use Approaches for Artificial Intelligence. Recuperado de: https://ec.europa.eu/info/funding-tenders/opportunities/docs/2021-2027/horizon/guidance/ethics-by-design-and-ethics-of-use-approaches-for-artificial-intelligence_he_en.pdf.

Fecher, B. & Friesike, S. (2014). Open Science: One Term, Five Schools of Thought. Opening Science. The Evolving Guide on How the Internet is Changing Research, Collaboration and Scholarly Publishing. Springer.

Feenberg, A. (2005). Teoría crítica de la tecnología. Revista Iberoamericana de Ciencia, Tecnología y Sociedad, 2(5), 109-123. Recuperado de: <https://ojs.revistacts.net/index.php/CTS/article/view/1018>.

Feng, P. & Feenberg, A. (2008). Thinking about Design: Critical Theory of Technology and the Design Process. En P. Kroes, P. Vermaas, A. Light, S. Moore (Eds.), Philosophy and Design (105-118). Dordrecht: Springer. DOI: https://doi.org/10.1007/978-1-4020-6591-0_8.

Floridi, L. (2023). The Ethics of Artificial Intelligence. Principles, Challenges, and Opportunities. Oxford: Oxford University Press. DOI: <https://doi.org/10.1093/oso/9780198883098.001.0001>.

Fuenmayor, R. (1991). Truth and openness: An epistemology for interpretive systemology. Systems Practice, 4(5). Recuperado de: <http://www.saber.ula.ve/bitstream/handle/123456789/15918/fuenmayor-truth.pdf>.

Giuliano, H. (2013). La teoría crítica de la tecnología: una aproximación desde la ingeniería. Revista Iberoamericana de Ciencia, Tecnología y Sociedad, 8(24), 63-74. Recuperado de: <https://ojs.revistacts.net/index.php/CTS/article/view/626>.

Google (2024). Principios de la IA. Objetivos para crear IA beneficiosa. Recuperado de: <https://ai.google/static/documents/ES-419-AI-Principles.pdf>.

Gordaliza, E. (2021). Inteligencia Artificial Responsable desde el diseño. Recuperado de: https://www.bbva.com/wp-content/uploads/2021/06/InteligenciaArtificialResponsable_EstherDelaTorre_BBVA.pdf.

Habermas, J. (1999). Teoría de la acción comunicativa. I. Racionalización de la acción y racionalidad social. Madrid: Trotta.

Hess, C. & Ostrom, E. (2016). Los bienes comunes del conocimiento. Madrid: Traficantes de Sueños.

Lauwaert, L. & Oimann, A. (2025). Moral Responsibility and Autonomous Technologies. Does AI Face a Responsibility Gap? En N. Smuha (Ed.), *The Cambridge Handbook of the Law, Ethics and Policy of Artificial Intelligence* (101-116). Cambridge: Cambridge University Press.

MacIntyre, A. (2013). *Tras la virtud*. Madrid: Crítica.

Microsoft (2025). Promoción de prácticas de inteligencia artificial responsables. Recuperado de: <https://www.microsoft.com/es-es/ai/responsible-ai>.

Microsoft (2022). Microsoft Responsible AI Standard. Recuperado de: <https://blogs.microsoft.com/wp-content/uploads/prod/sites/5/2022/06/Microsoft-Responsible-AI-Standard-v2-General-Requirements-3.pdf>.

Misselhorn, C. (2022). Artificial Moral Agents: Conceptual Issues and Ethical Controversy. En S. Voeneky, P. Kellmeyer, O. Mueller & W. Burgard (Eds.), *The Cambridge Handbook of Responsible Artificial Intelligence. Interdisciplinary Perspectives* (31-49). Cambridge: Cambridge University Press.

OpenAI (S.F.) OpenAI Charter. Recuperado de: <https://openai.com/charter/>.

215

Open Source Initiative (2025). The Open Source AI Definition – 1.0. Recuperado de: <https://opensource.org/ai/open-source-ai-definition>.

Pérez, C. (2020). Revoluciones tecnológicas y paradigmas tecnoeconómicos. En D. Suárez, A. Erbes & M. F. Barletta (Eds.), *Teoría de la innovación, evolución, tendencias y desafíos: herramientas conceptuales para la enseñanza y el aprendizaje* (134-159). Madrid: Ediciones Complutense & UNGS.

Roca, S. (2025). Defectuoso por diseño: Relaciones de poder en la esfera digital. *Teknokultura. Revista de Cultura Digital y Movimientos Sociales*, 22(1), 69-76. DOI: <https://doi.org/10.5209/tekn.95708>.

Suárez, J. L. (2023). *La condición digital*. Madrid: Trotta.

Terrones, A. (2020). Inteligencia artificial, responsabilidad y compromiso cívico y democrático. *Revista Iberoamericana de Ciencia, Tecnología y Sociedad*, 44(15), 253-276. Recuperado de: <https://ojs.revistacts.net/index.php/CTS/article/view/166>.

UNESCO (2021). Proyecto de recomendación sobre la ciencia abierta. Recuperado de: https://unesdoc.unesco.org/ark:/48223/pf0000378841_spa.

UNESCO (2022). Recomendación sobre la ética de la inteligencia artificial. Recuperado de: https://unesdoc.unesco.org/ark:/48223/pf0000381137_spa.

UNESCO (2023). Key facts UNESCO's Recommendation on the Ethics of Artificial Intelligence. Recuperado de: <https://unesdoc.unesco.org/ark:/48223/pf0000385082>.

Van den Hoven, J., Vermaas, P. & Van de Poel, I. (Ed.) (2015). Handbook of Ethics, Values, and Technological Design. Sources, Theory, Values and Application Domains. Reino Unido: Springer. DOI: <https://doi.org/10.1007/978-94-007-6970-0>.

Van der Velden, M. & Mörtberg, Ch. (2015). Participatory Design and Design for Values. En J. van den Hoven, P. E. Vermaas e I. van de Poel (Eds.), Handbook of Ethics, Values, and Technological Design. Sources, Theory, Values and Application Domains (41-66). Dordrecht: Springer.

Wade, M. & Yokoi, T. (2024). How to Implement AI — Responsibly. Harvard Business Review, 10 de mayo. Recuperado de: <https://hbr.org/2024/05/how-to-implement-ai-responsibly>.