Revista de Filosofía
Volumen 81 (2024) 137-160
Artículo
Doi: 10.5354/0718-4360.2024.74905

¿CÓMO INTEGRAR LA ÉTICA APLICADA A LA INTELIGENCIA ARTIFICIAL EN EL CURRÍCULO?

Análisis y recomendaciones desde el feminismo de la ciencia y de datos

HOW CAN WE INTEGRATE APPLIED ETHICS INTO ARTIFICIAL INTELLIGENCE CURRICULA?

An analysis and recommendations from data and science feminism

Gabriela Arriagada Bruneau[1]
Instituto de Éticas Aplicadas
Instituto de Ingeniería Matemática y Computacional.
Pontificia Universidad Católica de Chile
gcarriagada@uc.cl
ORCID: 0000-0002-0006-7024

Javiera Arias[2]
Instituto de Éticas Aplicadas
Pontificia Universidad Católica de Chile.
jcarias@uc.cl
ORCID: 0009-0001-2413-3241

Recibido: 06-06-2024 • Aceptado: 24-09-2024

1 Profesora asistente, Instituto de Éticas Aplicadas e Instituto de Ingeniería Matemática y Computacional, Pontificia Universidad Católica de Chile. Investigadora joven del Centro Nacional de Inteligencia Artificial.

2 Investigadora asociada, Instituto de Éticas Aplicadas, Pontificia Universidad Católica de Chile.

Resumen

Este artículo examina la incorporación de la ética aplicada a la inteligencia artificial (IA) en los currículos universitarios chilenos, destacando la urgencia de implementar un marco de acción integrado. Mediante un análisis documental, se evidencia que la mayoría de los programas de educación superior no declaran cursos de ética en IA en sus currículos, lo que alerta la necesidad de sistematizar esta integración institucionalmente. En respuesta, proponemos un enfoque basado en el feminismo en la ciencia y el feminismo de datos, promoviendo la inclusión de diversas perspectivas y experiencias en la enseñanza de la ética aplicada. Este marco busca mejorar la integración de la ética en el currículo y también preparar a los estudiantes para resolver dilemas éticos en contextos sociotécnicos complejos, reforzando la necesidad del razonamiento ético aplicado en la formación en disciplinas asociadas a la IA.

Palabras clave: ética de IA, ética profesional, sociotecnología, epistemología feminista, diversidad.

Abstract

This article examines the incorporation of applied ethics into artificial intelligence (AI) within Chilean university curricula, emphasizing the urgent need to implement an integrated framework of action. Through a documentary analysis, it becomes evident that most higher education programs do not explicitly include AI ethics courses in their curricula, highlighting the need for institutionalizing this integration systematically. In response, we propose an approach grounded in feminist science and data feminism, advocating for the inclusion of diverse perspectives and experiences in the teaching of applied ethics. This framework aims not only to enhance the integration of ethics into the curriculum but also to prepare students to address ethical dilemmas in complex sociotechnical contexts, underscoring the importance of applied ethical reasoning in AI-related disciplines.

Keywords: AI ethics, professional ethics, sociotechnics, feminist epistemology, diversity.

1. La percepción de la ética como necesaria pero insuficiente

En el contexto del desarrollo y la aplicación de la inteligencia artificial (IA), la ética ha cobrado una importancia creciente, consolidándose en un pilar esencial para el avance de esta tecnología. La vasta literatura sobre ética en IA abarca una variedad de preocupaciones, que van desde la privacidad, la justicia y la responsabilidad hasta la legitimidad de los algoritmos utilizados en

sectores críticos como la salud y la educación. Hoy, la ética en IA es una disciplina suficientemente robusta para que se integre sistemáticamente en la formación de profesionales del campo.

Dada la presencia de la IA en nuestra vida cotidiana, es imprescindible que los futuros profesionales comprendan y aborden en profundidad las implicaciones éticas de su desarrollo e implementación. En este sentido, la inclusión de la ética en los currículos de educación superior en disciplinas relacionadas con la IA (informática, ciencia de datos, ingeniería y estadística, entre otras) se perfila como una estrategia fundamental para enfrentar estos desafíos. Esta formación debe enfocarse en desarrollar habilidades integradas, como el pensamiento crítico ético aplicado a herramientas y metodologías técnicas (Fiesler et al., 2020). Sin embargo, un desafío clave radica en que, aunque se consodera que la ética en la IA es relevante, muchos profesionales no tienen claro por qué lo es o cómo se relaciona con su trabajo.

Estudios recientes han revelado un patrón común entre profesionales y desarrolladores de IA: aunque reconocen la importancia de la ética en este campo, prevalece la percepción de que los principios éticos establecidos no se aplican directamente a su labor, sino que es más una preocupación general que una responsabilidad profesional personal. Tanto el estudio exploratorio hecho por el Centro Nacional de Inteligencia Artificial en Chile (López et al., 2023) como la investigación realizada por Griffins et al. (2023), que abarca desarrolladores de diferentes partes del mundo, destacan que gran parte de los profesionales estiman que la aplicación de estos principios éticos es ajena a sus responsabilidades prácticas diarias. Este fenómeno refleja una disociación entre el reconocimiento teórico de la ética como un pilar fundamental en el desarrollo de la IA y su integración efectiva en la práctica profesional.

Esta discrepancia es comprensible si consideramos las críticas hacia los marcos y principios éticos en IA. Debido a brechas estructurales e institucionales, estos marcos son a menudo las únicas guías disponibles para que los desarrolladores se informen sobre ética, aunque su eficacia ha sido cuestionada. Mittelstadt (2019) critica que muchas iniciativas, especialmente las patrocinadas por la industria, son meras señales éticas que retrasan la regulación y estancan el debate. De manera similar, Hagendorff (2020) y Munn (2023) advierten que estos principios carecen de mecanismos de refuerzo y fomentan una brecha entre teoría y práctica, manteniendo la situación actual y priorizando intereses económicos sobre la ética.

En este contexto, donde la ética de la IA es valorada, pero a menudo percibida como abstracta, resulta crucial integrar de manera efectiva la ética aplicada en los currículos de educación superior. Es fundamental que los desarrolladores de IA no solo reconozcan su importancia, sino que también adopten prácticas concretas basadas en principios éticos. Si bien la inclusión de la ética en el currículo es esencial, no

basta; una formación sólida debe equipar a los profesionales con un conocimiento ético situado y relevante, garantizando su aplicación en la práctica profesional.

En virtud de lo anterior, sostenemos que la integración de la ética aplicada a la IA en el currículo de educación superior representa un aporte significativo para avanzar en su validación. En este artículo, evidenciamos, mediante un breve análisis documental, la carencia de cursos de ética aplicada a la IA en los currículos universitarios en Chile. No obstante, es sabido que las modificaciones curriculares son un proceso intrínsecamente lento y complejo, debido a diversos factores como la revisión exhaustiva de contenidos, estrategias de enseñanza, objetivos de aprendizaje y la posterior aprobación por múltiples niveles de autoridad.

En este contexto, la adopción de un marco de acción integrada, entendido como un enfoque estructurado y coordinado que alinee expectativas curriculares con prácticas docentes, resulta esencial para acelerar y optimizar el proceso de modificación del plan de estudios. Por lo tanto, nuestra propuesta ofrece un marco de acción integrada basado en el feminismo de la ciencia y los datos, con el objetivo de elaborar una estrategia unificada que coordine tanto los objetivos como las metodologías, facilitando una implementación más efectiva de los cambios curriculares previstos a mediano plazo.

A partir de la comprensión de la IA como un sistema sociotécnico (van de Poel, 2020) y los conceptos de conocimientos situados (Haraway, 1988), de objetividad fuerte (Harding, 1995, 2015), la perspectiva socialmente fundada (Rolin, 2006) y el marco de feminismo de datos (D'Ignazio y Klein, 2020), argumentaremos que la integración de la ética aplicada en IA debe desafiar pretensiones de universalidad, condicionando los análisis éticos y la traducción de los principios éticos a contextos específicos, incluyendo las realidades culturales, sociales y políticas en los que se desarrollan. Así, mostraremos cómo cada concepto puede inspirar contenidos y estrategias metodológicas para mejorar la integración de la ética en IA en el currículo, lo que podría promover entre los estudiantes considerar que la ética es necesaria para desarrollar e implementar la IA.

2. La ética aplicada a la IA en el currículo chileno

La urgencia de la formación ética en los programas relacionados con la IA es evidente debido al impacto y el potencial de esta tecnología para exacerbar y amplificar las desigualdades sociales, como señalan Raji et al. (2021). Por ello, es esencial que quienes desarrollen e implementen tecnologías de IA no solo sean técnicamente competentes, sino que también versen profundamente en la capacidad de discernir éticamente. Para desarrollar esta capacidad, los estudiantes deben comprender las

consecuencias de largo alcance de su trabajo y demostrar un compromiso con las consideraciones éticas, asegurando así que la IA se utilice para mejorar el bienestar social (Gorur et al., 2020).

En este contexto, una estrategia clave para abordar las complejidades éticas de las tecnologías emergentes es ofrecer formación ética a quienes las desarrollarán. No obstante, no todas las formas de educación ética son igualmente efectivas. Enseñar códigos de ética profesional o teorías éticas normativas es necesario, pero insuficiente. La ética enfrenta desafíos prácticos que requieren una aplicación contextualizada, por lo que asumimos que la formación ética en programas de IA debe centrarse en una ética aplicada.

La ética aplicada se enfoca en los desafíos específicos de cada disciplina y desde métodos filosóficos aborda problemas morales en contextos profesionales concretos (Beauchamp, 2005). Además, como sugiere Dignum (2021), esta aplicación debe ser transdisciplinaria, formando estudiantes con competencias en diversos marcos intelectuales, más allá de una simple multidisciplinariedad. Goetze (2023), por ejemplo, enfatiza la importancia de los proyectos transdisciplinarios en la educación en ciencias de la computación, señalando que integrar la ética requiere marcos conceptuales y metodologías específicas que permitan a los estudiantes abordar temas complejos.

Sin embargo, a pesar de un reconocimiento generalizado de la importancia de la educación ética, la pregunta clave es cómo hacerlo. La formación en ética de la IA debería, idealmente, encontrar un equilibrio entre el contenido técnico y filosófico, para ello debe combinar elementos prácticos y teóricos que preparen a los estudiantes con el conocimiento fundamental necesario para navegar futuros avances y desafíos en el campo (Tuovinen y Rohunen, 2021).

La literatura identifica dos enfoques para integrar la ética en el currículo: ofrecer cursos dedicados a la ética de la IA o incluir discusiones éticas en cursos técnicos. El primer enfoque se centra exclusivamente en la ética, mientras que el segundo incorpora aspectos éticos en los objetivos técnicos (Garrett et al., 2020). Los cursos independientes de ética podrían ser limitados en su efectividad si solo ofrecen una única exposición a discusiones éticas a lo largo de toda la formación e incluir aspectos éticos en cursos técnicos podría llevar a los estudiantes a percibir la ética como un aspecto adicional o complementario. Ambos enfoques tienen limitaciones, lo que subraya la necesidad de una integración transversal y transdisciplinaria que combine y fortalezca sus aspectos positivos, superando así sus deficiencias.

En lo que sigue, examinaremos la escasa incorporación de cursos sobre ética en inteligencia artificial en los currículos de educación superior en Chile, lo que nos permite fundamentar nuestra propuesta para encaminarnos a ese enfoque combinado.

2.1 Evidencia inicial para el marco conceptual

Para complementar nuestra propuesta, se realizó un breve análisis documental, centrado exclusivamente en la revisión de las mallas curriculares declaradas en los sitios web de diversas universidades chilenas. Este enfoque, útil para una visión general, no es exhaustivo y no refleja la realidad educativa en estas instituciones. Por ejemplo, se debe considerar la existencia de prácticas de currículum oculto –un concepto ampliamente discutido en la literatura que refiere a la inclusión no oficializada de contenidos educativos, como la enseñanza de ética, a través de la iniciativa individual de los docentes (Bergenhenegouwen, 1987; Margolis, 2001)–. Por lo tanto, los hallazgos de este análisis se deben interpretar con cautela, reconociendo las limitaciones inherentes al no examinar en que estas prácticas suplen, en parte, la falta de una integración formal de la ética en los currículos.

Nuestro punto de partida, aunque modesto, evidencia claramente la falta de integración sistemática de la ética aplicada a la inteligencia artificial en los programas académicos chilenos, especialmente en las carreras STEM (Science, Technology, Engineering y Mathematics) vinculadas al uso o implementación de IA. Los porcentajes que presentaremos señalarán esta deficiencia y servirán de base para proponer un marco conceptual que integre más profunda y sistemáticamente la ética en los currículos de educación superior en Chile, mediante prácticas docentes y actualizaciones curriculares.

2.1.1 Análisis documental

Para complementar e informar nuestra propuesta, utilizamos una metodología de análisis documental, que consiste en la revisión e interpretación de documentos para extraer datos significativos y relevantes. Siguiendo la estructura establecida por Bowen (2009), identificamos y recopilamos documentos pertinentes de los sitios web oficiales de cada universidad, cuantificando solo la cantidad de cursos de ética en IA declarados en mallas curriculares a nivel de pregrado y magíster[3].

[3] Para el análisis documental, consideramos programas de estudios bajo un paraguas amplio de la IA, lo que significa que incluye programas de ingeniería computacional y matemática, ingeniería civil en informática, ciencia y tecnología de datos, ingeniería en desarrollo de videojuegos y realidad virtual, ingeniería en tecnologías de la información y otras ingenierías relacionadas con la mecatrónica o robótica, programas con un componente de IA o datos en su currículo o como base de formación.

- Para este breve análisis, se consideraron 37 universidades chilenas, incluyendo 47 programas de pregrado y 24 de magíster. No se incorporaron centros técnicos o institutos, ni programas de doctorado en el estudio. Además, algunas universidades fueron omitidas intencionalmente, ya que no incluían programas que cumplieran con los criterios del paraguas general de IA.
- La recopilación de datos se llevó a cabo mediante el análisis de planes de estudios oficiales y disponibles públicamente en los sitios web de las universidades seleccionadas. La recopilación de datos concluyó el 1 de noviembre de 2023, por lo que los planes de estudios pueden haberse modificado después de esa fecha.

A continuación, presentamos una tabla que resume el porcentaje de programas que incluyen cursos de ética de IA, considerando los 47 programas de pregrado y los 24 programas de magíster analizados.

Tabla 1

Frecuencia y porcentaje de los cursos de ética de IA declarados en el currículo de programas de pregrado y magíster

Nivel	Frecuencia	Porcentaje
Pregrado	8	17,02%
Magíster	4	16,67%

Fuente: elaboración propia, sobre la base de datos extraídos de las mallas de estudio de carreras STEM.

3. Un marco para integrar la ética aplicada de IA en el currículo

Para abordar estos bajos porcentajes de integración de la ética aplicada en los currículos de programas relacionados con IA, ofrecemos una propuesta de cómo creemos que esta integración puede enfocarse desde el feminismo en filosofía de la ciencia y el feminismo de datos, concibiendo la IA como un sistema sociotécnico. Consideramos que esta perspectiva puede promover un desarrollo transdisciplinario y contextualizado, favoreciendo el entendimiento de la ética de la IA como parte íntegra del quehacer profesional.

3.1 La IA como sistema sociotécnico

Reconocer la inteligencia artificial (IA) como un sistema sociotécnico es esencial para integrar la ética en su desarrollo y aplicación. Este enfoque adopta una perspectiva crítica de las estructuras sociales, científicas y técnicas que subyacen en su creación. La idea de la IA como un fenómeno sociotécnico se origina en visiones constructivistas de la tecnología, donde los sistemas sociotécnicos emergieron como respuesta a las dinámicas laborales de la industria minera británica en la posguerra (Trist, 1981). Este cambio de paradigma transformó la comprensión y el diseño de las organizaciones laborales, destacando la importancia de lograr un desarrollo industrial integral y productivo. La investigación sociotécnica, en sus inicios, se caracterizó por los beneficios bidireccionales que surgieron de la interacción entre elementos sociales y técnicos, subrayando la reciprocidad entre tecnología y sociedad.

Con el tiempo, la noción de sistemas sociotécnicos se ha refinado para incluir conexiones más profundas entre los aspectos sociales y técnicos de las tecnologías, que trascienden el nivel organizacional o de gestión. Un ejemplo es el marco de la construcción social de la tecnología, que sostiene que el desarrollo tecnológico es un proceso tanto técnico como social, influenciado por factores sociales, económicos y culturales (Pinch y Bijker, 1984). La noción de flexibilidad interpretativa de Bijker (1995) amplía esta visión, sugiriendo que un artefacto tecnológico puede tener diferentes significados y usos para distintos grupos sociales. Esto implica que el desarrollo y uso de la tecnología están sujetos a interpretaciones diversas, basadas en las necesidades, valores y contextos sociales de los usuarios.

En la actualidad, la adopción del concepto de sistemas sociotécnicos en el ámbito de la IA refleja las interdependencias y necesidades que vinculan su desarrollo con infraestructuras técnicas, comportamientos humanos y la participación de diversas instituciones. Es decir, este enfoque no reduce la comprensión y estudio de la IA a una simple tecnología aislada de los contextos sociales, sino que la interpreta en función de las dependencias e influencias mutuas que establece con ellos.

Existen varios ejemplos sobre cómo estas visiones sociotécnicas han sido adaptadas en el campo de la IA. Como analizan Arriagada-Bruneau et al. (en prensa), la mayoría de estas propuestas destacan diferentes conexiones entre IA y sociedad, lo que se traduce en diversidad propuestas metodológicas. Los autores señalan que el diseño basado en valores, propuesto por van de Poel (2020), implica que los artefactos técnicos, como la IA, pueden incorporar valores si se diseñan con intenciones específicas y si su uso contribuye a la realización de dichos valores. Por ejemplo, una IA diseñada para optimizar la eficiencia energética en sistemas de climatización de edificios puede programarse con el objetivo de minimizar el consumo de energía. Si sus algoritmos efectivamente reducen la huella de carbono

de las operaciones del edificio, entonces el artefacto técnico encarna el valor de la sostenibilidad ambiental. Esta traducción de valores en normas técnicas no es un mero añadido, sino un aspecto fundamental del diseño ético de la IA, crucial para garantizar que las tecnologías funcionen de manera beneficiosa y estén alineadas con principios éticos.

Además del diseño basado en valores, los autores destacan la relevancia de la adaptabilidad (Arriagada-Bruneau et al., en prensa), que es un concepto central en la teoría de los sistemas sociotécnicos y se refiere a la necesidad de que los sistemas evolucionen en respuesta a nueva información y normas cambiantes. Ejemplifican esto con lo propuesto por Chopra y Singh (2018), quienes sugieren directrices para una gobernanza robusta que se enfoque en la toma de decisiones morales, considerando no solo los agentes individuales y los problemas éticos aislados, sino también el contexto sociotécnico en el que operan los sistemas de IA. Esto permite que incorpore y se adapte a demandas y necesidades cambiantes, entendiendo el impacto transversal de la IA.

Asimismo, la adopción de una perspectiva sociotécnica enfatiza y promueve el diseño inclusivo y participativo, involucrando a diversas partes interesadas, lo que asegura que las tecnologías desarrolladas sean inclusivas y equitativas. Sasha Costanza-Chock (2020) utiliza la noción de ubicación contextual (*situatedness*) para transformar el diseño de soluciones tecnológicas, abarcando una tarea más amplia que simplemente atender los problemas específicos relacionados con datos o con la IA. La autora critica cómo los valores, prácticas y narrativas prevalecientes en el diseño reproducen desigualdades sistémicas y mantienen las distribuciones actuales de poder. Por ejemplo, se ajustan las funcionalidades tecnológicas para grupos sociales dominantes y rentables, mientras se descuidan las necesidades de las minorías, creando cargas adicionales para ellas.

Creemos, por tanto, que una de las incorporaciones conceptuales fundamentales en los programas educativos es el uso de estudios de caso específicos y ejemplos reales que reflejen las diversas aplicaciones y contextos de la IA. Considerar la concepción de la IA como un sistema sociotécnico, permitirá a los estudiantes comprender cómo los contextos sociales, culturales e históricos influyen en la creación y aplicación de la tecnología. Así, podrán desarrollar una visión más crítica y contextualizada de los problemas éticos, fomentando la comprensión de que la ética es parte integral de su quehacer profesional, puest está vinculada al desarrollo de esta tecnología, y no es solo un aspecto relevante, adicional o externo.

Adoptar una definición de la inteligencia artificial como un sistema sociotécnico implica reconocer que no es simplemente una tecnología autónoma o aislada, sino un conjunto de sistemas y prácticas integrado en contextos sociales, económicos y culturales, incluyendo contextos y posiciones epistémicas de los

desarrolladores. Esta perspectiva enfatiza que la creación, implementación y uso de la IA está intrínsecamente vinculada a factores humanos y sociales. Así, no podemos entender el desarrollo de la IA sin entender los conocimientos situados que la crean.

3.2 Los conocimientos situados

El concepto de conocimientos situados, propuesto por Haraway (1988), nos invita a reconocer que todo conocimiento es producido desde una posición particular, criticando una visión descontextualizada. De esta forma, propone que todos los conocimientos son situados y, por lo tanto, parciales. La idea base es que el conocimiento no es neutral ni universal, sino que está intrínsecamente vinculado a la posición particular del sujeto que lo produce. En el caso de la IA, se trata de la posición de los mismos desarrolladores, pero también la de los usuarios u otros sujetos de conocimiento involucrados en la implementación del sistema.

En su momento, Haraway presentaba esta idea en respuesta a la noción tradicional de objetividad asumida en círculos científicos, donde se aspiraba a hacer ciencia como una vista 'desde la nada', libre de sesgos y presunciones que pudiesen entorpecer el desarrollo científico. De manera, este fenómeno se puede analogar a lo que ocurre con las percepciones de los desarrolladores de IA, quienes al no percibir la IA como sociotécnica, tienden a ver su rol como técnica pura, donde la ética parece relevante pero solo en instancias posteriores o lejanas, es decir, cuando la IA es aplicada o usada, y no cuando es creada. Por esto, la noción de conocimientos situados es relevante, al desafiar las nociones tecnicistas y objetivistas de la IA. Así, si seguimos a Haraway, entenderemos que todo conocimiento está influido por la ubicación social, cultural y personal del desarrollador. Y, en lugar de aspirar a una falsa objetividad, debemos aspirar a contextualizar el conocimiento situado, que reconoce y utiliza la posición específica del sujeto como un recurso en la producción de conocimiento.

En el contexto de la ética aplicada en IA, esto se traduce, además, en que la formación ética debe considerar las experiencias y contextos específicos de los estudiantes, así como las particularidades del entorno social y técnico en el que operan los sistemas de IA. Esto también incluye un reconocimiento de los sesgos inherentes que los estudiantes pueden desarrollar por su propia experiencia de vida, tanto personal como profesional. Haraway nos llama a reconocer la parcialidad y la ubicación de nuestro conocimiento desde una humildad intelectual, que es necesaria para entender la complejidad del mundo. Por ejemplo, en lugar de tratar

de eliminar el sesgo en IA ciegamente, porque es uno de los 'ideales' objetivos tradicionales, deberíamos reflexionar sobre nuestras propias posiciones y cómo influyen en nuestro entendimiento. Este enfoque puede enriquecer nuestras investigaciones al hacer explícitos los contextos desde los que observamos y analizamos el mundo y, además, pone énfasis en un aspecto poco explorado en la literatura en ética de IA, que son los sesgos profesionales que pueden surgir en los equipos de desarrollo.

La adopción de los conocimientos situados como base de la ética de IA puede revelar cómo las tecnologías no son neutrales, sino que reflejan y refuerzan las estructuras de poder existentes. Por ejemplo, al estudiar cómo se diseñan y utilizan los algoritmos de reconocimiento facial, es crucial considerar cómo las perspectivas y sesgos de los diseñadores –a menudo provenientes de contextos demográficos específicos– afectan la precisión y equidad del algoritmo para diferentes grupos de personas, como ocurrió con lo denunciado por Buolamwini y Gebru (2018)[4].

La inclusión de estudios de caso y ejemplos específicos de dilemas éticos puede ayudar a situar el conocimiento, logrando que la formación ética sea más relevante y aplicable para los estudiantes. Esto, sin embargo, debe orientarse, idealmente, a dos grandes objetivos:

- Internalizar cuestionamientos éticos y conceptos básicos desde los inicios de la carrera. Esto es, no esperar a que los estudiantes tengan cursos de ética para hablar de ética.
- Incorporar cuestionamientos éticos en diferentes cursos, en particular los que integran la creación de códigos y desarrollo de modelos. De esta forma, se internaliza la simbiosis entre decisiones técnicas y éticas.

Respecto del primer objetivo, es crucial que los cuestionamientos éticos y conceptos fundamentales se incorporen en cursos de alfabetización o introductorios. Esto promueve que los estudiantes desarrollen una conciencia ética desde el principio de su carrera. Esta integración temprana permite que los futuros profesionales de la IA no solo se familiaricen con los principios éticos, sino que también los consideren como parte integral de su proceso de pensamiento y toma de decisiones.

[4] Lo denunciado por Joy Buolamwini y Timnit Gebru se centra en sesgos de género y raza en los sistemas de reconocimiento facial. En su investigación "Gender Shades", analizaron cómo los algoritmos de reconocimiento facial de empresas líderes (Microsoft, IBM y Face++) presentan mayores tasas de error al identificar el género de personas negras, especialmente mujeres, en comparación con hombres blancos.

Por ejemplo, cursos introductorios sobre IA o programación podrían incluir módulos específicos sobre la historia de los dilemas éticos en tecnología, análisis de casos donde la falta de consideraciones éticas ha llevado a consecuencias negativas y discusiones sobre las responsabilidades sociales de los científicos de datos y desarrolladores de IA. Situar conocimientos implica seguir un enfoque basado en la teoría ética y, al mismo tiempo, discernir elementos, acciones y decisiones en contextos prácticos que implican cuestionamientos éticos.

El segundo objetivo pretende mostrar que la ética no debe ser vista como un área de conocimiento aislada. En cambio, debe entenderse como una consideración omnipresente en todas las fases del desarrollo de IA. Al respecto, la integración de casos específicos en cursos estratégicos puede ser una estrategia efectiva para lograrlo.

Por ejemplo, en un curso sobre aprendizaje automático, se podrían analizar casos en los que los algoritmos han perpetuado sesgos raciales o de género. Los estudiantes podrían trabajar en proyectos que requieran identificar y mitigar posibles sesgos en los conjuntos de datos y en los modelos que desarrollan, además de identificar aspectos sobre las decisiones técnicas de diseño que pueden causar faltas a la privacidad o transparencia. Asimismo, podrían discutir las implicaciones éticas de sus propias decisiones de diseño y cómo podrían afectar a diferentes grupos sociales. De esta forma, los estudiantes no solo aprenden a desarrollar modelos técnicamente competentes, sino también modelos que sean éticamente responsables.

3.3 La objetividad fuerte y una perspectiva socialmente fundada

Sobre la base de los conocimientos situados que nos plantea Haraway, nos referimos ahora al concepto de objetividad fuerte de Harding (1992, 1995, 2015). La autora reconoce que hay objetividades débiles, a saber, aquellas que ven posible una ciencia neutral. Para Harding, evitar esas debilidades implica desarrollar una comprensión más robusta y objetiva de la realidad, que se obtiene al incluir múltiples perspectivas, especialmente aquellas de los grupos tradicionalmente marginados o subrepresentados.

La teoría del punto de vista feminista de Sandra Harding propone que la diversidad entre los investigadores es una ventaja epistémica para la comunidad científica (Wylie, 2003), pero con una condición específica: esta ventaja se deriva de la diversidad en las posiciones sociales de los investigadores y participantes. Harding sostiene que las personas en posiciones sociales desfavorecidas, debido a su experiencia única, tienen una ventaja epistémica en la comprensión de la realidad social que les afecta. Según ella, estas posiciones generan perspectivas

menos distorsionadas que las de aquellos en posiciones privilegiadas. La inclusión de diversidades (grupos, personas, perspectivas y narrativas) no se traduce solo en una cuestión de justicia social, sino que sus beneficios epistémicos reflejan una mejora en la calidad y el alcance del conocimiento científico producido. Este concepto se denomina 'privilegio epistémico', el que está vinculado a una forma particular de objetividad que es la objetividad fuerte.De acuerdo con Harding, no se puede ignorar cómo se produce el conocimiento en el mundo real y cómo se practica la ciencia. Las organizaciones públicas y privadas influyen directamente en las formas en que se produce el conocimiento científico, y eso es aún más decisivo en el avance de la IA, cuyas grandes potencias son gigantes corporaciones privadas con intereses económicos y políticos. Si le sumamos la poca diversidad de personas en la participación activa en ciencia, tecnología, ingeniería y matemáticas, y en comunidades de IA, se genera una situación donde se replican valores e intereses de grupos privilegiados.

La diversidad en las posiciones sociales dentro de una comunidad que busca conocimiento es valiosa porque hay muchas formas en las que una persona puede estar en desventaja. Sin embargo, hay diversas críticas que se le hacen a la postura de Harding y que es necesario atender antes de integrar este concepto en nuestra propuesta.

Las filósofas feministas Helen Longino (1993) y Louise Antony (1993) han criticado una posible contradicción entre los asumidos conocimientos situados que adopta Harding y la tesis de privilegio epistémico –como la llama Kristina Rolin (2006)–. La tesis del privilegio epistémico sugiere que algunas perspectivas sociales serían más objetivas o menos sesgadas que otras respecto de ciertas experiencias (en particular de grupos marginalizados que las experimentan). Pero la tesis del conocimiento situado desafía esta idea, ya que todo conocimiento está influenciado por el contexto social en el que se genera y esto no nos permitiría atribuir una objetividad fuerte a esas experiencias epistémicas. No podríamos entonces hablar de una perspectiva completamente imparcial, ya que cada punto de vista está condicionado por la posición social de quien lo sostiene.

Esta contradicción, entre la idea de que algunas perspectivas son menos parciales y la idea de que todo conocimiento está condicionado por su contexto social, Louise Antony lo ha llamado la paradoja del sesgo. En otras palabras, al afirmar que todo conocimiento es parcial, la epistemología del punto de vista feminista pone en duda la posibilidad de imparcialidad, lo que complica la capacidad de criticar o evaluar objetivamente diferentes perspectivas.

Aunque Harding hace referencia a estas dificultades (Harding, 1991), no las resuelve por completo. Sin embargo, Kristina Rolin (2006) ofrece una posible solución a esa paradoja, al argumentar que una teoría contextualista de la justificación epistémica puede resolver la paradoja del sesgo al permitir evaluar las perspectivas

sociales sin depender de una visión completamente imparcial o neutral, ya que el contextualismo ofrece un marco para comparar y valorar diferentes puntos de vista, reconociendo que todo conocimiento está influido por su contexto.

Basado en la definición que entrega Michael Williams (2001), Rolin explica que la justificación epistémica, en el marco del contextualismo, se produce en un contexto de suposiciones que actúan como derechos predeterminados. Estas suposiciones se aceptan inicialmente sin cuestionamiento, pero pueden desafiarse y reevaluarse cuando surgen nuevas circunstancias o contextos (no son meramente relativas o subjetivas a la experiencia epistémica). La recontextualización de estas suposiciones es un proceso iterativo, que se anticiparía a un posible regreso vicioso en la justificación con intenciones de alcanzar una conclusión definitiva.

Este enfoque contextualista permitiría evaluar perspectivas sociales en un contexto específico, sin necesidad de recurrir a un estándar universal o una "visión desde ningún lugar", tal y como lo defiende también Harding con su idea de objetividad fuerte, pero sin entrar en conceptualizaciones que puedan causar la paradoja del sesgo.

Así, Rolin plantea que la epistemología feminista del punto de vista ha dependido históricamente de metáforas visuales y espaciales, usando nociones como 'perspectiva' y 'puntos de vista'. Pero hablar de un punto de vista sugiere una posición fija desde la que se observa un objeto de estudio, evocando que necesitamos esa visión desde ningún lugar para comparar y evaluar diferentes perspectivas. Por tanto, la propuesta de Rolin ofrece un estándar de imparcialidad situado como el que se defiende en diversas teorías feministas y que permite evaluar los méritos relativos de distintas perspectivas socialmente fundamentadas. Nos parece que esta distinción es necesaria, ya que debemos situar nuestros procesos epistémicos y, por tanto, en vez de atribuir objetividades epistémicas a ciertas comunidades marginalizadas, podemos construir visiones contextualizadas que nos acerquen a sus experiencias epistémicas situándolas, pero exigiendo una debida justificación para su aceptación.

Nos parece crucial, por ende, incorporar esta perspectiva revisada que propone Rolin, ya que el contextualismo ofrece un marco para esta evaluación al situar la justificación epistémica en un contexto de suposiciones predeterminadas, que pueden ser desafiadas y reevaluadas cuando surgen nuevas críticas o perspectivas. Esto nos ayuda a cumplir los objetivos fundacionales de las posturas feministas que buscan fomentar la diversidad y la participación en la comunidad científica, ya que responden también al desarrollo propio del avance científica, a saber, que estas suposiciones y perspectivas epistémicas sean defendidas, modificadas o abandonadas, para así consolidar un punto de vista, que resulta de un compromiso comunitario con la diversidad y la apertura al escrutinio externo.

Traducir la adopción de estas teorías feministas al desarrollo de la IA, por tanto, significa adentrarse en el diseño y creación de sistemas de IA entendiendo las posiciones epistémicas y sociales de sus creadores, usuarios, financiadores y otros incumbentes que pueden verse directa o indirectamente afectados por estos sistemas. Lo que nos lleva también a considerar cómo debemos entender estas dinámicas en las comunidades epistémicas marginalizadas.

3.4 Comunidades epistémicas y pertenencias de grupo

La adopción de una contextualización y posicionamiento de las perspectivas epistémicas permite rastrear las relaciones de poder de maneras epistémicamente significativas. Las limitaciones que encontramos en la falta de acceso a datos y estudios que reflejen esas diversidades afectan inevitablemente la calidad ética y técnica de los sistemas de IA. Las experiencias y conocimientos de los grupos marginados a menudo no se consideran en el desarrollo de tecnologías de IA; sin embargo, estas perspectivas pueden proporcionar información valiosa sobre cómo y por qué ciertos datos o suposiciones pueden estar sesgados. Por ejemplo, un equipo de desarrollo diverso podría cuestionar las suposiciones implícitas en un modelo de predicción del crimen, proponiendo métodos alternativos que consideren factores socioeconómicos e históricos que no se reflejan en los datos disponibles.

Otro caso ilustrativo es el de los algoritmos de contratación utilizados en procesos de selección de personal. Estos sistemas, entrenados con datos históricos sesgados, tienden a perpetuar las desigualdades preexistentes, favoreciendo a candidatos similares a los históricamente contratados. La inclusión de un equipo de desarrollo diverso podría mitigar estos efectos, al introducir perspectivas que reconocen y contrarrestan los prejuicios históricos, y al proponer criterios de selección más inclusivos que reflejen una gama más amplia de habilidades y experiencias. Además, para abordar estos sesgos de manera más efectiva, se podrían implementar otras estrategias, como la consulta directa con grupos minoritarios durante el proceso de desarrollo del algoritmo. Involucrar a estos grupos en la revisión de los criterios utilizados y en la interpretación y validación de los resultados del algoritmo, podría ayudar a garantizar que los sistemas reflejen una comprensión más completa y matizada de las experiencias y desafíos de personas de diferentes orígenes.

Por esto, es fundamental iniciar el diseño de un sistema de IA a partir de las preguntas de las vidas de estos grupos sociales excluidos, y tomar sus testimonios como punto de partida para formular problemas. En este contexto, se trata de los grupos de los cuales no tenemos datos –o buenos datos– y de grupos que no participan

en la toma de decisiones sobre el diseño de los sistemas de IA, aun cuando sean los usuarios finales. Sus preguntas no solo serían novedosas y valiosas, sino que también los procedimientos que se emplearían para responderlas diferirían de los utilizados por personas pertenecientes a grupos sociales privilegiados o dominantes. Al tener procesos de desarrollo inclusivo que consideren estas brechas y debilidades, que surgen desde nuestra realidad social, es posible alcanzar lo que Harding denomina otra lógica de la investigación científica (2015), que en este caso podrían traducirse a otra lógica para desarrollar IA.

Aplicar esta idea a la educación en ética de la IA involucra incorporar una variedad de puntos de vista en el currículo y fomentar la participación activa de estos grupos, junto con entender sus narrativas y contextos en el desarrollo y la implementación de la IA, lo que se logra mediante la inclusión de lecturas y recursos de autores diversos, así como la creación de espacios de diálogo y la reflexión crítica sobre cómo las tecnologías de IA pueden afectar de manera diferencial a distintas comunidades.

Lo anterior es necesario, debido a que, como enfatiza Intemann (2010), el que una comunidad epistémica incluya a un miembro de un grupo marginado no implica automáticamente ventajas epistémicas. Para lograr un punto de vista que represente las necesidades y realidades de ese grupo marginado –y obtener el privilegio epistémico de incluirse en el entendimiento–, se requiere un pensamiento crítico y una reflexión rigurosa, que es colectiva, ya que el privilegio epistémico no es alcanzado por un individuo, sino por la comunidad epistémica en su conjunto (Intemann, 2010).

Por lo tanto, para acceder al punto de vista de un grupo oprimido, no es necesario ni suficiente ser miembro de ese grupo. No es suficiente porque los miembros deben tomar conciencia de su identidad de grupo y lograr una comprensión compartida de las relaciones de poder que causan la opresión. Tampoco es necesario porque los valores e intereses del grupo oprimido son públicamente accesibles, de modo que cualquiera puede teorizar fenómenos en relación con los valores e intereses del grupo y, en consecuencia, malinterpretarlos.

Por otra parte, el punto de vista adquirido debe alcanzarse mediante el autoconocimiento de agentes autónomos que participan en la concienciación y, así, el privilegio epistémico puede desplazarse al grupo, que llega a definirse a sí mismo como un agente político colectivo. Esto implica, necesariamente, que para ahondar en las desigualdades, inequidades y limitaciones que conlleva el impacto social de una IA, el grupo privilegiado (en este caso, los desarrolladores) deben ser agentes partícipes de una concienciación que les permita desplazar su privilegio epistémico para darle espacio y poder a los grupos minoritarios. Solo así se adoptaría una objetividad fuerte. Por tanto, incluir contenidos o motivar acciones que busquen una inclusión o reconocimiento de diversidades mediante formalizaciones (por

ejemplo, reducir sesgos con herramientas de mitigación) o de adopción vaga de principios éticos, sería fomentar una objetividad débil.

La formación ética es una oportunidad, en tanto posibilita un proceso de concienciación que trasciende la pertenencia a un grupo. El desarrollo del discernimiento ético –habilidad que un curso de ética debería cultivar– permite reflexionar sobre cómo representar de manera justa los intereses de diversos grupos. Como señala Young (1989), aunque la pertenencia a un grupo puede influir en la perspectiva de una persona, no la determina por completo. El mantenimiento de la identidad grupal y la influencia de nuestras experiencias personales, como afirma Young, no impide el desarrollo de una actitud pública, referida a la apertura para escuchar las demandas y necesidades de otros, más allá de nuestros propios intereses. Lo que se requiere es, entonces, que una persona logre tomar una distancia crítica de sus intereses. En este contexto, se podrían fomentar en clases estrategias metodológicas como debates y mesas redondas, en que se promueva la actitud pública. En estas actividades, se pueden abordar aspectos clave como la representatividad de datos, la integración de contextos para priorizar variables en el entrenamiento de modelos y levantamiento de necesidades en etapas de formulación de problemas.

Además, Young (1989) señala que el representar de manera explícita las perspectivas de diferentes grupos fomenta la distancia crítica sin intentar adoptar una postura imparcial. La representación de voces diversas puede lograrse con lecturas de autores de diferentes orígenes y con la reflexión guiada con ejemplos concretos: por un lado, mostrar casos donde no se han considerado necesidades de ciertos grupos y, por otro, destacando casos en los que se ha implementado un proceso de participación inclusiva en el diseño de sistemas de IA.

3.5 El feminismo de datos

En una versión actualizada de las posturas de Haraway y Harding, el feminismo de datos, conceptualizado por D'Ignazio y Klein (2020), añade otra capa a este marco conceptual para integrar la ética en la IA. Las autoras enfatizan la necesidad de una sensibilidad crítica hacia los datos y los procesos de recopilación y análisis de los mismos, reconociendo cómo las estructuras de poder y las desigualdades sociales se reflejan y se perpetúan a través de los datos –entendiendo, además, que este es el sustrato para entrenar modelos de IA–.

D'Ignazio y Klein (2020) destacan que las experiencias vividas son esenciales para comprender y dar forma a los usos y limitaciones de los datos. Los datos son

representaciones necesariamente reductivas de estas experiencias, donde las vidas de las personas se traducen en números, palabras o imágenes específicas, omitiendo sus experiencias. A la vez, otras personas, usualmente más privilegiadas, tienen roles influyentes en la recopilación, etiquetado, análisis y uso de estos datos. Ambos actores –tanto aquellos cuyos datos se utilizan como aquellos que los usan– operan en un contexto donde coexisten y se combinan múltiples estructuras de opresión, como género, raza, clase y capacidad, entre otros. El feminismo de datos enfatiza la importancia de examinar el contexto en el que se producen los datos y los objetivos de su uso desde una perspectiva interseccional.

Por ejemplo, consideremos el caso del algoritmo de evaluación de riesgos Compas, utilizado en el sistema de justicia penal estadounidense para evaluar el riesgo de reincidencia de los acusados. Este algoritmo ha sido criticado ampliamente por sus sesgos raciales, pues algunos estudios han demostrado que tiende a predecir incorrectamente que los acusados negros tienen mayor probabilidad de reincidir en comparación con los acusados blancos (Angwin et al., 2016). Este algoritmo perpetúa las desigualdades raciales existentes en el sistema de justicia penal. El feminismo de datos nos invita a cuestionar por qué existen más datos sobre acusados negros que blancos y cómo esto se relaciona con una vigilancia históricamente más intensa en vecindarios habitados mayoritariamente por personas de color. Esta reflexión crítica es esencial para entender las limitaciones y sesgos de los datos utilizados, donde no se trata de identificar aspectos del modelo que puedan llevar a sesgos por errores técnicos, sino de examinar la fuente misma de la desigualdad, que se refleja en el tipo de datos y las referencias históricas que poseen.

El caso de Compas también sirve para ilustrar la importancia de estudiar los problemas de equidad presentes en los datos desde el marco de feminismo de datos, que destaca la interseccionalidad como un aspecto crucial para entender y abordar las desigualdades en los sistemas de IA. El sesgo racial en Compas no solo afecta a los acusados negros de manera desproporcionada, sino que también interactúa con otras formas de opresión. Por ejemplo, una mujer latinoamericana enfrenta una combinación única de desafíos debido a su clasificación étnica y de género (Hamilton, 2019a, 2019b). El algoritmo de Compas, al no considerar estas complejidades interseccionales, puede contribuir a decisiones judiciales injustas y desiguales.

Un paso crucial para integrar esto en el currículo es incorporar estudios de casos emblemáticos, como el algoritmo Compas, para que los estudiantes comprendan cómo los datos que alimentan estos sistemas reflejan y perpetúan desigualdades estructurales. Según el feminismo de datos, los datos no son neutrales, sino producto de relaciones de poder que afectan a diferentes grupos de manera dispar. Estos casos no deben ser un simple instrumento ilustrativo, sino un puntapié inicial para ahondar en cómo decisiones técnicas y prejuicios sociales, estructuras institucionales,

y variados otros estímulos se interconectan para generar discriminación mediante sistemas algorítmicos o de IA.

Los estudiantes deben aprender a identificar y cuestionar diferentes elementos éticos involucrados en el desarrollo de la IA, como los sesgos en los conjuntos de datos y en los modelos, entendiendo cómo estos sesgos perpetúan y amplifican las desigualdades. En este caso, además, correspondería identificar diferentes tipos de sesgos que influyen en el proceso: no solo sesgos técnicos (computacionales o estadísticos), sino que también sociales (prejuicios estructurales) y cognitivos (como sus propios sesgos profesionales) (Arriagada-Bruneau, 2024). Asimismo, la participación activa y el aprendizaje basado en proyectos son fundamentales para consolidar esta formación ética. Los estudiantes deben ser desafiados a diseñar algoritmos con un enfoque explícitamente ético, como un modelo de evaluación de riesgos que evite los sesgos raciales presentes en COMPAS y considere las complejidades interseccionales de los datos, para ello es útil complementar o integrar razonamiento ético con talleres de programación que fomenten la diversidad epistémica y la enseñanza de enfoques críticos en el análisis de datos.

Para abordar estos problemas de manera efectiva, es crucial que la enseñanza de la ética de la IA incluya estas perspectivas. Los estudiantes deben aprender a reconocer que los datos no son neutrales, sino producto de relaciones sociales desiguales, y que las interseccionalidades son parte de los desafíos de trabajar con sistemas sociotécnicos y de impacto social como las inteligencias artificiales. De este modo, deben ser conscientes de que los datos que alimentan los algoritmos no solo reflejan la realidad social, sino que también la construyen. Esta comprensión crítica de los datos ayuda a identificar y mitigar los sesgos inherentes que perpetúan las desigualdades, incluyendo los sesgos que provienen desde sus propias posturas epistémicas y las estructuras de privilegio que rodean el avance de la IA.

Los datos, en su esencia, son abstracciones de la realidad, que buscan simplificar y cuantificar experiencias humanas complejas. Esta reducción, sin embargo, es inevitablemente incompleta. La epistemología feminista nos recuerda que cualquier proceso de abstracción deja fuera aspectos vitales de la experiencia humana, particularmente aquellos relacionados con grupos marginados. Este sesgo no es simplemente una falta técnica; más bien, es una manifestación de las dinámicas de poder que determinan qué datos se recopilan, cómo se etiquetan y para qué se utilizan.

Consecuentemente, un currículo que adopte la perspectiva del feminismo de datos puede fomentar la reflexión crítica sobre las fuentes de datos, las metodologías de análisis y las implicaciones éticas de las decisiones algorítmicas. Al hacerlo, se prepara a los futuros desarrolladores de IA para ser técnicamente competentes y, al mismo tiempo, ser agentes de cambio ético, capaces de diseñar tecnologías que promuevan la equidad y la justicia social.

4. Conclusiones

En este artículo, planteamos un marco conceptual basado en el feminismo de la ciencia y el feminismo de datos para guiar la integración de la ética aplicada en los currículos universitarios. Este enfoque subraya la importancia de considerar la IA como un sistema sociotécnico, donde los conocimientos y prácticas no son neutrales, sino que están profundamente influidos por contextos sociales, culturales y políticos. Este marco busca desafiar las pretensiones de universalidad en los análisis éticos, promoviendo una visión crítica y contextualizada de la ética en IA, puesto que esta contextualización complementada con innovación docente para enseñar la ética aplicada a la IA ofrece una forma de avanzar en su validación como una disciplina de estudio que es necesaria en el currículo de educación superior chileno.

Esta necesidad surge, en parte, porque las estrategias pedagógicas más tradicionales –que se limitan a enseñar una ética general o a incluir cursos éticos aislados en el currículo– resultan insuficientes para lograr una verdadera integración y, por lo tanto, socavan su validación. Esto se debe a que, en su mayoría, estas estrategias no son capaces de formar a los estudiantes con herramientas técnicas y de pensamiento crítico que les ayuden a identificar contextos propios de un sistema sociotécnico como lo es la IA. Un aspecto crucial de esta identificación es que la naturaleza no técnica de gran parte del contenido ético debe reflejarse en los métodos de enseñanza y evaluación, lo que puede requerir tanto a estudiantes como a profesores adoptar formas de pensar poco familiares o ajenas, ya que los cursos diseñados para desarrollar habilidades en ingeniería generalmente no los exponen a conceptos o diversos métodos filosóficos (Tuovinen y Rohunen, 2021).

Las discusiones actuales en la academia giran en torno a las estrategias más efectivas para enseñar ética en IA, ya sea mediante cursos independientes, seminarios, incorporando la ética en el último año de estudio, o integrando componentes éticos a lo largo del currículo. No obstante, sigue existiendo una gran brecha en la implementación real de estas estrategias, lo que indica la necesidad de una incorporación más estructurada y generalizada de la ética en la educación de IA. Esta estructura, sin embargo, siguiendo el marco propuesto en este artículo, no ha de ser estándar ni universalizable, pues depende de las demandas contextuales de diferentes países e instituciones, considerando sus tradiciones educativas y limitaciones, las posibilidades de innovación, sus recursos, y los requisitos para los profesionales derivados de la legislación local y la cultura ética institucional.

Igualmente, es fundamental considerar el contexto chileno y latinoamericano. En América Latina, donde las sociedades son multiculturales y enfrentan vulnerabilidades debido a las altas tasas de pobreza y desigualdad, es fundamental considerar la dependencia de tecnologías extranjeras. Esta dependencia puede forzar a la población a elegir entre aceptar sistemas que pueden no comprender

totalmente o correr el riesgo de quedar rezagados (Mancilla-Caceres y Estrada-Villalta, 2022). De esta forma, entender los contextos propios de la IA en Chile, y en la región latinoamericana, es clave.

En virtud de lo expuesto, nos parece que esta contextualización en los planes curriculares para incluir la ética aplicada a la IA debe seguir los lineamientos del feminismo de la ciencia y el feminismo de datos sobre la base de las distinciones ofrecidas anteriormente, las que pueden resumirse en la inclusión de voces y experiencias de grupos marginados en la enseñanza de la ética aplicada a la IA. Esto no solo implica la representación demográfica en las aulas, sino también la inclusión de estudios de caso y materiales de enseñanza que reflejen las experiencias de diversas comunidades subrepresentadas y diversidades metodológicas para integrar diversidades epistémicas.

Además, una pedagogía inspirada en estos feminismos debe fomentar la reflexión crítica y la participación activa de los estudiantes. Esto puede lograrse mediante metodologías de enseñanza que promuevan el pensamiento crítico, como el aprendizaje basado en proyectos, debates y talleres participativos. Los estudiantes podrían trabajar en proyectos que requieran la identificación de sesgos en conjuntos de datos, el desarrollo de algoritmos equitativos y la creación de políticas éticas para la implementación de IA. O, como sugieren algunos investigadores, también se puede fomentar el aprendizaje en torno a habilidades técnicas centrado en consideraciones sociales o en lo que algunos instructores llaman 'tecnología para el bien social' (Garrett et al., 2021).

El feminismo de la ciencia y el feminismo de datos llaman también a una colaboración transdisciplinaria que reconozca la complejidad de los sistemas sociotécnicos. En la educación de IA, esto podría traducirse en cursos que combinan los conocimientos técnicos con disciplinas como sociología, filosofía, lingüística, estudios de género y ciencia política, explorando módulos de codocencia para fortalecer la narrativa ética a lo largo del currículo. Esta aproximación permitiría a los estudiantes entender mejor el impacto social de la tecnología y las implicaciones éticas de sus decisiones técnicas.

La integración efectiva de la ética aplicada a la IA, por ende, requiere un enfoque que vaya más allá de los cursos tradicionales y aislados de ética. Sugerimos que esta educación debe estar profundamente arraigada en la promoción de una educación inclusiva, crítica y reflexiva. Al seguir estos lineamientos, podemos preparar a los futuros profesionales que trabajen en áreas de desarrollo e implementación de sistemas de IA para ser éticamente responsables, asegurando que el desarrollo de tecnologías beneficie a la sociedad, respetando la diversidad de experiencias humanas.

Referencias

Angwin, J., Larson, J., Mattu, S., y Kirchner, L. (2016). Machine Bias: There's software used across the country to predict future criminals. And it's biased against blacks. *Pro Publica*. https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing.

Antony, L. (1993). "Quine as feminist: Th e radical import of naturalized epistemology". En L. Antony y C. Witt (eds.), *A mind Of One's Own.*

Arriagada-Bruneau, G. (2024) *Los sesgos del algoritmo: la importancia de diseñar una inteligencia artificial ética e inclusiva.* La Pollera.

Arriagada-Bruneau, G., López, C., Mendoza, M. (en prensa) *Ethics of Artificial Intelligence and Information Technologies.* Taylor and Francis. CRC Press.

Beauchamp, T. L. (2005). The Nature of Applied Ethics. En *A Companion to Applied Ethics* (pp. 1-16). John Wiley & Sons. https://doi.org/10.1002/9780470996621.ch1.

Bergenhenegouwen, G. (1987). Hidden curriculum in the university. *Higher Education*, *16*(5), 535–543. https://doi.org/10.1007/BF00128420

Bijker, W. E. (1995). *Of Bicycles, Bakelites, and Bulbs: Toward a Theory of Sociotechnical Change.* MIT Press.

Bowen, G. A. (2009). Document Analysis as a Qualitative Research Method. *Qualitative Research Journal, 9*(2), 27-40. https://doi.org/10.3316/QRJ0902027.

Buolamwini, J., y Gebru, T. (2018). Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. *Proceedings of the 1st Conference on Fairness, Accountability and Transparency,* 77-91. https://proceedings.mlr.press/v81/buolamwini18a.html.

Chopra, A. K., y Singh, M. P. (2018). *Sociotechnical Systems and Ethics in the Large*. Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society, 48-53. https://doi.org/10.1145/3278721.3278740.

Costanza-Chock, S. (2020). *Design Justice: Community-Led Practices to Build the Worlds We Need*. MIT Press.

D'Ignazio, C., y Klein, L. F. (2020). *Data Feminism*. MIT Press.

Dignum, V. (2021). The role and challenges of education for responsible AI. *London Review of Education, 19*(1), Article 1.

Fiesler, C., Garrett, N., y Beard, N. (2020). What Do We Teach When We Teach Tech Ethics? A Syllabi Analysis. En *Proceedings of the 51st ACM Technical Symposium on Computer Science Education* (pp. 289-295). Association for Computing Machinery. https://doi.org/10.1145/3328778.3366825.

Garrett, N., Beard, N., y Fiesler, C. (2020). More Than 'If Time Allows': The Role of Ethics in AI Education. *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, 272-278. https://doi.org/10.1145/3375627.3375868.

Goetze, T. S. (2023). Integrating Ethics into Computer Science Education: Multi-, Inter-, and Transdisciplinary Approaches. *Proceedings of the 54th ACM Technical Symposium on Computer Science Education V*. 1, 645-651. https://doi.org/10.1145/3545945.3569792.

Gorur, R., Hoon, L., y Kowal, E. (2020). Computer Science Ethics Education in Australia – A Work in Progress. *IEEE International Conference on Teaching, Assessment, and Learning for Engineering* (TALE), 945-947. https://doi.org/10.1109/TALE48869.2020.9368375.

Griffin, T. A., Green, B. P., y Welie, J. V. M. (2023). The ethical agency of AI developers. AI and Ethics. https://doi.org/10.1007/s43681-022-00256-3.

Hagendorff, T. (2020). The Ethics of AI Ethics: An Evaluation of Guidelines. *Minds and Machines, 30(*1), 99-120. https://doi.org/10.1007/s11023-020-09517-8.

Hamilton, M. (2019a). The Biased Algorithm: Evidence of Disparate Impact on Hispanics. *American Criminal Law Review*, 56(4), 1553-1577.

___. (2019b). The sexist algorithm. *Behavioral Sciences & the Law, 37*(2), 145-157. https://doi.org/10.1002/bsl.2406.

Haraway, D. (1988). Situated Knowledges: The Science Question in Feminism and the Privilege of Partial Perspective. *Feminist Studies, 14*(3), 575-599. https://doi.org/10.2307/3178066.

Harding, S. (1991). *Whose Science? Whose Knowledge?: Thinking from Women's Lives.* Cornell UP. https://www.jstor.org/stable/10.7591/j.ctt1hhfnmg.

___. (1992). Rethinking Standpoint Epistemology: What Is "Strong Objectivity"? En *Feminist Epistemologies*. Routledge.

___. (1995). 'Strong Objectivity': A Response to the New Objectivity Question. *Synthese, 104*(3), 331-349.

___. (2015). *Objectivity and Diversity: Another Logic of Scientific Research*. University of Chicago Press. https://press.uchicago.edu/ucp/books/book/chicago/O/bo19804521.html.

Intemann, K. (2010). 25 Years of Feminist Empiricism and Standpoint Theory: Where Are We Now? *Hypatia, 25*(4), 778-796. https://doi.org/10.1111/j.1527-2001.2010.01138.

Longino, H. E. (1993). Feminist Standpoint Theory and the Problems of Knowledge. *Signs*, *19*(1), 201-212.

López, C., Arriagada Bruneau, G., y Davidoff, A. (2023). ¿Cómo navegar el camino hacia la ética en IA? *Bits de ciencia*, 25, 36-43. http://www.dcc.uchile.cl/difusion/revista/25.

Mancilla-Caceres, J. F., y Estrada-Villalta, S. (2022). The Ethical Considerations of AI in Latin America. *Digital Society, 1*(2), 16. https://doi.org/10.1007/s44206-022-00018-y.

Margolis, E. (2001). *The Hidden Curriculum in Higher Education*. Routledge. https://doi.org/10.4324/9780203901854.

Mittelstadt, B. (2019). Principles alone cannot guarantee ethical AI. *Nature Machine Intelligence, 1*(11), 501-507. https://doi.org/10.1038/s42256-019-0114-4.

Munn, L. (2023). The uselessness of AI ethics. *AI and Ethics, 3*(3), 869-877. https://doi.org/10.1007/s43681-022-00209-w.

Pinch, T. J., y Bijker, W. E. (1984). The Social Construction of Facts and Artefacts: Or How the Sociology of Science and the Sociology of Technology Might Benefit Each Other. *Social Studies of Science, 14*(3), 399-441.

Poel van de, I. (2020). Embedding Values in Artificial Intelligence (AI) Systems. *Minds and Machines, 30*(3), 385-409. https://doi.org/10.1007/s11023-020-09537-4

Raji, I. D., Scheuerman, M. K., y Amironesei, R. (2021). You Can't Sit With Us: Exclusionary Pedagogy in AI Ethics Education. *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 515-525. https://doi.org/10.1145/3442188.3445914.

Rolin, K. (2006). The Bias Paradox in Feminist Standpoint Epistemology. *Episteme: A Journal of Social Epistemology, 3*(1), 125-136.

Trist, E. L. (1981). *The Evolution of Socio-technical Systems: A Conceptual Framework and an Action Research Program*. Ontario Ministry of Labour, Ontario Quality of Working Life Centre.

Tuovinen, L., y Rohunen, A. (2021). Teaching AI ethics to engineering students: Reflections on syllabus design and teaching methods. *Proceedings of the Conference on Technology Ethics, Tethics*, Turku, Finland, October, 20-22.

Williams, M. (2001). *Problems of Knowledge: A Critical Introduction to Epistemology*. Oxford UP.

Wylie, A. (2003). Why Standpoint Matters. En R. Figueroa & S. G. Harding (Eds.), *Science and other cultures: Issues in philosophies of science and technology* (pp. 26-48). Routledge. https://philarchive.org/rec/WYLWSM.

Young, I. M. (1989). Polity and group difference: A critique of the ideal of universal citizenship. *Ethics*, *99*(2), 250-274.