

# Neuroética, neuroderecho e inteligencia artificial

JORGE ALBERTO Álvarez Díaz  
JUAN DANIEL Macías Sierra  
MARÍA DE JESÚS Medina Arellano  
FRIDA Romay Hidalgo





INSTITUTO DE INVESTIGACIONES JURÍDICAS  
Serie Enseñanza Transversal en Bioética y Bioderecho:  
Cuadernos Digitales de Casos, núm. 15



COORDINACIÓN EDITORIAL

**Lic. Raúl Márquez Romero**  
*Secretario técnico*

**Lic. Karla Beatriz Templos Nuñez**  
*Jefa de la Biblioteca Jurídica Virtual*

Violet Anzures Martínez  
y Karla Beatriz Templos Nuñez

APOYO EDITORIAL

Edith Aguilar Gálvez

ELABORACIÓN DE PORTADA

**ISBN: 978-607-30-6440-8**

# 15. NEUROÉTICA, NEURODERECHO E INTELIGENCIA ARTIFICIAL

ENSEÑANZA TRANSVERSAL EN BIOÉTICA Y BIODERECHO:  
CUADERNILLOS DIGITALES DE CASOS

Jorge Alberto ÁLVAREZ DÍAZ

Juan Daniel MACÍAS SIERRA

María de Jesús MEDINA ARELLANO

Frida ROMAY HIDALGO

# CONTENIDO

## I. Agradecimientos

## II. Introducción

## III. Las neurociencias y el derecho ¿Neuroderecho?

1. ¿Qué es la neuroética?
2. ¿Qué es el neuroderecho?
3. México a través de su jurisprudencia con relación al uso lúdico de la marihuana

## IV. Bioética y evolución: sociobiología y neuroética

## V. Mejoramiento y neuroética ¿pueden las neurociencias hacernos mejores agentes morales?

1. ¿Qué es el mejoramiento? ¿Para qué? ¿Para quiénes?
2. El uso de drogas inteligentes: ¿mejores agentes morales?
3. Ética en la Investigación de neuroimágenes
4. ¿Más inteligentes, libre de enfermedades o mejores agentes morales?

## VI. Inteligencia artificial y agencia moral

1. ¿Qué es la inteligencia artificial?

2. Inteligencia artificial aplicada
3. Inteligencia artificial fuerte
4. Superinteligencia
5. Temas relacionados con la inteligencia artificial
6. La agencia moral y las condiciones necesarias para IA

## **VII. Caso práctico para el debate grupal**

1. Intervenciones “neurológicas” no probadas

## **VIII. Fuentes consultadas**

1. Bibliohemerográficas
2. Jurisprudencia y sentencias relevantes
3. Legislación
4. Páginas Web

## I. AGRADECIMIENTOS

La producción de nuestros cuadernillos digitales está asistida por el financiamiento del Programa de Apoyo a Proyectos para la Innovación y Mejoramiento de la Enseñanza (DGAPA-PAPIME, UNAM), con clave de proyecto PE304119, titulado ‘Enseñanza Transversal en Bioética y Bioderecho: Cuadernillos Digitales de Casos’. Agradecemos al equipo de proyectos digitales del Instituto de Investigaciones Digitales por el apoyo en la digitalización del diplomado Bioética, Salud y Bioderecho, en especial a Diana Teresa Ruíz Hernández, por el apoyo en elaboración de infografías de todos los cuadernos de esta serie. De igual forma, agradecemos al departamento de publicaciones del Instituto de Investigaciones Jurídicas de la UNAM por aceptar el reto y apoyo en la edición de un proyecto editorial digital.

## II. INTRODUCCIÓN

Este cuaderno de enseñanza corresponde al módulo titulado uso de drogas inteligentes y regulación de estas como parte del Diplomado en Bioética, Salud y Bioderecho de la UNAM. La intención de integrar las reflexiones ya existentes sobre el uso de drogas inteligentes para el mejoramiento cognitivo y moral, además del uso lúdico de las mismas es contribuir a un área de creciente investigación, por tanto, se trata de presentar un breve estado del arte. En este texto también se abordará brevemente el desarrollo jurisprudencial del uso lúdico de la marihuana en México, y que se trata de una revisión importante en lo que respecta a la posibilidad de regular no solamente el uso de la marihuana, sino de otras drogas también. Además, se ha incluido el tema de la inteligencia artificial y los retos para la bioética y el bioderecho con relación a la responsabilidad moral y jurídica de agentes inteligentes, de igual manera se trata de una breve aproximación que no pretende abordar a profundidad todas las aristas que implica el estudio especializado de estos temas.

En las últimas tres décadas, la investigación en neurociencias ha explotado de manera prolífica y nuestro conocimiento del cerebro se ha incrementado inmensamente. El avance de la ciencia básica de esta área, tan pronto como se aplica en el ámbito clínico ha supuesto diversos retos tanto para la bioética y el bioderecho, pero también para los estudios sobre la ciencia y la sociedad, los desafíos y oportunidades de su aplicación son dilemáticos todos. En este texto se presentarán las nociones esenciales para conceptualizar y analizar las problemáticas que plantea el desarrollo neurocientífico y su aplicación a diversas especies, además se discuten algunos ejemplos y se plantean casos concretos para el debate grupal.

### III. LAS NEUROCIENCIAS Y EL DERECHO ¿NEURODERECHO?

A finales del siglo XX, la genética era considerada la disciplina que tenía más potencial para modificar prácticas y paradigmas éticos y legales. Se hicieron múltiples investigaciones, entre las que destacan: un estudio con gemelos en el que analizaron genes ligados al comportamiento; en el cual descubrieron el rol esencial que tienen los genes en aspectos de la psicología humana (Farah, 2012; Parens, 2004). Sin embargo, las investigaciones arrojaron datos sobre la genética de la psicología humana, que implica patrones extremadamente complejos de interacción entre los genes, y entre los genes y el entorno, lo que limitó el desarrollo de teorías e intervenciones para hacer cambios en el comportamiento de las personas (Van Gestel y Van Broeckhoven, 2003).

Lo anterior contribuyó a que más investigaciones explorarán y se entusiasmaran por las neurociencias, que son el conjunto de disciplinas que estudian la estructura, función, desarrollo, bioquímica, funcionamiento neuronal, interacción de sus elementos y la patología del sistema nervioso, que dan las bases biológicas de la conducta (Blakemore y Frith, 2011). El acercamiento a la neurociencia no inició en ese periodo, en 1848 el Doctor John Martyn Harlow documentó el caso de Phineas Gage, un trabajador de la industria ferroviaria, que tuvo un accidente en el que se golpeó la cabeza y se dañó el córtex prefrontal cerebral. Veinte años después, en 1868, el Doctor Harlow presentó el caso en la reunión anual de la *Massachusetts Medical Society* titulado *Recovery from the*

*passage of an iron bar through the head*, en el que describió los cambios de comportamiento que tuvo el trabajador después del accidente y hasta que falleció (García-Molina, 2010).

Casi un siglo después, a principios de 1980, Michael Posner y Marcus Raichle adaptaron la tomografía por emisión de positrones (*PET* por sus siglas en inglés), para realizar las primeras imágenes del cerebro, lo cual revolucionó el estudio de la psicología y las neurociencias. Además, otra técnica no invasiva de neuroimagen, que sigue siendo muy utilizada, es la Resonancia Magnética Funcional (RMNf) con la que se comenzó a obtener imágenes del interior del cerebro en las que se detectaron los cambios fisiológicos ocurridos relacionados con diferentes procesos mentales (Farah, 2012; Morse, 2017; y Álvarez-Díaz, 2013).

Entre las disciplinas que son parte de las neurociencias, se encuentran la neuroética y al neuroderecho. La relación entre ambas es muy estrecha, dado que tanto la ética como el derecho son dos ramas del conocimiento que marcan pautas de comportamiento, tanto en la esfera pública como privada. Ambas ramas establecen razones y argumentos para explicar comportarnos de ciertas formas en comunidad. La ética consiste en la justificación racional y argumentada de las acciones preferibles ante las acciones alternativas (véase cuaderno sobre bioética de esta serie, en: Hincapié Sánchez y Medina Arellano, 2019).

### 1. ¿Qué es la neuroética?

Para poder comprender a la neuroética, es importante hacer la distinción de esta por el periodo que se analiza: de 1973 a 2002, la neuroética era considerada una rama de la bioética que planteaba soluciones a los problemas éticos en el campo de la investigación neurocientífica; y del 2002 en adelante, se entiende a la neuroética como una neurociencia de la ética (Roskies, 2012 en Álvarez-Díaz, 2015).

La distinción de la neuroética con la bioética, fue propuesta por William Safire, quien fungía como Director de *The Dana Foundation* y columnista del *The New York Times*, y presentada ante un grupo de 150 personas (en donde se encontraban profesionales de diversas disciplinas, entre las neurociencias, bioética, psiquiatría, psicología, filosofía y



docentes de derecho y políticas públicas, por ejemplo) en la conferencia *Neuroethics: Mapping the Field* organizada por la Universidad de Standford y la Universidad de California en San Francisco, los días 13 y 14 de mayo de 2002.

Posteriormente, Roskies (2002) publicó un comentario en el que profundizo en la distinción conceptual que se señaló al principio de este apartado. Por un lado, la neuroética entendida como ética de la neurociencia, la divide en dos grupos:

- I. Ética de la práctica. - estudia los dilemas y consideraciones éticas que surgen en el diseño, y desarrollo de estudios neurocientíficos. Por ejemplo, se pudiese analizar cómo se garantizaría el derecho a la privacidad, bajo los principios de la bioética, de los resultados de pruebas diagnósticas para identificar enfermedades neurodegenerativas como el Alzheimer.
- II. Implicaciones éticas de la neurociencia. - evalúan el impacto social y ético de los resultados de los estudios que tienen como objetivo entender los mecanismos funcionales del cerebro y su relación con la sociedad. Por ejemplo, se elaboran los siguientes cuestionamientos: ¿si es ético el hecho de que por medio de neuroimágenes se pueda identifica cuándo una persona tiene memorias falsas y reales? ¿en cuáles circunstancias se podría usar dicha información?, y ¿qué dilemas en torno a la privacidad surgirían?

Los planteamientos y dilemas de dichos grupos pueden ser estudiados y resueltos a partir de los principios bioéticos, por lo que la primera distinción no constituye, como tal, una disciplina que no existiese anteriormente.

Por otro lado, la segunda distinción, la cual define a la neuroética como una neurociencia de la ética, implica que a través de la neuroética se estudien los cambios fisiológicos del cerebro cuando las personas elaboran razonamientos éticos. Dicha conceptualización marcó un cambio de paradigma en la ética, y se ha ido desarrollando con cierta autonomía (Álvarez-Díaz, 2013). Asimismo, la neuroética se divide en dos ramas: empírica y teórica. La primera se enfoca en el análisis de los datos neurocientíficos que se vinculan con conceptos éticos, los cuales son obtenidos a partir

del método científico. En cambio, la segunda, tiene como eje las cuestiones metodológicas que permitan relacionar hechos neurocientíficos con conceptos éticos desde dos dimensiones: la descriptiva y la normativa (Álvarez-Díaz, 2013).

Álvarez-Díaz (2015) señala que la filosofía plantea cuatro formas para elaborar juicios morales: el intelectualismo moral, el emotivismo moral, el prescriptivismo moral, y el intuicionismo moral. Para fines de este apartado, se ahondará en el intuicionismo moral, dado que ha sido el supuesto en el que se han basado gran parte de las investigaciones en neuroética. Este considera que la base de todo juicio moral es la mera conciencia moral por lo que cuando hacemos un juicio ético se lleva a cabo de forma inmediata al diferenciar lo bueno de lo malo; deja a un lado a la razón y los sentimientos. Señala que los juicios morales se basan en intuiciones, que están programadas en cada una de las personas en un plano neurobiológico. Entre las investigaciones empíricas de la neuroética más trascendentes están las aportaciones de J. Haidt (2001) y Greene (2002), quienes elaboraron una revisión sistemática, en la que identificaron las áreas cerebrales que se activan ante dilemas éticos. De acuerdo con Haidt y Greene, las áreas del cerebro implicadas en la ética son: giro frontal medial; corteza cingulada posterior, precuneal y retrosplenial; surco temporal superior y lóbulo parietal inferior; corteza orbitofrontal y corteza frontal ventromedial; polo temporal; amígdala; corteza dorsolateral prefrontal; y lóbulo parietal.

Los dilemas que la neuroética busca resolver son innumerables, pero desde sus orígenes ha tratado resolver cuestiones referentes al propio entendimiento de los seres humanos, y su desarrollo. Entre las interrogantes se encuentran: I) ¿cómo se puede entender la relación cuerpo y cerebro?; II) ¿qué es lo que distingue al ser?; III) ¿la vida está condicionada por el determinismo o existe la libertad de elección?; IV) ¿cuál es la esencia de lo moral o ético?; y V) ¿es posible que exista una ética universal? Si bien en este texto, no se les dará respuesta a dichos dilemas, si se busca que las lectoras se tomen un momento para reflexionar en lo anteriormente planteado.

Para finalizar, cabe señalar que faltan más estudios para poder realmente comprender desde la neuroética la relación de la ética con la neuroanatomía funcional, y la interacción de las redes neuronales entre sí y con otras redes. En un análisis hecho

por Szucs y Loannidis (2016) señalaron que más de la mitad de todos los estudios neurocientíficos podrían ser inválidos y no reproducibles. No obstante, no por eso se tiene que desacreditar a la neuroética más bien se tienen que redoblar esfuerzos desde la investigación para realizar estudios que puedan ser replicados en cualquier latitud del mundo. Como bien menciona Sthephen J. Morse (2018), *“la neurociencia es una ciencia puramente mecanicista. Las neuronas, las redes neuronales y el conectoma no tienen razones. No tienen aspiraciones, no tienen sentido del pasado, el presente y el futuro. Son propiedades de los agentes. La ética y el derecho y el derecho se dirigen a los agentes. Por tanto, siempre habrá un problema de traducción entre el mecanismo puro de la neurociencia y la psicología tradicional de la ética y el derecho.”* Sin duda, la comunidad académica interesada continuará explorando la neuroética y el neuroderecho para contribuir a resolver los problemas de traducción que puedan suscitarse entre la neurociencia, el bioderecho y la bioética.

### 2. ¿Qué es el neuroderecho?

El bioderecho y la neurociencia tienen una relación que comenzó a mediados del siglo XX, con el surgimiento y desarrollo de técnicas no invasivas para medir la actividad eléctrica del cerebro como el electroencefalograma (EEG), y la cirugía psiquiátrica. A la par la teoría de criminología clínica propuesta por Lombrosiano, que se basaba en que las conductas criminales tenían una explicación neurobiológica y fisiológica, en su tiempo era muy popular. La emergente relación entre el bioderecho y las neurociencias ha sido ampliamente discutida por académicas de diversas disciplinas (Vincent, 2013; Pardo y Patterson, 2013; Morse y Roskies, 2013; Shen, 2016).

No obstante, el término neuroderecho fue acuñado por Taylor, Harp y Elliott (1991) en la revista *Neuropsychology* en un trabajo conjunto e interdisciplinar titulado *«Neuropsychologists and Neurolawyers»* en 1991 con el fin de explicar el conjunto de dilemas que se suscitaban en los litigios civiles en los que había habido lesiones cerebrales temporales o permanentes (Chandler, 2018). Pero fue hasta el 2007 cuando comenzó el auge de la disciplina, en gran medida gracias a la iniciativa internacional *Law and Neuroscience Project* impulsada por la *MacArthur Foundation*.

El neuroderecho se puede definir como la disciplina que busca comprender la relación entre el cerebro y el derecho teniendo como base los avances de la neurociencia (Pardo y Patterson, 2013). El fin del neuroderecho es entender cómo la neuroanatomía puede provocar ciertos comportamientos para proponer o mejorar la regulación existente en los diversos sistemas jurídicos.

Muchos de los descubrimientos en el neuroderecho son el resultado de los experimentos hechos por las investigaciones neurocientíficas; las cuales brindan resultados para la abogacía, desde esa evidencia se apoya el neuroderecho para vincular neuroimágenes con efectos jurídicos que beneficien o perjudiquen en la menor medida posible a las personas (Petoft, 2015). Por ejemplo, a través del uso de la técnica de Resonancia Magnética Funcional (RMNf) se pueden obtener imágenes del cerebro del presunto culpable o presunta víctima, dichas neuroimágenes posteriormente pueden ser analizadas por un especialista como un psiquiatra o neurólogo para probar si es una persona neurotípica o tiene algún trastorno mental o daño cerebral que impida que pueda actuar conforme a un raciocinio que tiene como base el marco legal de la sociedad a la que pertenece; después de haberse determinado lo anterior, ahora si la persona impartidora de justicia podría determinar si tuvo o no responsabilidad civil y/o penal, solamente como ejemplo práctico de la aplicación de las neurociencias al neuroderecho.

Jennifer A. Chandler (2018) propone que el neuroderecho se pueda dividir en tres grupos y estos a su vez tienen respectivas subdivisiones para una mejor comprensión de la disciplina. La clasificación es la siguiente:

<b>Divisiones del neuroderecho</b>	<b>Subdivisiones</b>
I. Derecho de la neurociencia	i. Respuesta jurídica a intervenciones del cerebro; ii. Respuesta jurídica a lesiones o deterioro cerebral; iii. Respuesta jurídica a la obtención y uso de información relativa al cerebro.
II. Neurociencia del derecho	i. Impacto de las neurociencias en categorías y conceptos jurídicos;

	<ul style="list-style-type: none"> <li>ii. Impacto de las neurociencias en las políticas y prácticas dentro del sistema jurídico;</li> <li>iii. Entendimiento y mejora de la toma de decisiones por los actores del sistema jurídico.</li> <li>iv. Diagnóstico y detección, legalmente relevante, de distintos comportamientos y estados psicológicos.</li> </ul>
III. Cuestiones autorreflexivas y crítica de estudios jurídicos.	<ul style="list-style-type: none"> <li>i. Cuestiones o dilemas autorreflexivos;</li> <li>ii. Neuroderecho crítico.</li> </ul>

**Tabla 1.** Adaptación de las autoras de la tabla propuesta por Chandler (2018)

De la misma manera que la neuroética, el neuroderecho también tiene cierta autonomía, lo cual hace que sea una disciplina y no una rama del derecho. Las diferencias entre las neurociencias y el derecho hacen que el entendimiento y aplicación del neuroderecho sea complejo. Existen discrepancias lexicológicas, términos homófonos que pueden tener distintos significados dependiendo desde cuál de las dos disciplinas se esté analizando. En el caso, de que se haga uso de información neurocientífica como prueba, ésta debe ser precisa, razonable y bien documentada. Por ello, dichas pruebas periciales deben ser elaboradas por personas expertas en la materia que, a su vez, tengan la habilidad de plasmar sus hallazgos de forma que cualquier persona que lea el expediente pueda tener un entendimiento general de lo que representa la prueba, y las implicaciones que pudiese tener en el proceso legal en curso (Picozza, 2015).

Para concluir este apartado, es menester precisar que la neuroética normativa y el neuroderecho buscan resolver muchos de los problemas a los cuales la neurociencia ha intentado dar una respuesta (Spranger, 2012). Uno de los dilemas es la mejora de funciones normales corporales y mentales, y sus implicaciones éticas, legales, sociales, políticas y económicas. Diversos autores han intentado determinar bajo qué supuestos la mejora de ciertas funciones no debería de ocurrir. Como todo tema controversial, hay

posturas contrarias. Entre los principales oponentes está Michael Sandel (2007) y sostiene que las mejoras, *per se*, amenazan la esencia misma de lo que nos hace humanos; partiendo de la base que todas las personas son seres imperfectos. En cambio, Allen Buchanan (2011) opina que la mejora puede contribuir al desarrollo de los seres humanos siempre y cuando esté apropiadamente regulado. De igual manera existen posiciones morales que debaten aún el tema de que sí las neurociencias en sí buscan el mejoramiento genético o bien la regeneración neuroética (Chan y Harris, 2008).

El debate es complejo, ya que las mejoras físicas o cognitivas pueden impactar en diversos ámbitos de la vida de las personas. Por ejemplo, es ampliamente conocido que deportistas profesionales deben de someterse a pruebas diagnósticas de forma rutinaria para descartar el uso de sustancias como esteroides y drogas que, en algunos casos, tienen un efecto de mejora en sus funciones cerebrales para tener mejores tiempos o mayor resistencia para los entrenamientos y competencias. La lista de las sustancias prohibidas está en el sitio de la *World Anti-doping Agency* (WADA, 2021), y en el informe podemos notar que el uso de cannabinoides no está permitido (con excepción del cannabidiol), los cuales son todas aquellas sustancias químicas, independientemente de su origen o estructura, que se enlazan con los receptores cannabinoides del cuerpo y del cerebro, y que tienen efectos similares a los producidos por la planta *Cannabis Sativa L.*, conocida comúnmente como marihuana.

Dicha prohibición de los cannabinoides ha sido controversial en los últimos meses porque la velocista estadounidense Sha'Carri Richardson perdió su lugar para representar a EUA en los juegos olímpicos de Tokio 2021 tras haber dado positiva a marihuana (Beauregard, 2021). Lo cual es un poco contradictorio, ya que en el estado donde ella reside, Oregón, el consumo de marihuana es legal, y también no hay suficientes ensayos clínicos concluyentes que comprueben que dicha sustancia produce una mejora en el desempeño de las deportistas.

Por tanto, afirmamos que la neuroética y el neuroderecho, tienen un campo de oportunidad para hacer más estudios y una revisión sistémica de los efectos en el cerebro de las sustancias prohibidas por la WADA, y por ende su impacto en el comportamiento.

Para así determinar los alcances éticos y jurídicos, y proponer un marco regulatorio adecuado sustentado en evidencia científica y acorde a la realidad de muchos países, donde por lo menos, el consumo de ciertos cannabinoides ya es legal, e incluso de uso habitual como controlador de dolor, además del uso lúdico.

### 3. México a través de su jurisprudencia con relación al uso lúdico de la marihuana

El Pleno de la Suprema Corte de Justicia de la Nación (SCJN), el 28 de junio del 2021 invalidó con efectos generales los artículos 235, último párrafo, 237, 245, fracción I, 247, último párrafo y 248 de la Ley General de Salud (LGS) que establecen la prohibición absoluta para consumir cannabis y THC con fines recreativos o lúdicos al resolver la Declaratoria General de Inconstitucionalidad 1/2018 (DGI 1/2018) publicada en el Diario Oficial de la Federación el 15 de julio de 2021.

Con la DGI 1/2018 se eliminan de los últimos párrafos de los artículos 235 y 247 de la Ley General de Salud, lo que conlleva a que la Secretaría de Salud por medio de la Comisión Federal para la Protección contra Riesgos Sanitarios (COFEPRIS) pueda emitir autorizaciones para uso lúdico por personas adultas de la marihuana. Si bien dicha declaratoria representa un hito en México, sigue quedando un largo camino por recorrer, ya que la DGI 1/2018 no despenalizó la siembra, cosecha, posesión, transporte, transformación y comercialización del cannabis.

Para que la SCJN, pudiese invalidar la prohibición absoluta para consumir cannabis y THC con fines recreativos o lúdicos, se tuvo que asentar precedentes jurisdiccionales y jurisprudencia previamente. Los fallos jurisdiccionales relativos al cannabis y THC son los siguientes:

I. Amparo en Revisión 237/2014 (Caso SMART)

II. Amparo en Revisión 623/2017

III. Amparo en Revisión 1115/2017

IV. Amparo en Revisión 1163/2017

V. Amparos en Revisión 547/2018 y 548/2018.- al ser resueltos se sumaron a cinco fallos que declaran inconstitucional la prohibición de realizar actividades relacionadas con el uso personal o consumo de cannabis en México como lo son el

sembrar, cultivar, cosechar, preparar, poseer, transportar, entre otras. Lo que permitió generar jurisprudencia por reiteración el 31 de octubre de 2018.

Asimismo, con base en las siguientes ocho tesis que emanaron de cinco amparos otorgados entre 2015 y 2018, la Primera Sala de la SCJN inició el expediente de la DGI 1/2018 en el que notifico al Congreso para modificar o derogar las normas declarados inconstitucionales:

Tesis “PROHIBICIÓN ABSOLUTA DEL CONSUMO LÚDICO DE MARIHUANA. NO ES UNA MEDIDA PROPORCIONAL PARA PROTEGER LA SALUD Y EL ORDEN PÚBLICO”

Número de Registro: 2019382

Localización: [J]; 10a. Época; 1a. Sala; Semanario Judicial de la Federación; 1a./J. 9/2019 (10a.); Publicación: Viernes 22 de Febrero de 2019 10:24 h

Tesis “PROHIBICIÓN ABSOLUTA DEL CONSUMO LÚDICO DE MARIHUANA. ÉSTA PERSIGUE FINALIDADES CONSTITUCIONALMENTE VÁLIDAS.”

Número de Registro: 2019381

Localización: [J]; 10a. Época; 1a. Sala; Semanario Judicial de la Federación; 1a./J. 7/2019 (10a.); Publicación: Viernes 22 de Febrero de 2019 10:24 h

Tesis “INCONSTITUCIONALIDAD DE LA PROHIBICIÓN ABSOLUTA AL CONSUMO LÚDICO O RECREATIVO DE MARIHUANA PREVISTA POR LA LEY GENERAL DE SALUD.”

Número de Registro: 2019365

Localización: [J]; 10a. Época; 1a. Sala; Semanario Judicial de la Federación; 1a./J. 10/2019 (10a.); Publicación: Viernes 22 de Febrero de 2019 10:24 h

Tesis “DERECHOS DE TERCEROS Y ORDEN PÚBLICO. CONSTITUYEN LÍMITES EXTERNOS DEL DERECHO AL LIBRE DESARROLLO DE LA PERSONALIDAD.”

Número de Registro: 2019359

Localización: [J]; 10a. Época; 1a. Sala; Semanario Judicial de la Federación; 1a./J. 6/2019 (10a.); Publicación: Viernes 22 de Febrero de 2019 10:24 h



Tesis “DERECHO A LA PROTECCIÓN DE LA SALUD. DIMENSIONES INDIVIDUAL Y SOCIAL.”

Número de Registro: 2019358

Localización: [J]; 10a. Época; 1a. Sala; Semanario Judicial de la Federación; 1a./J. 8/2019 (10a.); Publicación: Viernes 22 de Febrero de 2019 10:24 h

Tesis “DERECHO AL LIBRE DESARROLLO DE LA PERSONALIDAD. SU DIMENSIÓN EXTERNA E INTERNA.”

Número de Registro: 2019357

Localización: [J]; 10a. Época; 1a. Sala; Semanario Judicial de la Federación; 1a./J. 4/2019 (10a.); Publicación: Viernes 22 de Febrero de 2019 10:24 h

Tesis “DERECHO AL LIBRE DESARROLLO DE LA PERSONALIDAD. LA PROHIBICIÓN PARA EL AUTOCONSUMO DE MARIHUANA CONTENIDA EN LA LEY GENERAL DE SALUD INCIDE PRIMA FACIE EN EL CONTENIDO DE DICHO DERECHO FUNDAMENTAL.”

Número de Registro: 2019356

Localización: [J]; 10a. Época; 1a. Sala; Semanario Judicial de la Federación; 1a./J. 3/2019 (10a.); Publicación: Viernes 22 de Febrero de 2019 10:24h.

Tesis “DERECHO AL LIBRE DESARROLLO DE LA PERSONALIDAD. BRINDA PROTECCIÓN A UN ÁREA RESIDUAL DE LIBERTAD QUE NO SE ENCUENTRA CUBIERTA POR LAS OTRAS LIBERTADES PÚBLICAS.”

Número de Registro: 2019355

Localización: [J]; 10a. Época; 1a. Sala; Semanario Judicial de la Federación; 1a./J. 5/2019 (10a.); Publicación: Viernes 22 de Febrero de 2019 10:24h.

Como resultado, el Congreso tendría que haber modificado la Ley General de Salud el 30 de octubre de 2019; desafortunadamente, se le venció el plazo y la SCJN le concedió tres prórrogas: una para el 15 de diciembre de 2020 y la otra para el 30 de abril de 2021. Como tampoco, el Congreso modificó la LGS, dentro de ese plazo, la SCJN fue

la que terminó emitiendo la Declaratoria General de Inconstitucionalidad 1/2018 el 28 de junio de 2021 (Ibarra y Ravelo, 2021).

Lo que sigue es que el Congreso de la Unión legisle en concordancia con el principio de justicia social para que haya un marco regulatorio *ad hoc* a los determinantes sociales del país, y que, sobre todo, deje a un lado la criminalización. El Estado tiene la responsabilidad de proteger a las personas, regulando con evidencia científica y con un enfoque bioético, feminista e intercultural.

### Un enfoque bioético sobre la legalización de la marihuana

En una sociedad democrática, se basa en un derecho inalienable de las personas adultas y competentes para tomar sus propias decisiones



Las decisiones que un individuo toma se basan en el principio bioético de libertad de autonomía



Los ciudadanos tienen derecho de elegir sus propias formas de vida, incluso aquellas que le puedan producir daño directo



El bienestar de cada uno puede promoverse socialmente mediante las instituciones del Estado pero éste no puede imponer



La posible legislación de la producción, distribución y consumo de la marihuana puede ser gradual en un proceso ordenado y regulado por el Estado



El estado mexicano tiene la obligación de advertir a la sociedad sobre los riesgos a la salud, pero no puede prohibir legalmente todo tipo de producción y distribución de la marihuana



Índica

Sativa

Ruderalis

INSTITUTO DE INVESTIGACIONES PSICOLOGICAS

Linaires, J. (2016) Un enfoque bioético sobre la legalización de la marihuana. En Animal Político. México. Recuperado de <https://www.animalpolitico.com/blogueros-una-vida-examinada-reflexiones-bioetica/2016/03/02/las-sinrazones-del-combate-a-la-marihuana/>

#### IV. BIOÉTICA Y EVOLUCIÓN: SOCIOBIOLOGÍA Y NEUROÉTICA

Como actualmente es ampliamente conocido, “La década del cerebro” fue una iniciativa que patrocinó la Biblioteca del Congreso y el Instituto Nacional de Salud Mental de los Estados Unidos. En la proclamación presidencial 6158, George Bush declaraba la década de 1990 como Década del Cerebro. Se presentó públicamente el 17 de julio de 1990 y se centró en cuatro elementos básicos: 1) el aumento y surgimiento de enfermedades cerebrales y mentales (congénitas, traumáticas y degenerativas); 2) los avances tecnológicos en microscopía y neuroimagen; 3) los avances teóricos para comprender algunos procesos fisiopatológicos, así como el desarrollo de ciencias básicas (genética, bioquímica, por ejemplo); y 4) los avances en disciplinas intermedias (entre las que destacan biología molecular, genética molecular). En el discurso también se mencionaban las adicciones. El avance fue enorme teniendo apoyo político (que se transforma en apoyo económico).

La investigación no se dirigió solamente a entender las intervenciones médicas tradicionales, tales como establecer la causa de las enfermedades neurológicas, cómo diagnosticarlas, cómo evaluar los resultados de un tratamiento, o comparar tratamientos para establecer cuál es el mejor. La investigación también quiso conocer más sobre los seres humanos. Si estaba claro desde la década de 1930 que en el cerebro radica el control motor y el registro de los procesos sensitivos, ¿por qué no estudiar otras funciones que se han entendido como específicamente humanas? Como uno de los puntos de desarrollo era la neuroimagen, se perfeccionó la imagen por resonancia magnética nuclear (fMRI), que consiste en obtener imágenes cuando el cerebro no desarrolla una tarea determinada, y luego tomar otras imágenes cuando se asigna una determinada tarea. De este modo, se investigaron actividades como la creación artística, o la elaboración de juicios éticos. Con estos dos ejemplos surgieron términos como “neuroestética” o “neuroética”.

Para la investigación en neuroética se fue consolidando una metodología. En términos simples: se tomaba la fMRI del sujeto de investigación sin hacerle una pregunta, y luego se tomaba la fMRI tras hacerle una pregunta que presentara un dilema ético. Se comparaban las imágenes, y se determinaban cuáles áreas se activaban mientras se

resolvía el dilema ético. Esto era una verdadera novedad: que la ciencia se encargara de analizar un fenómeno que tradicionalmente lo había hecho la filosofía. También se compararon grupos de personas que creían en Dios con grupos que no creían, por lo que se ha hablado de “neuroteología”; asimismo, se compararon personas budistas entrenadas para hacer meditación con grupos que no meditaban, y al no haber un dios en el budismo, se habla en este caso de “neurorreligiosidad” (Álvarez-Díaz, 2014).

Lo anterior ha representado unas ventajas claras: se sabe más sobre el cerebro, y con las posibilidades actuales de comunicación, la información se difunde rápidamente. También se han presentado problemas. Por ejemplo, al socializar el conocimiento, se ha visto que los medios de comunicación suelen interpretar los hallazgos científicos con un neuroesencialismo (algo así como “todo está en el cerebro”), y esto impacta en la generación de políticas públicas (por ejemplo, política pública en salud mental) (Racine, *et.al.*, 2010). La investigación también ha mostrado que para mejorar la comunicación de la neurociencia: 1) se requiere precaución, ya que hay períodos de mayor atención (llamados “oleadas de noticias”) en los cuales la información que se brinda tiende a ser más optimista; 2) se debe cuidar no ser demasiado optimistas en temas del desarrollo neurocientífico, a la vez de no ser demasiado escépticos en temas legislativos y de políticas; 3) los investigadores deben tener en cuenta que la precisión con que se comunica la información suele ser baja, por lo que deberían ser muy activos para evitar la difusión de conceptos erróneos (Van Atteveldt, *et.al.*, 2014). Lamentablemente esto no suele ser común, por lo que las noticias falsas o explicadas de modo incorrecto se difunden rápidamente.

Otro punto, positivo al inicio, se ha transformado en un problema. Con las nuevas tecnologías, como la fMRI, nuevas metodologías, nuevas reflexiones, se generaron palabras para demarcar terrenos que antes no existían; algunas ya mencionadas fueron neuroestética, neuroética, neuroteología o neurorreligión; también aparecieron neuroeconomía, neuroeducación, neuropolítica. A estos términos, palabras nuevas, neologismos neurológicos, Judy Illes (neurorradióloga) les denominó “neurologismos”. Sin embargo, ahora parece que las neurociencias son un manantial que ofrece legitimidad pretendidamente científica a cualquier tipo de saber solamente poniendo el

prefijo neuro: se habla de “neuro-oratoria”, “neurogastronomía”, “neurofeminismo” y un largo etcétera. Parece que se está llegando a un espacio peligroso de pseudociencia, donde el discurso neurocientífico se descontextualiza y sirve para intentar legitimar casi cualquier cosa (y además... vende). El problema es tal, que Gina Rippon propone que en muchas ocasiones se trata simplemente de “neurotonterías” (Rippon, 2019).

La neuroética tiene un problema común a otras disciplinas novedosas de esos neurologismos, y un problema específico. El problema general es que la investigación también ha mostrado que un estudio científico si se le acompaña de tablas y cuadros, tiene menos aceptación, menos credibilidad, que si se acompaña de una neuroimagen (McCabe y Castel, 2008). La neuroimagen resulta muy atractiva porque hace muy simple algo que en realidad es muy complejo, y porque de cierta forma le coloca en algún sitio. El problema específico es metodológico, en donde se hacen dos puntualizaciones: la filosofía no cuenta con suficiente conocimiento neurocientífico (¿se están dejando llevar por el atractivo de las neuroimágenes?), y la comunidad académico-científica no tiene bastantes conocimientos de bioética (cuando utilizan los términos: moral, ética, intuiciones morales, razonamientos morales, juicios morales, por ejemplo ¿quieren expresar lo mismo que la filosofía?, ¿solamente se refieren a los términos en un sentido coloquial y no especializado?).

La neuroética se ha desarrollado mucho en los últimos cinco lustros, en buena medida gracias a la neuroimagen. Hay que reflexionar mucho, ¿qué se ve en la neuroimagen? ¿Qué no se ve? Son problemas epistemológicos. Imre Lakatos escribe al inicio de su libro *La Metodología de los Programas de Investigación Científica*: “¿Qué distingue al conocimiento de la superstición, la ideología o la pseudo-ciencia? La Iglesia Católica excomulgó a los copernicanos, el Partido Comunista persiguió a los mendelianos por entender que sus doctrinas eran pseudocientíficas. La demarcación entre ciencia y pseudociencia no es un mero problema de filosofía de salón; tiene una importancia social y política vital.” La neuroética tiene que afrontar esos problemas epistemológicos y analizar las consecuencias para la disciplina y para la sociedad.

Estos retos están cada vez mejor delineados. Desde hace al menos una década se viene dejando claro que la neuroética es un campo que presenta retos epistemológicos,

metodológicos y de aplicación (como se verá más adelante). La neuroética tiene una historia reciente, aunque se monta en una larga tradición (Álvarez-Díaz, 2014a). Aunque en su nombre tiene inserto el término “ética” y podría entenderse que sigue la tradición de la vida buena, o de la vida mejor (Álvarez-Díaz, 2019), lo cierto es que ha cursado una suerte de biologización, en la cual las neuronas y el encéfalo son omnipresentes, a pesar de los siglos de tradición filosófica (Álvarez-Díaz, 2013a).

Las formas básicas de haber entendido la bioética (aunque no sea la única clasificación) han sido dos de acuerdo con Roskies: como una ética de la neurociencia (Álvarez-Díaz, 2015), y como una neurociencia de la ética. Con una ética de la neurociencia no tendría aparentemente mucha novedad, dado que se trataría de una rama de la bioética dedicada al estudio de la ética en las neurociencias, tanto en sus aspectos de investigación básica, como en la aplicación en la práctica clínica. Lo verdaderamente revolucionario parecía estar en la segunda acepción, entendiendo la neuroética como una neurociencia de la ética. Esto es, estudiar la ética entendida como una rama de las neurociencias, de modo que los supuestos epistemológicos llevarían a que se estudiara con métodos empíricos que corroboraran o falsearan hipótesis, y no con las tradicionales argumentaciones del saber filosófico. Incluso, con correlatos de otras ramas científicas basadas en datos empíricos, como las ciencias médicas.

En medicina la relación estructura-función es esencial. Si un brazo tiene una estructura y de pierde por una fractura, pierde la función. Algo así sucedería con el sistema nervioso, de modo que sería posible estudiar a personas que no tuviesen pérdidas de estructuras encefálicas y compararlas con personas que tuviesen sus estructuras intactas. Evidentemente, no sería bioéticamente posible lesionar el sistema nervioso de seres humanos para estudiar su comportamiento posterior, pero sí se podría hacer con aquellos que perdieran alguna región nerviosa por otra condición intencional (accidentes, tumores, por ejemplo) (Álvarez Díaz, 2014b). Así, los temas fundamentales para la ética, como la libertad y muchos otros, se analizaron desde una perspectiva experimental neurocientífica adicional a toda la tradicional filosófica (Álvarez-Díaz, 2014c).



Evidentemente, no existe saber o disciplina humana que se pueda fundamentar a sí misma. Ninguna; la neuroética no es una excepción. Para poder conformar el corpus teórico que ha hecho posible la neuroética, se han tenido en cuenta varios saberes por parte de quienes han realizado las aportaciones pioneras en el campo. Tales pueden ser los trabajos muy conocidos de Patricia Churchland en su libro *Neurophilosophy*, o las consideraciones de Michael Gazzaniga, quien escribe un texto monográfico dedicado al tema que ha resultado otro hito.

Para considerar que es posible realizar unos análisis neurocientíficos de la ética este par de posturas (y otras más) tienen algunas consideraciones teóricas compartidas. Si bien es cierto que existen otras referencias, la expuesta a continuación permite entender los orígenes de la neuroética. Las consideraciones teóricas inician con la animalidad de los seres humanos: en tanto que primates, poseen un encéfalo que ha evolucionado hasta ser lo que es en el *Homo sapiens*. Siendo este órgano el asiento de la ética, es posible pensar en una ética universal, tal como la denomina Gazzaniga. Aquí el inicio de las consideraciones sociobiológicas. Los seres humanos surgen en la evolución como presas en tanto que mamíferos, que deben buscar refugio y defensa de depredadores, como los grandes carnívoros. Por ello, los grupos de los seres humanos primitivos tenían que desarrollar habilidades adaptativas para poder sobrevivir. Esto hace plantear a las neurociencias que existen universales antropológicos, y con ello, universales éticos, toda vez que el encéfalo tuvo que moldearse desde esa época hasta la actual (que parece no haberse modificado grandemente).

Siendo una misma especie, con una evolución común, existirían algunos instintos morales universales, unas intuiciones morales comunes que llevan a juicios morales universales. Gazzaniga no solamente propone que existan universales éticos por compartir las mismas estructuras encefálicas, sino que va más allá y propone que hay que buscar cuál es esa ética universal. Francisco Mora neurocientífico español, propone que no solamente debe hacerse eso para la ética, sino que debe buscarse la base encefálica de la cultura para poder vivir de acuerdo con la forma como está moldeado evolutivamente el encéfalo humano. Entre sus propuestas están, por ejemplo, abandonar la vida en ciudad y regresar a la ruralidad, ya que los seres humanos primitivos vivían en

grupos pequeños de unos pocos cientos de personas. En esos grupos se cuidaban entre sí, porque cuidando a un integrante, se cuidaba al grupo. Rechazando a extraños, también se protegía el grupo.

Es posible seguir avanzando en estos planteamientos, y aparecerán más ideas que suenan, cuando menos, extrañas y reduccionistas. Hasta aquí, la filosofía, el origen de la ética, brilla por su ausencia. Por ello por lo que las reacciones a estas posturas biologicistas y reduccionistas han tenido respuestas diversas. Ante los autores citados, que pueden considerarse como neuroreduccionistas (por combinar un neuroesencialismo con un neurodeterminismo), aparecen dos respuestas. Una que puede definirse como neuroescéptica: hay autores que consideran que las neurociencias en nada pueden aportar a la dimensión característica de la ética, que es la normativa. Otra postura, que podría llamarse neurocrítica, considera que las neurociencias pueden explicar toda la ética, pero tampoco deben despreciarse porque conocer mejor al ser humano siempre ayuda; desde el “Conócete a ti mismo” en Delfos, hasta el momento actual.

Los autores neuroescépticos ni siquiera critican la sociología, asiento para los desarrollos teóricos de la neuroética, porque no lo consideran relevante. Los autores neurocríticos suelen provenir de las humanidades y las ciencias sociales, realizando análisis que critican ese sociobiologismo de base darwiniana. Sin la teoría de la evolución de Darwin como telón de fondo no sería comprensible la sociobiología, que apunta justo a explicar, como se ha esbozado, los comportamientos sociales de la especie humana desde la evolución como la entendió Charles Darwin.

Las críticas ponen a pensar; como muchas veces ocurre, quienes son criticados en ocasiones no las responden, ya que en lugar de contra argumentar, defienden los mismos supuestos de los cuales partieron. Con ello, llegan al mismo sitio: el propio. A manera de ejemplo, Gazzaniga propone que el no matar, o el cuidado de los niños, son consideraciones que han realizado todas las culturas del mundo. Dice que en ninguna de ellas se ha pensado que matar sea un acto éticamente muy bueno, o que descuidar a la niñez sea algo deseable. Sin embargo, el argumento de críticos es: cierto, pero en todas las culturas hay seres humanos que matan, y hay seres humanos que abusan de



diferentes formas a personas en situación de vulnerabilidad. Si algo está inscrito en la biología no se puede cambiar; el pH del cuerpo humano oscila normalmente entre 7.35 a 7.45, y por arriba o debajo de esos límites se cae en estados de alcalosis o de acidosis respiratoria, que, si se alejan mucho de estos valores o si no se tratan, pueden llevar a la muerte. Nadie puede siquiera sobrevivir con un pH de 3 o de 11. Si en el encéfalo estuviese inscrito el no matar, ¿por qué hay personas que matan? Por ciento, muchas. Desgraciadamente no es algo del todo extraño.

Por otra parte, si se piensa en el contexto de la niñez, UNICEF considera que en México el 60% de menores ha sufrido alguna forma de maltrato o abuso. Bajo esas consideraciones encefálicas y sociobiológicas, ¿Se habrá tenido una evolución diferente? ¿Cuáles son las ventajas adaptativas de maltratar a la niñez? Estas preguntas lucen absurdas al no considerar la cultura y con ella, los procesos socioculturales. Los análisis de los neurocríticos suelen tener alguna fundamentación filosófica, explícitamente en muchas ocasiones, o implícitamente en otras; otros suelen considerar otras ciencias como las sociales, y dar relevancia a las humanidades, como la filosofía o alguna de sus ramas. La forma como intentan relacionar los aportes neurocientíficos a lo que abonan estas disciplinas no biológicas es lo que marca muchas otras diferencias entre distintos autores.

Es difícil encontrar argumentos que analicen o cuestionen supuestos básicos compartidos. Un texto reciente intenta hacer esto: repensar la neuroética desde sus cimientos (Álvarez-Díaz, 2019a). Algo fundamental para el desarrollo de la neuroética ha sido el desarrollo de la neuroimagen; Álvarez Díaz (2019a) se pregunta en ese texto ¿qué se observa cuando se ve una neuroimagen funcional? Un punto de partida de la neuroética en tanto que neurociencia de la ética, es que puede ser posible estudiarse mediante métodos empíricos, como cualquier otra ciencia. Álvarez Díaz (2019a) formula las siguientes interrogantes: ¿Todos los neurocientíficos estarán entendiendo lo mismo por los términos “moral” o “ética” en el momento de plantear la parte experimental por analizar? ¿Todos los filósofos tienen claras las bases del funcionamiento encefálico al momento de proponer soluciones? En la literatura neuroética, ¿Se entiende lo mismo cuando se habla del “cerebro” que cuando se habla de la “mente”? Estos

cuestionamientos son analizados por Álvarez Díaz (2019a), junto a otro que es el motivo de esta sección: la sociobiología.

Varios autores han hecho explícita esta relación entre neurobiología como consecuencia sociobiológica de la evolución en sentido darwiniano. Álvarez Díaz (2019a) hace otra serie de cuestionamientos a este tipo de evolucionismo. No es que lo niegue, de ninguna manera, pero establece limitaciones a sus alcances. Si absolutamente todo puede explicarlo el darwinismo, sería como haber cambiado la teología medieval por un nuevo dogma. Todas las teorías científicas tienen alcances y limitaciones. Y es que precisamente en el problema complejo de la relación entre el binomio mente/cerebro con el de moral/ética es uno de los varios temas que no pueden explicarse del todo mediante el evolucionismo darwiniano.

El hueco está en que hay elementos que podrías explicarse no desde la genética (y su correlato con la sociobiología y el darwinismo), sino desde factores epigenéticos. No como a veces se entienden de modo excesivamente amplio y saltan desde los genes hasta la cultura. Epigenético se refiere a elementos diversos que regulan la expresión del genoma. Y algunos pueden encontrarse en el citoplasma de la misma célula. Esto es lo que podría originar herencia transgeneracional; aunque no lo enunció así Lamarck, se le suele atribuir la herencia “de los caracteres adquiridos”. Se ha estudiado en varias especies; recientemente ha aparecido una gruesa revisión sobre estos temas (Tollefsbol, 2019), posterior a la aparición del libro de Álvarez Díaz. Es cierto que hay complicaciones metodológicas para el trabajo con seres humanos, además de otras de tipo bioético que impiden hacer estudios prospectivos, que son los que aportan más evidencia. Sin embargo, con las investigaciones que son posible realizar, hay datos para empezar este análisis de forma sistemática y seria. De hecho, los análisis contemporáneos a propósito de las neurociencias tienen claro que el mismo Darwin no solamente conoció los planteamientos de Lamarck, sino que los compartió, y que de alguna manera los amplió dentro de su teoría evolucionista (Han, 2019).

Todas estas reflexiones sobre la Sociobiología y el darwinismo, con sus aportes y limitaciones, aunados a las propuestas Lamarckianas, son relevantes no solamente para la neuroética, sino para algunas otras interfaces. Por ejemplo, la interfaz entre neuroética

(entendida como neurociencia de la ética) y la bioética. Lamentablemente no se presta mucha atención en la actualidad a la gestación de la bioética, pero en sus orígenes están instituciones como el Hastings Center, fundado por Daniel Callahan, por desgracia fallecido recientemente, y por Willard Gaylin, nacido en 1925. Gaylin es un psiquiatra y psicoanalista; el psicoanálisis tiene un papel importante en el inicio de la bioética (Álvarez-Díaz, 2015a). Existen autores con orientación psicoanalítica que aún antes de los avances biológicos recientes en epigenética ya consideraban que la ética podría tener características Lamarckianas (De Meira Penna, 1985), existiendo otras corrientes de pensamiento contemporáneos que han afirmado que efectivamente podría ser así (Rensma, 2013).

El concepto de dignidad es complejo, pero necesario para la ética (Álvarez-Díaz, 2014d) y el derecho (por ejemplo, como un modo de fundamentar los derechos humanos). Gaylin es un ejemplo de reunir todo lo comentado hasta el momento. Para él, el concepto de dignidad humana puede entenderse hasta el medioevo como una creencia religiosa de que el ser humano posee una propiedad exclusiva por haber sido creado a imagen y semejanza de un Dios; filosóficamente y tras la secularización ilustrada, Kant consideraría que la autonomía otorga un valor peculiar al ser humano. Gaylin consideró que los seres humanos podrían ser peculiares para la ética por al menos cinco atributos: el pensamiento conceptual, la capacidad tecnológica, una gama amplia de emociones con su consecuencia en los sentimientos, lo que denominó como “genética ambiental lamarckiana” (la epigenética, entendida más bien a nivel cultural), y la libertad que tienen los seres humanos de cambiarse y modificarse a sí mismos (Gaylin, 1984). Como queda claro, el problema no queda resuelto del todo (como muchos otros de corte filosófico).

Otra de las interfaces relevantes entre este análisis crítico de la Sociobiología darwiniana a través de la contemporánea epigenética y sus postulados que muy bien pueden considerarse como neolamarckianos, es precisamente con el derecho. Hace apenas un lustro una revista dedicaba un número monográfico al tema. Aunque es muy reciente, aclaraba que hablar sobre epigenética podría resultar casi esotérico antes que práctico, ya que había que incorporar un léxico nuevo, con términos y conceptos

científicos dentro de un marco de mecanismos dinámicos novedosos para muchos. Se preguntaba abiertamente sobre la relevancia de la epigenética para el trabajo especializado (la revista está enfocada en las ciencias del comportamiento y la ley) (Gunter y Felthous, 2015). El volumen inicia con un ensayo general, prácticamente a modo de estado del arte (Gunter, 2015). Entienden que la epigenética no es una “ciencia más” con relevancia legal, sino que será una contribución decisiva en temas esenciales tales como: la ética clínica y forense, la comprensión de la naturaleza y causas del funcionamiento psicológico necesario para competencias civiles y penales, así como la comprensión de la naturaleza y causalidad de variedad de traumas psíquicos.

Considerando que la ley respalda prácticas éticas, y que las políticas públicas que consideren información epigenética se desarrollarán en el futuro a medida que surjan nuevas necesidades, se dedican artículos a la problemática ética en el campo de niñez (Thomas, 2015), protección de datos epigenéticos en salud mental (Terry, 2015), así como pruebas epigenéticas de uso de drogas legales e ilegales (Erwin, 2015). Otro trabajo presenta explícitamente la herencia epigenética transgeneracional, advirtiendo que no deben darse saltos en la investigación. Esto es, si hay datos en animales no humanos, deben corroborarse en seres humanos, ya que no es adecuado aplicar o modificar cuestiones legales por observaciones realizadas solamente en animales no humanos. Al mismo tiempo, abre una posibilidad futura de establecer si es posible la transmisión epigenética de rasgos o síntomas causados por conductas criminales o negligentes de una o dos generaciones previas, lo que plantearía preguntas interesantes sobre la proliferación de víctimas (DeLisi y Vaughn, 2015).

Siguiendo esta línea que une la investigación epigenética y sus saberes con las consecuencias de estos, otro trabajo enfatiza las preguntas que surgen por temas esenciales para el derecho, tales como la libertad y la responsabilidad. La ley no debería exigir que los seres humanos hagan algo que no pueden; no deberían socavarse principios fundamentales necesarios para un sistema jurídico pragmático y eficaz. Por ello, ofrecen una advertencia sobre la biologización de rasgos problemáticos de la personalidad hasta el punto de que se puedan ignorar factores psicológicos, sociales y culturales (Tamatea, 2015).

Dado que es relevante esa relación entre epigenética y derecho, sin que un saber quede reducido o subsumido en el otro, resulta importante saber que el perfil epigenético puede decir cosas sobre el comportamiento actual con relación a la exposición de sustancias de abuso y tóxicas. Es importante conocer el impacto del tabaquismo agudo y crónico (Philibert y Erwin, 2015), así como los mecanismos epigenéticos de la metilación en muestras poblacionales y las correcciones que deben hacerse a los hallazgos comunitarios (Gunter y Philibert, 2015).

Desde la definición de salud de la OMS, que la considera como el estado de completo bienestar físico, mental y social y no solamente la ausencia de enfermedad o invalidez, han existido grandes preguntas. Conceptualmente es importante el avance de la definición para no reducir la salud solamente a aspectos biológicos: se establecen bien los límites, aportes, pero las interacciones entre estas tres esferas quedan algo desdibujadas. La epigenética proporciona elementos para seguir pensando en esa relación entre “naturaleza vs. cultura”. No existe en la actualidad una teoría unificadora de la mente que sea universalmente aceptada; otro trabajo en ese número monográfico examina el impacto del trauma vivido en situación de guerra y un posible marcador epigenético del trastorno de estrés postraumático (Ramo-Fernández, *et.al.*, 2015).

Estas investigaciones, como queda claro, no se limitan al mero conocimiento epigenético, sino que abren discusiones para determinar en el futuro si un análisis epigenético es evidencia admisible o no lo es. La admisibilidad de la evidencia forense debe ser monitoreada por expertos y las ciencias del comportamiento, a medida que la ciencia y sus aplicaciones continúan desarrollándose.

### V. MEJORAMIENTO Y NEUROÉTICA ¿PUEDEN LAS NEUROCIENCIAS HACERNOS MEJORES AGENTES MORALES?

#### 1. ¿Qué es el mejoramiento? ¿Para qué? ¿Para quiénes?

La siguiente conversación no es hipotética; tuvo lugar en una red social con estudiantes de una universidad pública mexicana:

Estudiante 1: ¿Alguien que consiga Ritalin® a buen precio por aquí?

Estudiante 2: ¿Qué es eso compañero?

Estudiante 3: Una pastilla mágica que te quita lo pendejo, como en la peli de *Sin límites*.

Estudiante 4: ¿Es de la que dejó tartamudo a Porky?

Estudiante 1: Es la que hizo que Porky hablara bien.

Desde luego, hubo más comentarios. ¿De qué estaban hablando? El Ritalin® es un nombre comercial de un fármaco, el metilfenidato [2-fenil-2-(piperidin-2-il)etanoato de metilo]. Si bien fue patentado desde la década de 1950 teniendo como indicaciones trastornos del sueño (actualmente está indicado en narcolepsia y en hipersomnia primaria), su uso creció cuando se habló de la “disfunción cerebral mínima” o “hiperactividad”, que se transformó en “trastorno por déficit de atención” (TDA, que puede ser con o sin hiperactividad; en inglés es el *attention deficit disorder* o ADD). Uno de los problemas esenciales (no el único) que tiene una niña o niño con TDA es que una función mental superior, tal como la atención, no es posible mantenerla del modo adecuado y por el tiempo necesario para tener una vida escolar y social promedio. Un efecto tras consumir el fármaco es que se aumenta la atención (no es el único; hay muchos otros, por ejemplo, tiene un efecto anorexígeno, lo que quiere decir que quita el apetito, las ganas de comer). La cultura popular ha construido la misma asociación que se hizo en la época de los antidepresivos: si un menor a quien le va mal en la escuela porque no pone atención adecuadamente ni por el tiempo necesario se le puede indicar el medicamento, ¿qué sucede? Mejora el rendimiento escolar. Está bastante estudiado que si el diagnóstico es adecuado y el manejo es completo hay una franca mejoría (no se trata solamente de administrar el fármaco, ya que debe ir acompañado de aspectos conductuales, escolares, sociales, por ejemplo.). Sin embargo, ¿qué sucedería si un estudiante (adulto joven) quiere mejorar su rendimiento escolar y toma el medicamento? La tendencia es a creer que “mejora”. La idea de la mejora es esencial. No se trata de un “tratamiento”, sino de que, al no tener un diagnóstico o un problema atencional, se haría uso del fármaco por una persona que en principio es sana (como muchas personas en la etapa estudiantil). Por ello se habla de “mejorar”, ya sea el ánimo con el caso de los antidepresivos, o la atención con este tipo de fármacos (Chan y Harris, 2018).

Lo primero que hay que preguntarse es si esto es así o no. Puede haber muchas hipótesis, pero la ciencia intenta probarlas para determinar si son ciertas o no, si se pueden falsar o no. Para hacer ciencia se requiere de observaciones, teorías, y en general, conocimientos previos. La ciencia no parte de la nada; la mejor prueba en el mundo contemporáneo es que antes de hacer cualquier investigación deben buscarse los antecedentes y elaborar un estado del arte sobre el problema a tratar. Se debe tener claro qué es lo que se sabe para determinar qué es lo que no se sabe.

En el caso de las aplicaciones a personas humanas, para las clásicas intervenciones médicas de diagnóstico o tratamiento, se requiere investigación con seres humanos. De hecho, metodológicamente se parte del supuesto de que la intervención (en este caso el fármaco) no sirve, no tiene actividad. Si esto es lo que debe mostrarse tras la investigación, se toma criterios lógicos y matemáticos: se postula una hipótesis nula y una hipótesis alternativa. La hipótesis nula plantea que no se tiene un determinado efecto, y la alternativa que sí se tiene; matemáticamente se obtienen datos para negar el postulado de la hipótesis nula, de modo que la negación de una negación da como resultado una afirmación.

Para postular que un fármaco sirve (o no) como tratamiento de alguna condición, debe realizarse una investigación que se denomina como “ensayo clínico”. Hay muchos tipos de investigación, pero en medicina la más utilizada con propósitos de saber cuál es el tratamiento para una enfermedad (o una prueba diagnóstica), tiene que hacerse investigación mediante ensayo clínico. Sí existen ensayos clínicos donde hayan participado pacientes con diagnóstico de TDA, en los cuales se muestra que el metilfenidato es un fármaco que puede ayudar mucho en el tratamiento (como ya se indicó, no es la única intervención). Sin embargo, no hay ensayos clínicos con estudiantes sanos (al menos, sin TDA) de donde pudiera concluirse que el metilfenidato mejore su rendimiento escolar.

Se han hecho algunas investigaciones con uso de metilfenidato y algunos otros fármacos para explorar si efectivamente tienen un efecto objetivamente medible en el aprendizaje o en la cognición humana. Los resultados han sido verdaderamente decepcionantes. Por ejemplo, se sabe que se consumen de una forma más común de lo



que se supone (Partridge, *et.al.*, 2011), y que una de las motivaciones principales es el estrés asociado a una sobrecarga de trabajo (Eickenhorst, *et.al.*, 2012). Se sabe que los estudiantes recurren tanto a fármacos que son considerados drogas lícitas como a drogas ilícitas (Maier, *et.al.*, 2013).

Si se analizan algunas investigaciones con más detalle, los resultados no solamente son alentadores: deberían ser disuasivos. Una investigación tuvo como objetivo evaluar el efecto de la administración aguda de dosis variables de metilfenidato (10 mg, 20 mg, 40 mg y placebo) en una amplia gama de funciones cognitivas en jóvenes sanos. Son dosis terapéuticas en caso de necesitar el medicamento. Participaron 36 jóvenes universitarios y graduados; se sometieron a pruebas objetivas para evaluar atención y memoria episódica y funcional. En los resultados no se observaron diferencias en el rendimiento en ninguna de las pruebas. Hubo un efecto dependiente de la dosis (40 mg > placebo) sobre el bienestar autoinformado, pero no en el objetivamente demostrado por las pruebas. De acuerdo con la literatura reciente, los medicamentos como el metilfenidato, mejoran el rendimiento cuando los procesos cognitivos están por debajo de un nivel óptimo (en caso de un trastorno de la atención), lo que no fue el caso para los sujetos del presente estudio (Batistela, *et.al.*, 2016).

Otra investigación tuvo por objetivo determinar si el rendimiento en el ajedrez puede mejorarse con metilfenidato, modafinilo o cafeína. El modafinilo es otro fármaco que se utiliza en el tratamiento del déficit de atención. La cafeína es un estimulante que cualquiera que ha deseado dormir menos y acostarse un poco más tarde habrá escuchado y hasta normalizado. Los métodos científicos de este estudio fueron robustos, ya que se diseñó con metodología de ensayo clínico en fase IV (fármaco vigilancia), aleatorizado, doble ciego, y controlado con placebo. Participaron 39 jugadores de ajedrez, varones, quienes recibieron dos tabletas de 200mg de modafinilo, o dos tabletas de 20mg de metilfenidato, dos tabletas de 200mg de cafeína o bien dos tabletas con placebo (al ser doble ciego, ni quien administraba ni quien recibía las tabletas conocía su contenido). El diseño fue cruzado, de modo que en un momento del tiempo quienes recibían una intervención se cambiaba para recibir otra. Jugaron veinte juegos de 15 minutos durante dos sesiones contra un programa de ajedrez (Fritz 12) y completaron



varias pruebas neuropsicológicas; se esta manera se puede estandarizar tanto el estímulo de jugador como el desempeño neuropsicológico de modo objetivo. Los resultados muestran que los tres fármacos aumentaron significativamente el tiempo promedio de reflexión por juego en comparación con el placebo (mayor número de juegos perdidos a tiempo con los tres tratamientos.) Los efectos del tratamiento no se observaron si se analizaran todos los juegos (n = 3059). Se concluyó que conjuntamente con los resultados de las pruebas neuropsicológicas, los efectos modificadores de los estimulantes en tareas cognitivas complejas pueden resultar en particular de procesos de toma de decisiones más reflexivos. Cuando no está bajo presión de tiempo, tales efectos pueden dar como resultado un mejor rendimiento. Sin embargo, debido a limitaciones de tiempo, una toma de decisiones más reflexiva puede no mejorar o incluso tener efectos perjudiciales sobre el rendimiento de la tarea compleja (Franke *et al.*, 2017).

Otra investigación reciente tuvo como objetivo el revisar prácticas recientes de dopaje cognitivo farmacéutico y la relación riesgo/beneficio de dicho comportamiento y discutir temas relacionados. Los resultados mostraron un cambio generacional importante: los estudiantes actuales son cada vez más permisivos en su percepción a la toma de fármacos; la prevalencia del dopaje farmacológico entre estudiantes se ha informado del 1,3% al 33% en todos los estudios (variaciones según país y definición de dopaje; la prevalencia, por cierto, suele ser más alta entre profesionales de la salud). Las clases terapéuticas citadas con más frecuencia con fines de dopaje son psicoestimulantes y fármacos que se consideran nootrópicos aunque no haya evidencia que los catalogue realmente como tales (metilfenidato, modafinilo, piracetam), corticosteroides, sedantes y betabloqueantes. Algunas sustancias ilegales como el cannabis (en algunos países), las anfetaminas y la cocaína también se consumen. En productos de venta libre se reportan las tabletas a base de cafeína, las bebidas energéticas, el alcohol, e incluso combinaciones entre varias de estas categorías. El estudio concluyó que la efectividad esperada de estos productos a menudo es fantaseada, y no corresponde a la realidad de un impacto siquiera modesto en el rendimiento cognitivo. Parece que actualmente hay un perfil emergente de un estudiante más inclinado al comportamiento de dopaje (Carton *e.al.*, 2018).

Otra investigación reciente tuvo como objetivo examinar la frecuencia del uso definido como “no médico” (de dopaje, con pretensiones de mejora) de estimulantes recetados (anfetaminas y metilfenidato) en una muestra de universitarios y sus correlatos asociados de salud física y mental. Los métodos incluyeron una encuesta anónima en línea de 156 ítems enviada por correo electrónico a 9449 estudiantes universitarios. Se evaluó uso de alcohol y drogas, estado psicológico y físico y rendimiento académico, junto con medidas de impulsividad. Participaron 3421 estudiantes (59,7% mujeres); 6.7% de la muestra informó uso actual/reciente no médico de estimulantes recetados; 5.8% adicional informó abuso en el pasado (lo cual es indicativo, además, que hay médicos que están de acuerdo en extender una receta por este tipo de fármacos sin que exista un diagnóstico en quienes los van a consumir). El uso no médico de fármacos recetados se encontró asociado con promedios más bajos, con la toma de otras drogas (alcohol, nicotina, sustancias ilícitas y consumo de refrescos con cafeína), una impulsividad mayor (evaluada con la escala de Barratt), tratamiento previo para problemas de uso de sustancias, aparición elevada de juegos de azar, trastorno de estrés postraumático y ansiedad. También de modo preocupante, fueron significativamente más activos sexualmente (incluso a una edad más temprana), y tenían menos probabilidades de usar métodos de barrera. No se asociaron síntomas de depresión o trastorno por atracón compulsivo (aunque se asoció con el uso de medicamentos para perder peso) (Grant, et.al., 2018). Desde luego los calificativos no son moralina: las investigaciones muestran que no hay ningún beneficio, y sí muestran perjuicios objetivamente, como estas asociaciones, aunque subjetivamente los estudiantes se sientan tranquilos y mejor con el consumo de estos fármacos.

Como puede apreciarse, la investigación existente muestra datos preocupantes. Si el asunto del neuromejoramiento es un problema de neuroética, y los problemas en ética tienen que ver con lo que debe hacerse o no, ¿debería seguir buscándose la mejora por esta vía, es decir, farmacológica, con los medicamentos disponibles en la actualidad? ¿Es tiempo de empezar a diseñar nuevos fármacos? ¿Para quiénes serían estos fármacos? ¿Los podría adquirir cualquier persona de la sociedad o solamente las élites económicas? Estos cuestionamientos tienen respuestas varias, pero no constituyen aún

un problema para el momento actual. Es relevante analizarlos desde el punto de vista teórico, sirven para pensar y ordenar el futuro. Son importantes, pero tal vez no urgentes.

Probablemente sería más urgente pensar en la intervención del derecho ante la realidad: si estos fármacos no mejoran lo que se cree, sino que, por el contrario, afectan o se relacionan con conductas de riesgo, ¿qué debería hacer el derecho? ¿regularlo? ¿prohibirlo? A propósito de esta serie de cuadernillos de enseñanza transversal, ¿qué deberían hacer docentes en educación superior ante este hecho? En principio deberían informarse y comentarlo con sus estudiantes, reseñarles que no solamente no les benefician estos fármacos, sino que contrariamente a lo que creen, les puede perjudicar. También puede analizarse la temática en grupos de estudio especializados. En la UAM Xochimilco se llevó a cabo por varios años un Seminario de Neurofilosofía, coordinado por Sonia Ramírez Zapata; en la UNAM existe un Seminario de Neuroética, coordinado por Pedro Enrique García Ruiz. En ambos se ha tratado el tema del neuromejoramiento. Sin embargo, si el conocimiento y la discusión no sale más allá de estos espacios multidisciplinarios y especializados, los estudiantes pueden estar sometiéndose a riesgos innecesarios, sin saberlo, y creyendo que están haciendo algo que les va a ayudar. ¿Quién se beneficia entonces de la falacia de que el metilfenidato puede ayudar a mejorar el rendimiento cognitivo de los estudiantes? Sin duda la industria farmacéutica: una caja de metilfenidato con 30 cápsulas de 20 mg cuesta un poco más de 1200 pesos mexicanos (al momento de escribir este texto, aproximadamente 50 dólares norteamericanos). Para ponerse a pensar.

### *2. El uso de drogas inteligentes: ¿mejores agentes morales?*

La terminología que se ha utilizado para ir más allá del campo de la atención y el aprendizaje no es uniforme, y, además, exhibe algunos supuestos cuando se le propone. El término más utilizado probablemente es el de “neuroenhancement”, que suele traducirse por neuromejoramiento (o neurorealce). También se le ha llamado “cosmetic neurology”; esto tiene algunos supuestos, por ejemplo, que se trata de algo meramente estético, no real, en el campo neurológico. Incluso se ha hablado de “brain engineering”, que tiene los supuestos de que “encéfalo” y “mente” o “funciones mentales” son lo mismo:

si se interviene el encéfalo, en consecuencia, se obtienen cambios mentales, cognitivos o afectivos. Incluso se ha propuesto que no se hable de “mejoramiento humano” sino de “modificaciones corporales” (Rembold, 2014), si finalmente lo que se interviene es el cuerpo o una parte de él como el encéfalo. Ante estas propuestas, a veces convergentes y a veces divergentes, se continúa con el término de “neuromejoramiento” dado que es el que ha difundido más en lengua española.

Por otra parte, la forma de conseguir tal mejora tampoco es la misma, aunque en general se trata de cuestiones más bien teóricas. Aunque hay algunas aplicaciones, de ninguna manera son todavía de uso y/o acceso generalizado. Las intervenciones podrían ser neurotécnicas (no farmacológicas), a través de la estimulación encefálica profunda (“*Deep brain stimulation*”) (Pacholczyk, 2011), con interfaces encéfalo computadora (*brain-computer interfaces*), o bien con intervenciones neurobiónicas (*neurobionic interventions*). Esto abre panoramas con otros campos, por ejemplo, el nanotecnológico, para poder pensar en implantar chips nanométricos que pudieran intervenir y modificar funciones encefálicas, ya sea en caso de enfermedad o para mejora (Álvarez-Díaz, 2018).

Mucho de lo que se ha teorizado ha sido con fármacos, probablemente como consecuencia de lo ya expuesto previamente. Por ello, las intervenciones farmacológicas conducirían a diferentes tipos (teóricos) de mejoras: sensorial (“*sensory enhancement*”), del desempeño (“*performance enhancement*”), en habilidades cognitivas (“*cognitive enhancement*”), en el estado de ánimo (“*mood enhancement*”), y el tema de esta sección, la posibilidad de una mejora en el campo de la moral (“*moral enhancement*”).

La literatura de ficción, en ocasiones llevada al cine, puede ver el interés de la cultura del siglo XX por estas posibilidades. Puede apelarse a la novela de Aldous Huxley: *Brave New World*. Una novela distópica con el uso del “soma”, un fármaco necesario para el control de ese nuevo orden mundial, llevada a miniserie de televisión. También destaca una novela de Frank Herbert: *Dune*, llevada al cine con el mismo título. Futurismo de ficción, plantea que en un planeta llamado Arrakis existe una potente droga, “*melange*”, que incluso se utiliza para vuelos especiales, pero que tiene el efecto de prolongar la vida humana, por lo que es altamente codiciada individual y colectivamente. No podría

quedar fuera *A clockwork orange*, de Anthony Burgees; llevada al cine por Stanley Kubrick muestra el uso del “*moloko plus*” (“*knifey moloko*” o “*moloko drengo*”) que preparaba a las personas para una violencia extrema. Podrían nombrarse muchas más, pero se termina este recordatorio con *The dark fields*, novela de Alan Glynn que se llevó al cine bajo el título de *Limitless*. La novela plantea el uso de una droga, MDT-48, o NZT-48 en la película, que es capaz de producir una mejora cognitiva única. ¿Interés, deseo, necesidad o capricho de contar con fármacos de este tipo en el mundo no ficticio?

Los planteamientos de modificación para mejoramiento no se han dado solamente en la ficción. Corneliu E. Giurgea fue un rumano con una formación académica peculiar: psicólogo y químico. En 1964 consiguió la síntesis del piracetam, y en 1972 acuñó el término “nootrópico” (Giurgea, 1972). Bajo este nombre se designaría un grupo de fármacos que, bajo los criterios que propone Giurgea, deberían tener las siguientes características para ser considerados tales:

1. Deben favorecer el aprendizaje y la memoria.
2. Deben mejorar la resistencia de las memorias y habilidades aprendidas frente a aquellos agentes que puedan perturbarlas (hipoxia y electrochoques, por ejemplo).
3. Deben proteger al cerebro frente a daños físicos o químicos (barbitúricos, escopolamina).
4. Deben incrementar la eficacia de los mecanismos de control corticales y subcorticales.
5. Deben carecer de la farmacología usual de las drogas psicotrópicas (sedación o estimulación motora), deben tener muy pocos efectos secundarios y una toxicidad extremadamente baja.

Sabiendo que la evidencia que muestra la investigación científica indica que no hay en realidad fármacos nootrópicos, ya que los existentes tienen utilidad solamente clínica, es decir, en pacientes, ¿qué decir a propósito de intervenciones tecnológicas o farmacéuticas para conseguir mejoras morales? Todo lo que se pueda decir será teórico especulativo; cuando se avance en ciencia y tecnología lo teórico podría ponerse en

práctica en la investigación para obtener nuevo conocimiento. Mientras tanto, alguna propuesta.

Si existiera un nootrópico y se le administrara a un ser humano, ¿debería usarse? ¿quién debería usarlo? Si es para “mejorar”, ¿qué quiere decir “mejor”? ¿Mejor que quién? ¿Mejor para qué? ¿Mejor para quién? ¿Necesariamente una persona con mejores capacidades cognitivas es “mejor”? La idea misma de “mejora” es retadora. Hay quienes responden que no debería utilizarse, aunque existiera. Pero hay quienes consideran que sí. En el caso de que existiera y se utilizara, ¿se podría hablar de “otro tipo” de seres humanos? ¿“Mejorados” y “no mejorados”? ¿Serían más humanos o menos humanos? Más aún, ¿seguirían siendo humanos? Esto abre otras discusiones como el transhumanismo y el posthumanismo, que solamente se mencionan y por motivos de desarrollo del tema y de espacio no se avanza en esos temas. Sin duda, son parte del reto a futuro.

Además, existen otros autores que no solamente consideran que, de existir una posibilidad de mejoramiento moral debería utilizarse, sino que lo consideran propiamente como una obligación moral. Tal es el caso de Persson y Savulescu en su libro *Unfit for the future. The need for moral enhancement* (2012). Resulta ya de por sí interesante que el verbo que utiliza Darwin para hablar de la adaptación es precisamente *to fit*. Esto es, “ajustarse”: una especie está ajustada al medio o no lo está; tal como la llave a la cerradura, se ajusta o no se ajusta. Este par de autores analiza sobre todo los retos del mundo tecnológico contemporáneo y dicen que “... los seres humanos no están, por naturaleza, equipados con una psicología moral que les permita hacer frente a los problemas morales que crean estas nuevas condiciones de vida”.

El tema de la técnica y la tecnología, así como su relación con la ciencia, acompañan el pensamiento filosófico del siglo XX. José Ortega y Gasset escribe su *Meditación de la técnica* (un título con ese lenguaje que caracteriza a Ortega, en ocasiones casi ascético); Martin Heidegger escribe *La pregunta por la técnica*; (donde dice, entre otras cosas, que cuando los seres humanos no cuestionan la técnica es porque se encuentran ya arrojados a ella); siguiendo con la tradición alemana, un discípulo de Heidegger, Hans Jonas, es célebre en bioética por su libro *El principio de responsabilidad*. Ha existido una

creciente preocupación por los resultados de la aplicación de lo que terminó por llamarse tecnociencia, dado que no es posible disociarlas o entender una sin la otra.

Persson y Savulescu (2012) identifican amenazas del mundo tecnológico. Una es las armas de destrucción masiva, sobre todo si caen en manos de grupos terroristas. Otra sería el cambio climático y sus efectos a nivel global. Ambas habrían sobrepasado las capacidades morales reales actuales de los seres humanos. Las democracias liberales y las economías emergentes pueden enfrentar el primer problema con una restricción a las libertades individuales y una vigilancia intrusiva de todas las comunicaciones y actividades de las personas. Ante el segundo problema los seres humanos definitivamente no están preparados: no han sido capaces de poner límites a las libertades individuales de consumo supeditadas al crecimiento constante de la producción industrial y del comercio, que se basa en un alto consumo de energía y recursos naturales. Peor, parece ser que los seres humanos han causado este problema y lo han empeorado.

Para enfrentar esos problemas, Persson y Savulescu consideran que las personas del mundo deberían aceptar por voluntad propia una transformación del paradigma moral de las acciones mediante necesarias medidas precautorias y restrictivas del consumo y de las libertades individuales de sociedades liberales antes que sea demasiado tarde. El tiempo es crucial en el momento actual, en que parece ser que ya se acabó. Estos autores consideran que el poder tecnológico crece más rápido que la responsabilidad moral colectiva de los seres humanos. Resulta por ello urgente que los seres humanos tengan igual consideración moral por intereses y derechos propios, así como por los intereses y derechos de generaciones futuras y del resto de animales con quienes comparten ya el planeta. Recordando a Peter Singer y su consideración de la moralidad en círculos, habría que ir más allá de la tradicional moralidad egocéntricos, etnocéntricos y antropocéntricos (Figura 1) que suelen caracterizar a las culturas a lo largo de la historia. Por cierto, esto ya se ha intentado mediante los métodos tradicionales de la educación de la filosofía moral. De ahí, dicen, la urgencia de la mejora moral humana.



## Posibles centros de referencia moral

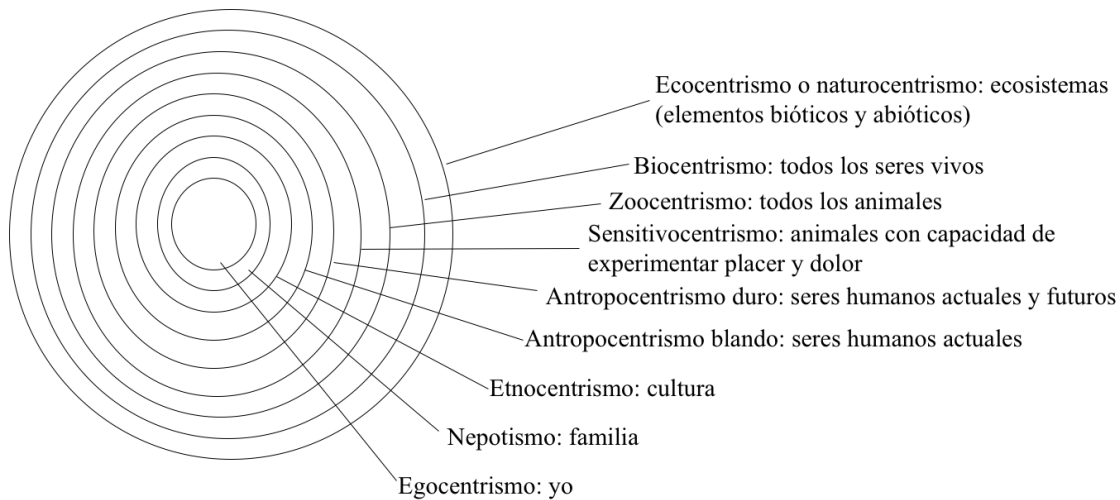


Figura 1.

Persson y Savulescu (2012) consideran que las limitaciones de la moralidad son muchas: es más fácil dañar a otros que beneficiarlos, el “sentido natural de responsabilidad” de los seres humanos estaría reducido a un corto alcance (la causalidad de efectos visibles entre los cercanos), las éticas han buscado más no dañar que beneficiar (probablemente con la excepción de Kant), la temporalidad es el corto plazo (no parece haber una verdadera planeación real a largo plazo), y la espacialidad es lo cercano. Como puede verse, van mucho en la línea sociobiológica darwiniana ya analizada.

La postura de Persson y Savulescu (2012) es que la tecnología genética y las neurociencias deberían perfeccionarse para poder conseguir ese objetivo de mejorar a los seres humanos. Lo denominan como “*moral bio-enhancement*”, ya que la apuesta es por la mejora directa a la biología humana. Aunque consideran que no sería una sustitución de la educación moral, lo entienden como un necesario complemento que debería hacerse con toda la humanidad (con lo que prácticamente se eliminan las críticas de que solamente estaría disponible para algunas personas). Para que efectivamente pudiera realizarse y mantenerse, habría que combinar terapia génica, ingeniería genética



(somática y germinal), neurofármacos, neurocéticos (administrados en la comida), estimulación cerebral transcraneana, por ejemplo. Esto sería un esquema para el “biomejoramiento moral”.

Persson y Savulescu (2012) consideran que la solución no sería exclusivamente tecnológica. Efectivamente debería continuarse con la internalización de motivaciones y razones morales para expandir responsabilidad individual y colectiva hacia seres humanos distantes, hacia otros animales, y hacia futuras generaciones de seres vivientes. El biomejoramiento moral lo consideran necesario, para modificar las emociones y así aumentar o potenciar simpatía y empatía (disminución de agresividad, alteridad positiva, sentido de la justicia como imparcialidad y como equidad). Consideran que habría que modificar la concentración y disponibilidad de oxitocina, dopamina, y serotonina (hormonas y neurotransmisores relevantes para el apego en seres humanos); además, mediante técnicas de estimulación encefálica transcraneal. Según estos autores, el biomejoramiento no afectaría personalidad, libertad, ni responsabilidad, ya que solamente mejoraría las posibilidades humanas.

Esta postura tan entusiasta como biologicista, evidentemente, cuenta con críticas. Se ha dicho que podría convertir los actos buenos en no elegidos sino automáticos, incluso compulsivos. También se les ha criticado que pudiera bloquearse la verdadera decisión voluntaria y autónoma. Persson y Savulescu (2012) dicen que no, ya que solamente aumentarían las motivaciones y la fuerza de la argumentación para realizar algo moralmente bueno. Como consideran que la educación moral debe continuarse, habría libertad y responsabilidad.

Sin embargo, aceptando como reales todos los supuestos, surgen también paradojas. Por ejemplo, quienes elijan someterse a un biomejoramiento moral ya no podrían elegir después, dado que, tras la mejora, elegirían el bien; es decir, no habría ya disyuntiva entre lo moralmente “bueno” y lo “malo”. ¿Se trataría de una “euro-prótesis moral”? Si fuese éste el caso, ¿el biomejoramiento estaría indicado para quien tenga una evidente pobreza de juicio moral (a veces llamados “idiotas morales”)?: ¿Quién(es) pone(n) los criterios de pobreza o riqueza moral? ¿Cómo establecer los criterios? ¿Siempre es “bueno” eliminar la agresividad? Otro reto dentro de todo ese planteamiento

es que, si las capacidades morales no se heredan, habría que diseñar cómo hacerlo de forma continua para que permaneciera en la población.

Ante estos planteamientos, parece ser que primero debería avanzarse en el conocimiento de que efectivamente hay intervenciones probadas para lograr esto, ya que, al no existir, las teorizaciones parecen meras especulaciones. Surgen entonces otras dudas que exceden el ámbito de la neuroética, ¿debería un país como México, con problemas urgentes como pobreza y violencia, dedicarse a estos temas? El libro de Álvarez Díaz (2019) sobre neuroética asume que sí: el saber es necesario para salir adelante como país y no ser un mero importador de tecnología, de ciencia, y de pensamiento.

### 3. *Ética en la investigación de neuroimágenes*

El desarrollo de la neuroética como campo del saber está muy relacionado con los estudios de neuroimagen, particularmente la neuroimagen funcional. Destaca en las investigaciones la imagen por resonancia magnética funcional (o fMRI por sus siglas en inglés). Abordar los problemas éticos alrededor de este tema es un punto de unión entre la ética de la neurociencia y la neurociencia de la ética. Esto es porque si se hace investigación con sujetos en principio sanos en protocolos de investigación sobre neurociencia de la ética, ¿qué debería hacerse si se encuentra algún hallazgo en la neuroimagen?

No es infrecuente que en la práctica clínica un paciente acuda por alguna condición y durante el proceso diagnóstico se encuentre otra que el propio paciente desconocía. Esto también puede ocurrir, y de hecho ocurre, cuando se hace investigación clínica: los sujetos de investigación pueden acudir suponiéndose sanos en tanto que asintomáticos, pero descubrirse que tienen alguna condición hasta ese momento desconocida. En 1982 aparece en la literatura médica un término para un hallazgo en la imagen de las glándulas suprarrenales: un “incidentaloma” (Geelhoed, 1982). En lengua española, un “incidente” es algo que surge o aparece durante el curso de algún evento; en este caso, el proceso diagnóstico para algún malestar. El sufijo “oma” denota el resultado de un proceso (como en “ateroma”, “glaucoma”, o “tracoma”); una especialización de este significado es la que

se refiere a un tumor (palabra que viene de tumefacción, hinchazón; un tumor es una masa, que puede ser benigna, o maligna como en el caso del cáncer). Así, el término quería dar a entender una masa que aparecía accidentalmente en un estudio de imagen.

En el caso de la neuroimagen, probablemente los incidentalomas más comunes se relacionan con la glándula hipófisis (Boguszewski, *et.al.*, 2019); desde luego, no son los únicos hallazgos que pueden encontrarse. Dado que la neuroética aparece como término en 1973, y como campo interdisciplinar en el 2002, el número de investigaciones que utilizaban neuroimagen (estructural y funcional, particularmente esta última) se incrementó a partir del inicio del nuevo milenio. Después de un lustro de desarrollo comenzó la preocupación: si se realiza una investigación sobre neurociencia de la ética, y en una neuroimagen de un participante que se sabe sano aparece un incidentaloma, ¿qué debería hacerse? Entre las primeras recomendaciones que se dan, se encuentran las siguientes (Nelson, 2008):

a) Los Comités de Ética en Investigación deben llegar a un consenso respecto a lo que debe indicarse en el proceso del consentimiento informado sobre los hallazgos accidentales, y este acuerdo debe ser conocido por todo investigador que use neuroimagen.

b) Las agencias de financiación deben exigir que todos los escáneres de resonancia magnética en este tipo de investigaciones los utilizara un neurorradiólogo, ya que en el caso de un paciente existe un historial clínico y datos que llevan a sospecha, mientras que en una investigación el participante suele ser un sujeto sano.

c) Lo ideal es que dentro del equipo de investigación se encuentre un neurorradiólogo certificado quien interpreta las neuroimágenes.

d) Idealmente debería hacerse una base de datos nacional a la cual se reportarán todos los hallazgos de todas las investigaciones.

e) Lo anterior permitiría hacer investigación sobre el tema y analizar repercusiones para la propia investigación en neuroimagen, así como los posibles usos de seguimiento clínico. A la base podrían tener acceso tanto clínicos como investigadores.

f) Debería analizarse prudentemente si una base de esta naturaleza sería necesaria para otros estudios de neuroimagen (estructurales, o estudios eléctricos).

g) Analizar las necesidades educativas en términos específicos de incidentalomas en neuroimágenes y su manejo ético por parte de los investigadores.

Posteriormente aparecieron reportes de caso; suelen tener aspectos que lleven a ejemplificar todo lo que debe hacerse (o no), todo lo que puede encontrarse (o no), para que aporten algo. Por ejemplo, pensar en una mujer estudiante, joven, con un hijo, presentarse como participante a una investigación de neuroimagen dentro de un protocolo de neurociencia de la ética, y encontrar una imagen compatible con una malformación arteriovenosa es una experiencia retadora (Underwood, 2012).

Posteriormente estas reflexiones básicas se extendieron para pensar si sería necesario hacer una base de datos paralela, o incluir en una sola, datos que no tuviesen solamente relación con las estructuras, sino que recabaran la funcionalidad de las mismas (Scott, *et.al.*, 2012). Esto permitiría en teoría, ir conociendo más acerca del funcionamiento normal/anormal de circuitos encefálicos diversos al poder hacer correlatos estructurales y funcionales con miles de participantes (al momento de hacer esta sugerencia, se calculaban ya unos dos mil artículos con neuroimagen funcional). En ese momento se mencionó que podría ser incluso una mala práctica el no informar a un sujeto de investigación sobre este tipo de hallazgos (Berlin, 2013). Además, que deberían tenerse consideraciones particulares con niñas y niños (Di Pietro e Illes, 2013); desde el punto de vista bioético por la maduración y autonomía progresiva en la toma de decisiones, y por la potencial repercusión que podría tener en el futuro de un menor un hallazgo de cualquier anomalía anatómica o funcional.

Las investigaciones más recientes muestran que los incidentalomas en neuroimagen pueden alcanzar hasta un 34%, de los cuales un 10% muestra significación clínica. En caso de un hallazgo incidental, se sugiere el siguiente enfoque (Stip, *et.al.*, 2019):

a) Quien realice el hallazgo debe informar al investigador principal para que envíen las neuroimágenes de forma anónima para que las revise un neurorradiólogo.

b) El neurorradiólogo, quien puede ser parte del equipo o ser contratado para el servicio específico (dependerá del presupuesto de la investigación), revisaría las neuroimágenes para generar un resumen descriptivo no clínico, dando su punto de vista respecto a si los hallazgos observados justifican un seguimiento por parte de un médico (neurólogo y/o psiquiatra).

c) La académica investigadora principal informaría al sujeto de investigación si los hallazgos encontrados ameritan seguimiento por parte de personal médico.

d) El sujeto de investigación tiene la libertad, una vez de ser informado, de hacer sus arreglos propios para el seguimiento con un médico de su elección.

e) Si los sujetos de investigación no tienen posibilidad de conseguir servicios clínicos de seguimiento por neurólogo y/o psiquiatra, el servicio debe ser ofrecido por el grupo de investigación. Con el consentimiento del participante, el resumen del neurorradiólogo se pondría de forma inmediata a disposición del especialista clínico correspondiente. Resulta esencial tener el proceso del consentimiento informado para investigación, y el documento firmado, además del proceso paralelo del consentimiento informado para atención clínica, y el documento firmado.

f) Si el sujeto de investigación prefiere atención fuera del grupo de investigación, puede enviarse una copia de las neuroimágenes y del resumen descriptivo al profesional de su elección, acompañados del formulario escrito del proceso del consentimiento informado.

Seguramente en el futuro surgirán investigaciones que sistematicen tanto los hallazgos encontrados en la neuroimagen con incidentalomas, como la conducta ética a seguir.

#### 4. *¿Más inteligentes, libre de enfermedades o mejores agentes morales?*

Otra propuesta de Álvarez Díaz (2019) en su reciente libro de neuroética es que los seres humanos buscaron lo humano de la humanidad en la genómica; tras el proyecto genoma humano y encontrar que no son muchos más los genes que tiene el *Homo sapiens* que la *Drosophila melanogaster* (la mosca de la fruta) hubo una cierta decepción.

Al no encontrar lo característicamente humano de la humanidad, voltearon la mirada hacia el encéfalo: ya no se trata de una aproximación genómica, sino neurómica, conectómica. La pregunta previa sería, ¿lo humano de la humanidad está en la biología? O bien, ¿qué porción de lo humano de la humanidad le corresponde a la biología?

Hace poco más de una década un autor tenía un planteamiento interesante sobre lo que Persson y Savulescu (2012) denominaron como biomejoramiento moral. Faust consideraba que los llamados padres comitentes (quienes desean o solicitan un hijo) tenían ya la posibilidad de seleccionar embriones conseguidos mediante fecundación *in vitro* (Faust, 2008). Esto ocurre en muchos países del mundo; prácticamente todos los latinoamericanos, ciertamente con peculiaridades (Álvarez-Díaz, 2020). Ante esta ubicua realidad, Faust toma en cuenta que es posible considerar el haplotipo mediante diagnóstico genético preimplantacional. Si se supiera que el haplotipo MoralKinder (MK +) natural (pero teórico, dado que no está descrito científicamente) predispondría a los individuos a un nivel de moral más alto que el promedio, ¿es permisible u obligatorio seleccionar el haplotipo MK +? Es decir, ¿es moral seleccionar por moralidad? Muy probablemente el mero hecho de considerar que tal haplotipo puede existir es suponer mucho. De hecho, las publicaciones que discuten estos temas son de ética, filosofía, ciencias sociales o humanidades; sin embargo, no hay artículo en revistas de genética o biología que muestren estas posibilidades.

Considerar que una especie biológica más como el *Homo sapiens* pueda quedar libre de enfermedad es meramente ilusorio: todas las especies tienen una serie de interacciones que llevan a la enfermedad, ya sea con elementos bióticos o abióticos del medio. La mejora de la agencia moral seguirá siendo campo de discusión por mucho tiempo, ya que lo esgrimido, a pesar de apelar a la biología y sus ramas, recae en argumentos que, en último de los casos, son epistemológicos y lógicos, para empezar; en resumen, de naturaleza filosófica.

## VI. INTELIGENCIA ARTIFICIAL Y AGENCIA MORAL

### 1. ¿Qué es la Inteligencia Artificial?

Abordar el tema de Inteligencia Artificial (IA) puede ser confuso. La cultura popular, especialmente la ciencia ficción, se ha encargado de transmitir la idea de que este tema es igual a *robots* que son capaces de actuar, pensar e incluso sentir como lo hacen los seres humanos. En otros casos, la idea de lo qué es la IA podría acotarse a algoritmos<sup>1</sup> que son capaces de identificar nuestras preferencias de consumo o de realizar cálculos matemáticos complejos, hechos que contribuyen a pensar que las computadoras son “inteligentes”, aunque tales actividades sean de lo más sencillo para ellos.

Hay algo de cierto en esas visiones, sin embargo, parecen estar lejos de ser ideas que pueda contribuir a un entendimiento más fundamentado de lo que es la IA. Aquí toca hacer nuestra primera advertencia: en la actualidad no existe una definición única de lo que debería entenderse por IA. En ese sentido, se explorarán brevemente algunas de ellas, con la intención de zanjar el terreno para una propia sobre la cual gire este trabajo de investigación.

La primera aproximación a la IA refiere a que es el campo dedicado a la construcción de artefactos que son capaces de mostrar, en entornos controlados y bien conocidos, y durante periodos de tiempo sostenidos, comportamientos que son considerados como inteligentes, o próximos a lo que se considera como tener mente (Arkoudas y Bringsjord, 2014, p. 34).

Otra definición refiere que la IA es el intento de hacer que las computadoras simulen el tipo de cosas que pueden hacer las mentes humanas y animales, ya sea con fines

---

<sup>1</sup> “La palabra algoritmo viene de Abu Abdulah Mhamad ibn Musa Alkhwairismi, matemático persa del siglo IX que escribió el primer libro de álgebra: *Al-Kitab al Mukhtasar fi Hisab al-jabr wa l-Muqabala*. El nombre de álgebra viene proviene de una palabra del título del libro: *al-jabr*. Tan pronto los escolásticos y filósofos medievales empezaron a diseminar la obra de Al-Khwarismi la traducción de su nombre por “algorismo” pronto empezó a describir cualquier método sistemático o automático de cálculo” (Monasterio Astobiza, 2017, p. 186).

tecnológicos y/o para mejorar la comprensión teórica que se tiene sobre los fenómenos psicológicos (Frankish y Ramley, 2014, p. 335).

Una definición más próxima a los elementos técnicos que comprende la IA, se encuentra en lo dicho por Kaplan y Haenlein (2019), cuando indican que es: “*a system’s ability to correctly interpret external data, to learn from such data, and to use those learnings to achieve specific goals and tasks through flexible adaptation*”. [la capacidad de un sistema de interpretar correctamente los datos externos, de aprender de ellos y de utilizar esos aprendizajes para lograr objetivos y tareas específicas mediante una adaptación flexible] (Bartneck, *et.al.*, 2021, p.2).

La IA también se entiende como la capacidad de una computadora o de un robot controlado por computadora para realizar tareas comúnmente asociadas a los seres inteligentes. Es un término utilizado frecuentemente al desarrollo de sistemas dotados de los procesos intelectuales que caracterizan a los seres humanos, por ejemplo, el razonar, descubrir significados, generalizar o aprender de la experiencia pasada (Copeland, 2020).

Stuart Rusell y Peter Norvig (2009) identifican una serie de definiciones en torno a la IA, y las clasifican en cuatro categorías: 1) en relación con el pensamiento, 2) en relación con el razonamiento, 3) en relación con el comportamiento humano, y 4) en relación con el comportamiento racional, según corresponda. Se trae a colación la clasificación de dichos autores con la finalidad de mostrar, de forma muy clara, los enfoques sobre los cuales puede construirse la definición de la IA.

<b>En relación con el pensamiento</b>	<b>En relación con el razonamiento</b>
<i>“The exciting new effort to make computers think . . . machines with minds, in the full and literal sense”</i> (Haugeland, 1985). [El nuevo y emocionante esfuerzo por hacer que los ordenadores piensen . . . Máquinas con mente, en el sentido pleno y literal]	<i>“The study of mental faculties through the use of computational models”</i> (Charniak y McDermott, 1985). [El estudio de las facultades mentales mediante el uso de modelos computacionales].



<p><i>"[The automation of] activities that we associate with human thinking, problem solving, learning . . ."</i> (Bellman, 1978) [La automatización de actividades que asociamos con el pensamiento humano, la resolución de problemas, el aprendizaje...]</p>	<p><i>"The study of the computations that make it possible to perceive, reason, and act"</i> (Winston, 1992). [El estudio de las computaciones que hacen posible percibir, razonar y actuar].</p>
<p><b>En relación con el comportamiento humano</b></p>	<p><b>En relación con el comportamiento racional</b></p>
<p><i>"The art of creating machines that perform functions that require intelligence when performed by people"</i> (Kurzweil, 1990). [El arte de crear máquinas que realizan funciones que requieren inteligencia cuando son realizadas por personas].</p> <p><i>"The study of how to make computers do things at which, at the moment, people are better"</i> (Rich y Knight, 1991). [El estudio de cómo hacer que los ordenadores hagan cosas en las que, por el momento, las personas son mejores].</p>	<p><i>"Computational Intelligence is the study of the design of intelligent agents"</i> (Poole, et.al., 1998). [La inteligencia computacional es el estudio de agentes inteligentes].</p> <p><i>"AI . . . is concerned with intelligent behavior in artifacts"</i> (Nilsson, 1998). [la IA . . . se ocupa del comportamiento inteligente en los artefactos].</p>

**Figura 1.** Categorización hecha por Stuart Rusell y Peter Norvig respecto a algunas definiciones de la IA.

Como es posible identificar —y como ya se había adelantado— la IA puede ser entendida de distintas maneras. Su definición sobre qué debería y que no debería incluir la IA ha cambiado a lo largo del tiempo (Bartneck, *et.al.*, 2021, p. 7). Todas, sin embargo, parecen coincidir en que la IA busca replicar a través de computadoras a la inteligencia humana.

En vista de todo lo anterior, la definición propuesta en este apartado para entender a la IA es la siguiente: la Inteligencia Artificial es una disciplina científica que se desprende de las ciencias de la computación, y que pretende dotar de autonomía y

adaptabilidad a las computadoras, con la finalidad de que realicen tareas de forma eficiente e innovadora y sin la necesidad de supervisión humana.<sup>2</sup>

Aquí es crucial clarificar lo siguiente. Cuando se hace referencia a la autonomía, es en el sentido vinculado a las ciencias de la computación, no a la bioética propiamente. Se habla de autonomía porque la computadora realiza tareas en entornos complejos sin la guía constante de un usuario, y no por el hecho de tener ciertas capacidades para elegir en tanto animales humanos o no como sería para el principio de autonomía en bioética.<sup>3</sup> Este sentido de la palabra autonomía poco tiene que ver con su definición desde el mirador bioético-filosófico —la cual se aborda más adelante—. La adaptabilidad, por su parte, implica la capacidad de la computadora de mejorar su rendimiento aprendiendo de la experiencia.

Una vez identificada la definición de IA, es momento de señalar tres clasificaciones en torno a la IA: IA aplicada, IA fuerte y Superinteligencia.

### 2. *Inteligencia Artificial aplicada*

También conocida como IA Débil o Estrecha, aunque la nomenclatura de IA Aplicada aparenta ser más precisa con relación a lo que busca describir. La IA Aplicada busca producir sistemas “inteligentes” comercialmente viables. Algunos ejemplos son los sistemas expertos utilizados en el sector médico y en el sector financiero (Copeland, 2020).

La IA aplicada está limitada a una única tarea muy definida. La mayoría de los sistemas actuales de IA caben en esta categoría: *“These systems are developed to handle a single problem, task or issue and are generally not capable of solving other problems, even related ones”* (Bartneck, *et.al.*, 2021, p. 10). Estos sistemas se

---

<sup>2</sup> Esta definición fue construida gracias a la parte 1 “Introduction to AI”, del curso “*Elements of AI*”, de la Universidad de Helsinki. University of Helsinki, Elements of AI, Part 1: *Introduction to AI*, 2019. Disponible en línea en: <https://www.elementsofai.com/>, fecha de consulta: 07 de mayo 2021.

<sup>3</sup> University of Helsinki, Elements of AI, “How should we define AI”, en Part 1: *Introduction to AI*, *op. cit.* Disponible en línea, en: <https://www.elementsofai.com/>, fecha de consulta: 07 de mayo 2021.

desarrollan para gestionar un único problema, tarea o cuestión y por lo general no son capaces de resolver otros problemas, ni siquiera los relacionados.

Otros ejemplos de IA Aplicada son los vehículos autónomos, la visión por computadora, el reconocimiento por voz y la traducción de idiomas, sólo por mencionar algunos de ellos (Chakraborty, 2017). Estas aplicaciones tienen un alcance limitado, en tanto que sólo pueden realizar la actividad para la cual fueron diseñadas (de ahí que también que se le conozca como IA Débil o Estrecha). La IA Aplicada simula —pero no replica— a la inteligencia humana (Kaplan, 2016, p. 68).

### 3. *Inteligencia Artificial fuerte*

En contraste con la IA Aplicada, la IA Fuerte<sup>4</sup> —también conocida como IA General— apuesta a construir máquinas que piensen, máquinas cuya capacidad intelectual sea indistinguible de la de un ser humano. Esta pretensión suscitó gran interés en las décadas de 1950 y 1960, sin embargo, fue desechada en vista de las extremas dificultades que supone lograr esto.

John Searle (1980, p. 417) define a la IA Fuerte como:

“Según la IA Fuerte, una computadora no es una mera herramienta en el estudio de la mente; más bien, la computadora adecuadamente programada es realmente una mente, que con los programas adecuados podría comprender y contar con estados mentales; los programas no son meras herramientas que nos permiten probar explicaciones psicológicas, sino que los programas son en sí mismos las explicaciones” [traducción propia].

En la actualidad, no hay evidencia de que la comunidad académica dedicada a la Inteligencia Artificial busque crear un agente que cumpla con las características de la IA Fuerte. La AI Fuerte, sin embargo, sigue siendo el “*santo grial*”, un objetivo que se nutre con las especulaciones en torno a la idea de la singularidad (Chakraborty, 2017).

---

<sup>4</sup> El término de IA Fuerte fue acuñado por John Searle, en 1986, para referirse a la idea de que una computadora puede ser una mente. Cabe señalar que su postura no es para apoyar la idea de que las máquinas pueden ser inteligentes, sino para refutar la misma, a través de su experimento denominado la habitación china (Searle, 2003).

#### 4. Superinteligencia

La Superinteligencia refiere a la idea de un sistema de IA que sobrepasa la inteligencia humana individual y colectiva (Mainzer, 2020, p. 223). Esta idea forma parte también de una postura vinculada a la narrativa distópica en la que la IA adquiere un nivel de inteligencia capaz de suponer un riesgo para la existencia de la civilización humana (Crawford, 2021, p. 214).

Se trata de una visión que describe un probable — quizá una muy lejana probabilidad, aunque no se descarta que la civilización humana se acerque a ella con la irrupción de la computación cuántica—<sup>5</sup> tipo de IA. Las primeras visiones en torno a la Superinteligencia fueron descritas por el matemático Irving John Good, a través de lo que llamó la “explosión de la inteligencia”. Esta explosión se refiere a una IA lo suficientemente inteligente como para entender su propio diseño, capaz de rediseñarse a sí misma y de crear un sistema sucesor, más inteligente, que a su vez también podría rediseñarse para ser aún más inteligente.

Nick Bostrom identifica a la Superinteligencia como uno de los “*riesgos existenciales*”, en el que un resultado adverso aniquilaría la vida inteligente originaria de la Tierra o reduciría su potencial de forma permanente y drástica. De cara a esta visión fatalista, también identifica una positiva, en la que la Superinteligencia contribuye a la preservación de la vida inteligente originaria de la tierra y ayuda a desarrollar su potencial (Frankish y Ramsey, 2014, p. 328). Bostrom también identifica como algo necesario que la IA aprenda los valores humanos, con la intención de formar una IA amistosa. De acuerdo con este autor, la Superinteligencia tendría que contribuir a la felicidad y plenitud de la humanidad.

---

<sup>5</sup> La computación cuántica busca aprovechar las propiedades cuánticas de los qubits — los cuales son sistemas cuánticos de dos niveles, que pueden representar estados intermedios entre el 0 y el 1, lo cual se conoce como superposición cuántica— con la finalidad de correr algoritmos cuánticos que utilizan la superposición y el entrelazamiento para ofrecer una capacidad de procesamiento mucho mayor que las computadoras clásicas. (Allende López, 2019). La combinación de la computación cuántica y la IA promete conseguir nuevos niveles de complejidad para el aprendizaje automático (Knight, 2019).

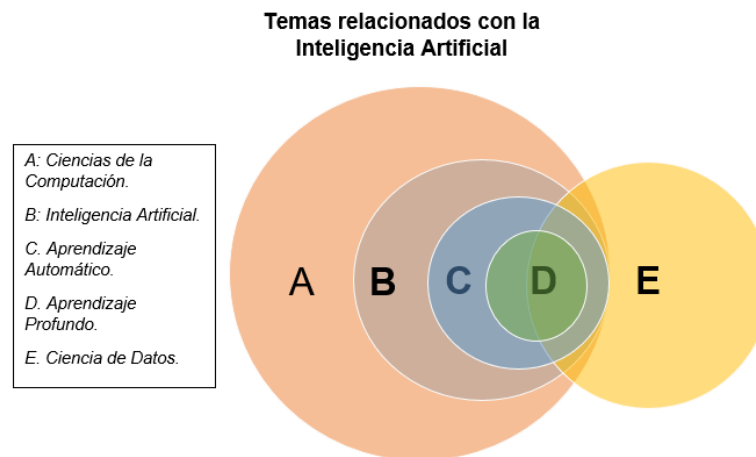
Algunos ejemplos de Superinteligencia son los siguientes:

- **Superinteligencia rápida:** un tipo de IA que puede hacer todo lo que hacen los seres humanos, pero mucho más rápido que ellos.
- **Superinteligencia colectiva:** un tipo de IA compuesto por una gran cantidad de subsistemas de IA, los cuales por sí solos son inferiores a las capacidades de los seres humanos, pero que en conjunto son superior a ellos.
- **Nueva Superinteligencia:** un tipo de IA que cuenta con capacidades intelectuales con las que ningún ser humano cuenta.

Pese a que la idea de la Superinteligencia resulta en la actualidad improbable, y que se aproxima más al campo de la ciencia ficción, existen análisis dentro de la academia —en donde Nick Bostrom es quizá el teórico más reconocido sobre el tema— que enfatizan la necesidad de poner sobre la mesa posibles soluciones frente a los dilemas bioéticos que pudiera traer consigo este tipo de IA.

### 5. Temas relacionados con la Inteligencia Artificial

La siguiente taxonomía sirve para identificar los subcampos que componen a la IA y el campo interdisciplinario con el cual tiene relación.



**Figura 2.** Taxonomía de los temas relacionados con la Inteligencia Artificial. Elaboración propia.

## 6. La Agencia Moral y las condiciones necesarias para IA

Para comenzar este apartado, cabe aclarar que no es la intención de este apartado profundizar en todo el espectro de definiciones existentes alrededor de este principio en bioética, puesto que este objetivo se ha desarrollado alrededor de los diversos textos que componen esta serie de enseñanza transversal en bioética y bioderecho. En todo caso, lo que se pretende aquí son dos cosas: construir una definición mínima para la IA e identificar las condiciones necesarias que ayuden a comprender qué es lo que configura a un agente moral, en relación con la IA.

Eso sí, para lograr lo anterior será necesario referir algunas ideas en torno al tema de algunas posturas en esta área. Con esto se busca proporcionar algunas pistas para entender a la agencia moral en IA y analizar si es posible entender una agencia moral artificial que corresponda a los sistemas artificiales autónomos.

### A. Descripción de los tipos de agencia que pudieran llegar a desempeñar los sistemas artificiales autónomos

Como nota previa al desarrollo de este apartado, es pertinente aclarar que tiene como base la exposición y crítica que Fabio Tollon (2019) hace en su artículo titulado: “*Moral Agents or Mindless Machines? A Critical Appraisal of Agency in Artificial Systems*” de las tres conceptualizaciones realizadas por Deborah Johnson (conocida por sus reflexiones en torno a la ética de la computación) y sus coautores, respecto a los roles potencialmente cargados de moralidad que pueden llegar a desempeñar los sistemas artificiales autónomos. Los tres tipos de agencia que pudieran ser acordes a dichos sistemas son los siguientes: la agencia de eficacia causal, la actuación en nombre de la agencia y la agencia autónoma.

Es conveniente también indicar la idea de agencia utilizada en el texto de Tollon, para el autor los agentes son entidades capaces de tener un cierto efecto en el mundo, y en donde dicho efecto suele corresponderse a ciertos objetivos, ya sea en forma de deseos, creencias e intenciones que tiene el agente.

El artículo de Tollon muestra lo que tal vez sean los principales argumentos para afirmar o negar la adscripción de una agencia moral a las máquinas dotadas con Inteligencia Artificial. Cabe señalar que las conclusiones a la que llega el autor en su texto es que el uso continuo del concepto de la palabra “*autonomía*” en las discusiones sobre la agencia artificial complica las cosas innecesariamente cuando se busca reflexionar sobre los papeles morales que las máquinas pueden llegar a desempeñar:

*Esto sucede de dos formas: en primer lugar, la condición de autonomía, incluso desde su enfoque ingenieril, no describe con precisión la forma en que los sistemas ‘autónomos’ se desarrollan en la práctica. En segundo lugar, el uso continuado del concepto de autonomía en el sentido moral introduce una carga metafísica innecesaria en el debate sobre el libre albedrío en las discusiones sobre la agencia moral [...] Afirmé que el uso continuado de un término tan cargado de metafísica complica nuestra capacidad para manejar bien nuestros conceptos y oscurece las formas en las que podemos pensar de forma coherente acerca del papel o los papeles morales que las máquinas pueden llegar a desempeñar [traducción propia] (Tollon, 2019, p. 21).*

No es el objetivo de este apartado confrontar esta postura. Como ya se ha ido adelantando, su intención es mostrar el marco conceptual descrito en el artículo, con la finalidad de ampliar el panorama en torno al estado actual de las discusiones sobre la agencia moral en la IA.

### *a. Agencia de eficacia causal*

Este tipo de agencia refiere a: “.... a la capacidad de algunas entidades — específicamente los artefactos tecnológicos— de tener una influencia causal en diversos estados de las cosas, como extensiones de la agencia de los seres humanos que los producen y utilizan” (traducción propia) (Tollon, 2019, p. 11). Esta categorización incluye artefactos que pueden separarse de los humanos, tanto en el tiempo como en el espacio, así como artefactos que son desplegados directamente por el ser humano.

Una buena pregunta es si tiene sentido pensar en estos artefactos como “*agentes*”, pues se asemejan más a una “*herramienta*”. La respuesta que Deborah Johnson tiene para esto es que los artefactos tecnológicos son “*agentes*” en tanto que en ellos se

encuentran programadas/codificadas las intenciones humanas. Una herramienta, como lo puede ser un martillo, no cuenta con las especificaciones de cómo debe ser utilizado (bien puede ser utilizado para golpear a una persona o martillar un clavo) en su propia composición.

Además, no puede realizar o representar la tarea de forma independiente de la manipulación humana: “La principal distinción entre una herramienta y un artefacto tecnológico, según Johnson, es que este último tiene una forma de intencionalidad como característica clave de su composición, mientras que el primero no” (traducción propia) (Tollon, 2019, p. 12).

En el sentido expuesto por Johnson, la intencionalidad de la tecnología pone de manifiesto que los artefactos tecnológicos estarían diseñados de determinadas maneras para lograr ciertos resultados. El ejemplo utilizado por explicar esto, es el siguiente:

*Consideremos el ejemplo de un motor de búsqueda: se pueden pulsar las teclas en un orden específico en una caja de texto, y luego se pulsa un botón. A continuación, el motor de búsqueda pasa por una serie de procesos que fueron programados por el ser humano. En efecto, las razones por las que el programa hace lo que hace están necesariamente ligadas a las interacciones del ser humano que lo creó (traducción propia) (Tollon, 2019, p. 12).*

Desde dicho mirador, tiene sentido pensar que los artefactos tecnológicos cuentan con “*agencia*”, en virtud de que suponen (debido a su diseño específico) una diferencia en los resultados efectivos de los que se dispone. Los artefactos hacen posibles formas innovadoras de alcanzar determinados fines y, por lo tanto, amplían el abanico de acciones disponibles para un agente concreto en una situación particular.

La eficacia causal del artefacto se engloba precisamente en el hecho de que contribuyen a la creación de ciertos estados de cosas nuevos, pero siempre en conjunción con las acciones de los seres humanos:

La razón por la que podemos pensar en estos artefactos con eficacia causal como agentes es el hecho de que hacen contribuciones causales sustanciales a ciertos resultados. Así, la eficacia causal conduce, en forma de una acción no trivial realizada por el artefacto, a un acontecimiento concreto (traducción propia) (Tollon, 2019, p. 13).



Los artefactos tecnológicos son, en efecto, agentes por el hecho de que sus fabricantes tienen ciertas intenciones al diseñarlos y crearlos, que adquieren sentido en relación con los seres humanos. En palabras de Tollon, la concepción de agencia dada a los artefactos tecnológicos refiere a entidades que actúan en conjunción causal con los seres humanos. Al reflexionar sobre el sentido de responsabilidad asociado a este tipo de agencia, Tollon señala que no es posible identificarla en el artefacto tecnológico, por lo que no es viable considerar al artefacto como un agente moral.

### *b. Actuando en nombre de la agencia*

Esta forma de agencia alude a los artefactos tecnológicos que actúan en nombre de operadores humanos. Es una especie de agencia subrogada, algo similar a lo que ocurriría cuando una persona actúa en nombre de otra. La persona representante se encuentra limitada por ciertas normas y responsabilidades (Tollon, 2019, p. 13).

Los artefactos que sustituyen o actúan en nombre de operadores humanos lo hacen en determinados ámbitos. El ejemplo utilizado en el texto de Tollon para describir este tipo de agencia es el de una plataforma de intercambio, que funciona a través de ordenadores y que se caracteriza por las operaciones de alta frecuencia que ahí ocurren. Todo esto ocurre sin la necesidad de que un agente de bolsa humano lleve un registro. Lo que al final busca dejar de manifiesto este ejemplo es que en la actualidad es posible que tareas que antes eran de dominio exclusivo del ser humano, son ahora realizadas por sistemas artificiales, sin necesidad de que los primeros intervengan. Si bien es cierto que la tarea de operar en el mercado de valores la realizan sistemas artificiales, también lo es que ésta sigue siendo la misma que realizaría un agente de bolsa humano. El artefacto, en efecto, actúa dentro de unos parámetros establecidos previamente por un operador humano (Tollon, 2019, pp. 13-14).

La distinción de un artefacto que actúa en nombre de la agencia y un agente que es causalmente eficaz, es que este último influye en los resultados junto con los seres humanos, mientras que el primero puede actuar a nombre de un agente humano, pero con la limitante de que las acciones que puede ejecutar dependen de las intenciones de sus programadores/diseñadores, es decir, se encuentra determinado.

Otra diferencia importante es que un artefacto que actúa en nombre de la agencia tiene un mayor grado de independencia de la intervención humana; el ejemplo de la plataforma de operaciones ayuda bastante a comprender este punto, pues para ingresar una orden de compra o de venta, y que esta se ejecute, no requiere de la intervención humana, pues los propios algoritmos sobre la que fue construida se encargarán de ello, por lo que tienen en mayor medida inserta la intencionalidad humana.

Sobre esto último, Deborah Johnson (2006, p. 222), afirma que al evaluar el comportamiento de los sistemas informáticos: hay una triada de intencionalidad en el trabajo, intencionalidad en el sistema e intencionalidad en el usuario], y que comprende estados mentales internos como las creencias y los deseos, en donde finalmente reside la responsabilidad de cualquier acción del sistema (Tollon, 2019, p. 14).

Estos artefactos, en efecto, pueden tener consecuencias en la vida moral de las personas. El ejemplo utilizado es el siguiente:

*Consideremos, por ejemplo, la tecnología de frenado automático de emergencia (AEB, por su sigla en inglés), que aplica automáticamente los frenos cuando detecta un objeto cerca de la parte delantera del vehículo. Este sencillo sistema ha tenido un enorme éxito, y las investigaciones indican que podrían reducir las colisiones con peatones, los giros a la derecha, los choques frontales, los choques por detrás y los choques con objetos fijos. Podemos pensar que el AEB nos ayuda a ser mejores y más seguros conductores, pues ayudan a reducir las muertes y lesiones en carretera. Estos sistemas artificiales, de los que el AEB es un ejemplo, pueden considerarse como tareas delegadas que pueden tener un significado moral. En efecto, podemos considerar que son entidades moralmente relevantes (traducción propia) (Tollon, 2019, p. 14-15).*

Pese a tener un impacto en el ámbito moral de los humanos, artefactos como el señalado no pueden ser considerados como agentes morales, ya que la intencionalidad deviene del diseñador y del usuario del artefacto. Johnson señala que sería un error afirmar que los seres humanos y los artefactos son componentes intercambiables en la acción moral. Para finalizar con esta conceptualización de agencia, Fabio Tollon señala que los artefactos pueden ser agentes que, al actuar en nombre de los seres humanos, participan en actos con consecuencias morales. Enfatiza también de esto no se infiere

que sean moralmente responsables de los actos, pues la responsabilidad está reservada para los seres humanos, y remata afirmando que, para ser moralmente responsable, un agente también debe contar con autonomía (Tollon, p. 16).

### *c. Agencia autónoma*

En este apartado, Fabio Tollon explica que en su relato se puede entender a la autonomía de dos formas: la primera, como algo que suele atribuirse a los seres humanos, y que cuenta con una dimensión claramente moral capaz de hacernos moralmente responsables de nuestras acciones, y la segunda, que puede definirse desde el campo de la ingeniería (de software). A decir de esta última:

*Los informáticos suelen referirse a los programas ‘autónomos’ para destacar su capacidad de realizar tareas en nombre de los humanos y, además, hacerlo de forma muy independiente. Un ejemplo simplista de un sistema de este tipo podría ser un algoritmo de aprendizaje automático que está mejor equipado para operar en entornos novedosos que un simple algoritmo preprogramado* (traducción propia) (Tollon, 2019, p. 16).

Sin embargo, a un algoritmo, aún con su capacidad operar y funcionar con independencia de un ser humano, no puede atribírsele agencia moral, pues en realidad éstos no eligen libremente cómo actuar —hasta el momento en el que se escribe este texto, no es posible que un algoritmo realice una tarea fuera del determinado ámbito para el que fue programado, por ejemplo, un algoritmo destinado a jugar ajedrez no puede identificar imágenes de gatos—.

Así pues, la “verdadera” autonomía “racional” queda reservada para las personas. Son las personas quienes tienen un control significativo sobre sus propias acciones: “un tipo de control que hace que una decisión o acción dependa del agente y no de otras circunstancias externas (Tollon, 2019, p. 17)”. Pero, qué ocurre si la decisión libre no es producto del agente específico en cuestión, sino más bien a una presión externa. Si es así, entonces la acción no es libre, y habría dificultad para determinar si el agente podría ser moralmente responsable de su acto.

La reflexión que hace Tollon al distinguir las dos concepciones de agencia autónoma —la primera inserta en una dimensión moral, y la segunda relacionada con la capacidad de un sistema artificial de actuar sin intervención humana— zanja el terreno para la siguiente conclusión: los artefactos tecnológicos no deberían entenderse como poseedores de una autonomía moralmente relevante, pues no se considera que sus decisiones sean elegidas libremente. Sus intenciones, están siempre vinculadas a las intenciones de sus diseñadores y usuarios finales.

### *B. La visión estándar y la visión funcionalista de la agencia humana como parte del debate en torno a la agencia moral artificial*

Una conceptualización más en torno al debate de la Agencia Moral Artificial (en adelante AMA) es la expresada por Dorna Behdadi y Christian Munthe (2020), en su artículo “*A Normative Approach to Artificial Moral Agency*”. A decir de los autores, existen dos visiones sobre el debate de la agencia moral artificial, y cuyos argumentos devienen de analizar qué es lo que compone a la agencia moral humana.

Las visiones identificadas en el artículo son la estándar y la funcionalista,<sup>6</sup> las cuales son incompatibles entre sí y cada una de ellas con distintas implicaciones frente a la idea de una posible AMA.

#### *a. Visión estándar*

Para esta visión, la agencia moral humana implica que los agentes morales cuenten con racionalidad, voluntad o autonomía y condiciones de conciencia fenomenal. Para Deborah Johnson —una de las principales proponentes de esta visión y quien es citada en el artículo de Behdadi y Munthe— la agencia moral de una entidad (E), debe estar en sintonía con las siguientes condiciones:

---

<sup>6</sup> A nota al pie de página, las autoras aclaran que cuando se refieren a “funcionalismo” y “funcionalistas” en su artículo, se refieren exclusivamente a una idea de la agencia moral, y no a las teorías funcionalistascognitivas, de la conciencia, del significado o cualquier otra área en dónde dichos términos sean utilizados.

1. E provoca un evento físico con su cuerpo. 2. E tiene un estado interno, I, que consiste en sus propios deseos, creencias y otros estados intencionales que, en conjunto, conforman una razón para actuar de una determinada manera (racionalidad y conciencia). 3. El estado I es la causa directa de 1. 4. El acontecimiento de 1 tiene algún efecto de relevancia moral (Behdadi y Munthe, 2020, p. 198).

Sobre estas condiciones, Behdadi y Munthe puntualizan que el número 1 (y posiblemente el 2) garantizan que E sea un agente, en tanto que 2 y 3 señalan lo que se requiere para la presencia de una agencia moral como la adscrita a los seres humanos. La particularidad número 4 asegura que el comportamiento es moralmente relevante, indicando que E ejerce su agencia moral en la situación correspondiente.

Deborah Johnson busca resumir su postura diciendo que todo comportamiento puede explicarse por sus causas, pero sólo la acción puede explicarse por un conjunto de estados mentales internos, los cuales comprenden las creencias, deseos y otros estados intencionales que puedan contribuir a explicar por qué un agente actuó.

Las objeciones a una posible AMA se localizan en las condiciones 2 y 3. En efecto, las entidades artificiales son incapaces de ser agentes morales dado que carecen de estados mentales internos que deriven en acontecimientos, los cuales a su vez son sus acciones. Es cierto que estas entidades pueden actuar, y que sus comportamientos se asemejan a la acción humana desde un punto de vista externo; sin embargo, estos comportamientos no bastan para conferirles cualidades morales, ya que carecen de estados mentales internos.

En efecto, la visión estándar aboga por una agencia moral con cualidades vinculadas a la naturaleza humana, y frente a la cual la idea de una agencia moral artificial no podría justificarse, pues es evidente que un robot no cuenta con estados mentales internos que orienten sus actuaciones.

### *b. Visión funcionalista*

La visión funcionalista tiene en Luciano Floridi y Jeff Sanders a sus principales representantes (Dorante, 2017). Entre sus posturas más relevantes, se encuentra el rechazo a criterios como el de conciencia y la adopción de la idea de una “*moralidad no*

mental”. Ambos autores señalan que las entidades pueden ser agentes morales dependiendo del Nivel de Abstracción elegido para dar cuenta de los componentes de la agencia moral humana. Behdadi y Munthe (2020, p. 198) explican:

*El nivel de abstracción aplicado por el punto de vista estándar es muy bajo, manteniendo los criterios cercanos al caso de un ser adulto, pero al elevar el nivel es posible contar con perspectivas menos antropocéntricas mientras se mantiene la consistencia y similitud relevante concernientes a los rasgos estructurales que subyacen en los agentes morales humanos (traducción propia).*

Floridi y Sanders ofrecen un conjunto de condiciones que permiten identificar a la agencia humana: Interactividad, Autonomía<sup>7</sup> y Adaptabilidad. Behdadi y Munthe (2020, p. 198) enuncian y describen dichas condiciones de la siguiente forma:

*1. Interactividad: E interactúa con su entorno. 2. Independencia: E tiene la capacidad de transformarse a sí y a sus interacciones independientemente de las influencias externas inmediatas. 3. Adaptabilidad: E puede cambiar la manera en que se actualiza 2 en función del resultado de 1 (traducción propia).*

De acuerdo a Behdadi y Munthe, al contrastar estas condiciones con las identificadas en la visión estándar de la agencia moral humana, puede identificarse que la condición 1 es correspondida con su contrapartida de la visión estándar, que la condición 2 se aleja de la visión estándar pues no es necesario contar con ningún tipo de estados mentales internos para contar con autonomía, en tanto que la condición 3 también es diferente.

Así, desde la visión funcionalista es posible justificar la existencia de la agencia moral artificial, pues basta que un ser cuente con los niveles de abstracción identificados para la agencia moral, para ser considerada como un agente moral. Esta visión permite pensar no sólo en la moralidad —comprendida como la capacidad del agente de regular su

---

<sup>7</sup> El término utilizado por Behdadi y Munthe es el de independencia, sin embargo, en el trabajo de Floridi y Sanders el término empleado es el de “autonomía”. La justificación que tienen los autores para este cambio es que el término “autonomía” no corresponde al uso estándar que se le da en la filosofía, y que se asemeja más a concepto de autonomía empleado en las ciencias de la computación y la robótica. (Behdadi y Munthe, 2020, 199).

conducta en torno a principios e ideas generales relativos a la bueno y lo justo— de los sistemas artificiales con autonomía, sino también en la consideración moral y obligaciones que como seres humanos deberíamos tener hacia ellos.

### *C. ¿Es posible adscribirles una agencia moral a las máquinas autónomas?*

El debate sobre la AMA es complejo, y tiende a serlo más conforme los debatientes incorporan nuevas variantes conceptuales, en donde la idea de agencia moral u otras que son clave en la discusión —como puede serlo el de responsabilidad— adquieren matices que facilitan la inclusión de una gama más amplia de tipos de agentes, por ejemplo.

Esto pudiera traer consigo la desventaja de que cuando el foco de discusión se centra en qué es un agente moral artificial, las cuestiones prácticas urgentes que involucran entidades artificiales cada vez más avanzadas se ensombrezcan, pero también la ventaja de que el debate en torno a este tema se nutre, y además permite imaginar soluciones a dilemas éticos que pudieran llegar a presentarse producto de la interacción entre robots y seres humanos, por ejemplo.

La intención de este apartado es en vías de contar con una posible respuesta a la pregunta de si es posible adscribirles una agencia moral a los sistemas artificiales autónomos, es la de contar con bases que faciliten la construcción de propuestas orientadas en el corto y mediano plazo al desarrollo ético de la Inteligencia Artificial, y a largo plazo, a reflexionar sobre la moral que pudiera regir en un AMA. Nuestra primera aproximación es que no es posible adscribirles una agencia moral a los sistemas con inteligencia artificial, al menos no en el presente.

Que se niegue la agencia moral no implica que no se quiera reflexionar sobre el impacto en el campo moral de las decisiones hechas por sistemas artificiales autónomos. Tampoco implica negar o restarle valor a una ética específica para reflexionar críticamente sobre la moral de dichos sistemas. Sobre esto último, aún en ausencia de una agencia moral artificial, resulta relevante seguir cuestionando desde la bioética el comportamiento de los sistemas con inteligencia artificial y si los principios, por ejemplo,

que han de seguirse por estos sistemas pueden serles codificados —en el sentido informático— (Nath, 2017).

De cara al desarrollo de la Inteligencia Artificial, no parece buena idea pensar que la ética solo deba reflexionar sobre la moral de los agentes a los que se les puede responsabilizar moralmente de sus acciones, en razón de que actúan con autonomía (en el sentido filosófico de la palabra) y con estados mentales que impulsan sus actos, porque llegará el momento en el que los sistemas artificiales autónomos se separen varios grados de la responsabilidad de su diseñador, de su creador o de quienes los despliegan.

Michael Anderson y Susan Anderson (2007) traen a escena un argumento que objeta la idea de que no es viable reflexionar sobre la moral de los sistemas artificiales autónomos, sólo porque estos no son agentes morales:

*Este tipo de objeción, sin embargo, muestra que el crítico no ha reconocido una distinción importante entre realizar la acción moralmente correcta en una determinada situación, incluyendo la posibilidad de justificarla apelando a un principio ético aceptable, y ser considerado moralmente responsable de la acción. Sí, la intencionalidad y el libre albedrío son necesarios en cierto sentido para considerar a un ser moralmente responsable de sus acciones, y sería difícil establecer que una máquina posee estas cualidades, pero ninguno de los dos atributos es necesario para realizar la acción moralmente correcta en un dilema ético y justificarla. Todo lo que se requiere es que la máquina actúe de una manera que se ajuste a lo que se consideraría la acción moralmente correcta en esa situación y sea capaz de justificar su acción citando un principio ético aceptable que esté siguiendo (traducción propia) (Anderson y Anderson, 2007, p. 19).*

Reiterando, es importante que la ética reflexione sobre la moralidad de los actos realizados por máquinas dotados de inteligencia artificial, para cual se requiere una ética específica, que se centre en los dilemas que se presentan cuando éstas interactúan con seres humanos y con el medio ambiente, por ejemplo. Según la evolución de la IA, seguiremos reflexionando más sobre la Ética de la Inteligencia Artificial.

Ahora bien, toca justificar la respuesta dada a la pregunta de si es posible adscribirles una agencia moral a las máquinas dotadas de inteligencia artificial. Como se adelantó,



la respuesta es no, pues en la actualidad ningún sistema con inteligencia artificial cumple con condiciones internas que son cruciales a la hora de adscribir una agencia moral.

Las máquinas, al no contar con condiciones internas como la autonomía o la intencionalidad, no están habilitadas para reaccionar por sí solas de forma correcta frente a una o varias circunstancias sociales que así lo requieran; tampoco cuentan con la capacidad de justificar sus acciones o elecciones ni mucho menos con la capacidad de actuar moralmente de forma consciente y también coherente con sus emociones —con las cuales tampoco cuenta— y con las prácticas de la comunidad en la que se desenvuelve.

Esta idea surge al pensar en los tipos de agencia moral descritos en el artículo titulado: “*Can artificial intelligences be moral agents?*”, aquí las autoras refieren a tres tipos de agencia: 1) El agente moral irreflexivo o superficial: alguien que posee la arquitectura emocional y cognitiva que le permite seguir “ciegamente” las reglas de la comunidad. Este tipo de comportamiento no requiere reflexión teórica, ni justificación alguna de sus acciones y elecciones realizadas. 2) El agente moral reflexivo: es un individuo que no sólo es capaz de seguir inconscientemente las reglas morales, sino que también puede justificar sus acciones y elecciones, adaptando de forma consciente una regla moral abstracta (aquí encajan la mayoría de los adultos humanos con buenas aptitudes para la comunicación), y 3) El agente moral sofisticado (o profundo): alguien que es capaz de tomar regularmente decisiones morales de forma consciente, comprendiendo que sus intuiciones pueden llevarle por el mal camino y que sopesar las razones y considerar las teorías morales abstractas debe formar parte del proceso de toma de decisiones morales. En este nivel de autoconsciencia moral se sitúan figuras como Gandhi o el Dalai Lama (Bartosz y Bartosz, 2019, p. 104).

Y la respuesta negativa hacia la pregunta continuará, en tanto los sistemas con inteligencia artificial sigan construyéndose bajo el paradigma actual. Por ejemplo, pensando en “*robots emocionales*” que sean contruidos para simular emociones, pero que sean incapaces de contar con una arquitectura emocional y cognitiva como la de las personas humanas, que les permiten participar en las prácticas morales de su comunidad:

*El problema no está relacionado con la sofisticación y la complejidad de esos robots, sino con algo más fundamental: los principios que guían el desarrollo de su arquitectura. El agente moral irreflexivo humano depende en gran medida de mecanismos emocionales fundamentales, mientras que las arquitecturas de Inteligencia Artificial existentes están lejos de hacerlo. Los robots verdaderamente emocionales no son aquellos equipados con un “módulo emocional” adicional, que calcula algunos parámetros que interpretamos como “el nivel de ira” o “la intensidad de alegría”. Los algoritmos emocionales implicarían más bien ‘una estructura de control que se basa en los estados emocionales en competencia y colaboración para inducir cambios de estado en el sistema para impulsar su comportamiento corporal y cognitivo, no algoritmos que calculan el contenido emocional como si fuera simplemente un resultado. (Más bien habría que prever) una arquitectura de la cognición en la que las implementaciones funcionales de las emociones sean el sub-estado computacional del que emerge la razón, motivando la manipulación de datos de distintas formas que engendran, entre otras actividades, la recopilación de datos (curiosidad, aburrimiento), el pensamiento recombinate (descubrimiento), la evitación de contradicciones (confusión) y la recuperación de errores (alegría). Mientras el planteamiento de la Inteligencia Artificial consista únicamente en ‘imitar’ el comportamiento emocional, no habrá agentes morales artificiales irreflexivos. Si este es el caso, también es imposible que una máquina (desarrollada con el uso de algoritmos no emocionales) sea un agente moral reflexivo o sofisticado, ya que los estratos de la agencia moral -como insistimos- forman una jerarquía, en la que el nivel superior depende de la existencia del nivel inferior (traducción propia) (Nath, 2017, p. 105).*

Así, mientras las máquinas estén diseñadas para imitar comportamientos, sin ninguna reflexión moral de su parte, no será posible pensar en una AMA. Otro argumento que fortalece la posición de este trabajo frente a la posible idea de la agencia moral artificial es el de Carissa Véliz (2021), quien refiere que los algoritmos al no contar con la capacidad de sentir no pueden ser autónomos (en el sentido filosófico de la palabra) ni capaces de responsabilizarse por sus acciones:

*Para tener una concepción del bien que queremos perseguir (autonomía), necesitamos tener un sentido de lo que conduce al placer, a la significación y a la satisfacción. Para guiar nuestras acciones mediante el reconocimiento de las pretensiones morales de las*

*demás personas, de lo que se siente al ser dañada, de lo que podemos hacer a las mentes y cuerpos de las demás a través de nuestras acciones (traducción propia) (Véliz, 2021).*

La anterior idea tiene sentido para este apartado, pues nos conduce a entender que para ser agente moral es indispensable ser sensible, ya que con ello se es consciente de lo que significa infringir dolor o causar placer. Esta autora también hace una interesante reflexión que vale la pena mencionar aquí:

*Puede llegar un momento en que la Inteligencia Artificial sea tan sofisticada que los robots puedan tener deseos y valores propios. Sin embargo, no será por su capacidad de cálculo, sino por su sintiencia, que a su vez puede requerir algún tipo de encarnación. En la actualidad, estamos lejos de crear algoritmos sintientes (traducción propia) (Véliz, 2021).*

Es cierto que al pensar en el futuro se cae en el terreno de la especulación. Si pensamos en el futuro de la Inteligencia Artificial, no sabemos con certeza hacia donde se dirigirá su desarrollo ni los impactos que pudiese tener en la sociedad.

Por la misma razón, no sería justo desechar las palabras de Véliz, las cuales se complementan bien si se piensa en la posibilidad de una superinteligencia artificial, entendida como: cualquier intelecto que supere ampliamente a los mejores cerebros de personas humanas en prácticamente todos los campos, incluyendo la creatividad científica, la sabiduría general y las habilidades sociales, que pueda ser reconocida como un agente moral y no como una mera herramienta o artefacto (Bostrom, 2009). A decir de esto, es valioso señalar lo que piensa Nick Bostrom:

*En la medida en que la ética es una búsqueda cognitiva, una superinteligencia podría hacerlo mejor que los pensadores humanos. Esto significa que las preguntas sobre ética, en la medida en que tienen respuestas correctas a las que se puede llegar mediante el razonamiento y la ponderación de las pruebas, podrían ser respondidas con mayor precisión por una superinteligencia que por los humanos. Lo mismo ocurre con las cuestiones de política y la planificación a largo plazo; cuando se trata de entender qué políticas conducirían a qué resultados, y*

*qué medios serían los más eficaces para alcanzar determinados objetivos, una superinteligencia superaría a los humanos (Bostrom, 2009, p. 378).*

Un escenario con agentes superinteligentes, que incluso sean capaces de sentir, abriría las puertas a una agencia moral artificial, que a su vez permita a las personas humanas exigirles que actúen conforme a los principios e ideas de lo que es bueno, pero también a que éstas últimas respeten posibles derechos con los cuales podrían contar agentes superinteligentes. El abrir las puertas a una agencia moral artificial desde esta perspectiva implicaría reflexiones muy serias desde la Ética de la Inteligencia Artificial, las cuales estén orientadas a asegurar el impacto benéfico de dichos seres en el planeta y en el universo.

Así, la justificación de por qué no es posible hablar en la actualidad de una agencia moral artificial va robusteciéndose. En el mismo sentido, otro punto clave que impide adscribir en la actualidad una agencia moral artificial es que, al hacerlo, la cadena de responsabilidad que vincula a la máquina con su creador se rompería. No está de más proporcionar argumentos adicionales por los cuales es importante conservar esa cadena de responsabilidad:

*Hay razones muy prácticas para ello. Sin poder establecer una cadena de responsabilidad, una sociedad sería menos libre y segura, ya que las trayectorias de los comportamientos de los agentes no serían descifrables ni predecibles (y, por tanto, no estarían sujetas a la culpa y a la responsabilidad). Pero este no es el final de la historia. Asignamos responsabilidades para aprender algo sobre nosotros mismos como parte de una sociedad, es decir, para trazar la distinción entre lo que es aceptable (incluido) y lo que no es aceptable (excluido, pero de alguna manera mantenido y preservado) en la sociedad. La asignación de responsabilidades (morales o legales) es un proceso humano que implica algunos grados de inclusión y también algunos grados de exclusión, no de la sociedad, sino dentro de ella (traducción propia) (Durante, 2017, p. 63).*

Con esta idea no se pretende excluir de ninguna responsabilidad a la máquina, pues finalmente sus actos podrían impactar la esfera moral de los individuos y de las

sociedades, pero sí excluir cualquier posibilidad que en la actualidad apueste por responsabilizar más allá del campo moral, como bien lo puede ser el campo jurídico, a una máquina y no a su creador o a su usuario, pues con ello estaríamos retrayéndonos a la época de la edad media en la cual se enjuiciaba a los animales por supuestamente cometer delitos.<sup>8</sup>

Así, este trabajo considera que los sistemas artificiales autónomos no pueden ser agentes morales en la actualidad. Hoy día, dichos sistemas se asemejan más a artefactos y sí cuentan con una agencia de eficacia causal, semejante a la descrita por Deborah Johnson. Sin embargo, este trabajo también resalta el valor de posturas como la de Luciano Floridi y Jeff Sanders, pues reflexiones así abren la puerta para debatir con profundidad sobre cuestiones como la moralidad distribuida a la que ellos hacen alusión, y que de cierta forma también se recoge como parte de la respuesta dada a la pregunta de este apartado, pues también se reconoce aquí que las máquinas son fuente del bien o del mal morales.

### VII. CASO PRÁCTICO PARA EL DEBATE GRUPAL<sup>9</sup>

Se sugiere el siguiente formato para la discusión grupal, de 1 a 2 horas de discusión grupal para caso hipotético en la modalidad de taller de discusión. De acuerdo con número de participantes en la sesión grupal, se sugiere organizar dos grupos con la intención de que se debatan posturas disidentes sobre el caso. Durante este taller deberán identificar las interrogantes y argumentos a favor y en contra del tema planteado. Tendrán que decidir cuáles de los argumentos identificados aplican mejor para su caso. Se recomienda que elijan quién del grupo será la relatora o bien tome nota de la discusión que se lleve a cabo en su grupo, para que después pueda estructurar el orden de la

---

<sup>8</sup> Una buena introducción a este tema, y algunos casos en donde animales fueron juzgados y sentenciados por tribunales, se localiza en: (Anaya Huertas, 2010).

<sup>9</sup> Este caso fue preparado originalmente por la Doctora Sarah Chan en el idioma inglés para la escuela de verano *EuroStemCell* que se lleva a cabo cada año en Hydra, Grecia. Esta versión se trata de una traducción autorizada por la Dra. Sarah Chan, con modificaciones y adaptaciones menores por parte de las autoras.

discusión oral ante todo el grupo. Al finalizar la discusión, deberán preparar una presentación oral de 5 minutos con los argumentos que apoyen su posición respecto del tema planteado. Podrán preparar una *power point* para utilizar la computadora disponible como apoyo a su presentación.

Grupo 1: A favor (5 minutos) / Grupo 2: En contra (5 minutos)

Seguido de discusión de todo el grupo en su conjunto para preguntas y debate por 15 minutos.

Grupo 2: Si (5 minutos) / Grupo 4: No (5 minutos)

Seguido de discusión de todo el grupo en su conjunto para preguntas y debate por 15 minutos.

### 1. Intervenciones “nerológicas” no probadas

Hechos hipotéticos del caso. Eres un neurólogo especializado en tratamientos para enfermedades neurodegenerativas. Recientemente algunos de sus pacientes le han estado preguntando acerca de las nuevas terapias con células troncales de las que han leído por medio del internet. Estas terapias consisten en la inyección de células troncales dentro del cerebro o médula espinal con el objetivo de reparar tejidos dañados/enfermos. Tal como sucede, la compañía que proporciona esta terapia hizo una solicitud para llevar a cabo un ensayo clínico de este tratamiento a nivel local, incluyendo a tu hospital. Dicha solicitud fue rechazada, aunque hasta donde se puede comprender, las bases científicas en general son sólidas y se cuentan con algunos resultados prometedores en la investigación con animales hasta el momento.

La respuesta de la compañía a esta situación ha sido el establecimiento de clínicas alrededor de fronteras en donde la jurisdicción sobre la investigación y tratamiento clínico tiene una regulación mínima y no tan estricta. Estas clínicas están ofreciendo tratamientos a pacientes en el ámbito privado, muchos de los cuales viajan desde tu país para recibir las terapias.

Un grupo de tus pacientes, mediante su red de apoyo local, ha decidido que quieren probar estos tratamientos. Tal y como uno de ellos dijo, quien se encuentra en las últimas

etapas de su enfermedad: “No tengo nada más que perder, y no existe alguna otra esperanza”. Ellos preguntan si tu firmarías un alta como tratamiento válido fuera de tu país como parte del servicio médico, es decir, que los gastos serían parcialmente reembolsables. Otro paciente también te señala, que: “No se trata solamente del dinero, la mayoría de nosotros puede y pagará el tratamiento si tienen que hacerlo, incluso si no lo apruebas --- Pero sé que hay algunos pacientes que no pueden pagarlos, aunque el resto de nosotros iremos de cualquier manera. Sin embargo, si tu aceptas un seguimiento a nuestra atención, sabremos que estaremos atendidos y aún si el tratamiento no funciona para mi o para los demás, posiblemente tú podrías aprender u obtener conocimiento del monitoreo que hagas sobre nosotros, mismo monitoreo que algún día podría ayudar a otros pacientes con nuestras condiciones”.

**DISCUSIÓN:** ¿Crees que los pacientes que viajan a otras jurisdicciones para recibir tratamientos no aprobados en sus países podrían ser apoyados en su decisión y ser incluidos en estudios de monitoreo una vez que regresen a casa? **Grupo 1: Sí- Grupo 2: No.**

## VIII. FUENTES CONSULTADAS

### 1. *Bibliohemerográficas*

ÁLVAREZ-DÍAZ, J.A. (2020). “Embryo donation among Latin-Americans who have attended assisted reproduction techniques: a first empirical approach”, *JBRA Assist. Reprod*, 24.

——— (2019). “Neuroética: avances científicos y vida mejor”, *Puertas Lumen Gentium*, 3:67-85.

——— (2019a). *Neuroética: relaciones entre mente/cerebro y moral/ética*. Ciudad de México: Universidad Autónoma Metropolitana.

——— (2018). *Aspectos éticos de la nanotecnología en la atención de la salud*. Ciudad de México: Universidad Autónoma Metropolitana.

- (2015). “Neuroética: una introducción”, *Revista Valenciana*, Vol. 15: pp. 157-187.
- (2015a). “Psicoanálisis y bioética”, *Rev Latinoam Bioet*, 15(1):64-79.
- (2014). “¿Es posible la transdisciplinariedad entre neurociencias y religiones?”, *Rev Mex Neuroci*. 2014;15(5):291-296.
- (2014a). “The death of neuroethics (as we once knew it)”, *Neurosciences and History*, 2(1);26-33.
- (2014b). “Lesión ventromedial: Un punto de encuentro entre neurociencias y psicoanálisis”, *Rev Latinoam Psiquiatr.*, 13(1):25-30.
- (2014c). “Libertad y ética. El trabajo de Benjamin Libet”, *Rev Bioética*, 22(3):434-440.
- (2014d). “¿Qué es la dignidad humana?”. En: ÁLVAREZ DÍAZ, J.A. y S. LÓPEZ MORENO, (coords.). *Ensayos sobre ética de la salud. Aspectos clínicos y biomédicos. Aspectos fundamentales y biomédicos, Vol. 1*. Ciudad de México: Universidad Autónoma Metropolitana. pp. 63-90.
- (2013). “Neuroética como neurociencia de la ética”, *Revista de Neurología*, Vol. 57: pp. 374-382.
- (2013a). “Hechos, valores, deberes... y neuronas (Carta al editor)”, *Rev Neurol.*, 57(12):576.
- (2011). “Retos de la bioética en la medicina del Siglo XXI”, *Rev Peru Med Exp Salud Pública*, 28(4):657-663.
- ALLENDE LÓPEZ, M. (2019). “¿Cómo funciona la computación cuántica”, en *Abierto al público, Banco Interamericano de Desarrollo*, 31 de mayo de 2019. Disponible en línea: <https://blogs.iadb.org/conocimiento-abierto/es/como-funciona-la-computacion-cuantica/>.
- ANAYA HUERTAS, A. (2010). “¿Animales ante los tribunales?”, *Nexos, el juego de la Suprema Corte*, 22 de junio de 2010. Disponible en: <https://eljuegodelacorte.nexos.com.mx/%C2%BFanimales-ante-los-tribunales/>.



- ANDERSON, M. y S. Anderson. (2007). "Machine ethics: creating and ethical intelligent agent", *AI MAG*, núm 28.
- ARKOUDAS, K. y S. BRINGSJORD, (2014). "Philosophical foundations", En: FRANKISH, K. y W. M. RAMSEY (eds.), *The Cambridge Handbook of Artificial Intelligence*, Reino Unido: Cambridge University Press.
- BATISTELA, S., BUENO, O.F.A., VAZ L.J. y J.C.F. GALDURÓZ. (2016). "Methylphenidate as a cognitive enhancer in healthy young people", *Dement Neuropsychol*, 10(2):134-142.
- BARTNECK, C., C. LÜTGE, A. WAGNER y S. WELSH. (2021). *An Introduction to Ethics in Robotics and AI*, Springer, Suiza: Springer.
- BARTOSZ, B. y J. BARTOSZ. (2019). "Can artificial intelligences be moral agents?", *New Ideas in Psychology*, vol. 54, pp. 101-106.
- BEAUREGARD, L.P. (2021). "Estados Unidos pierde a su atleta estrella para Tokio por una prueba positiva de marihuana", *El País*. Disponible en: <https://elpais.com/deportes/2021-07-07/estados-unidos-pierde-a-su-atleta-estrella-para-tokio-por-una-prueba-positiva-de-marihuana.html>.
- BEHDADI, D. y C. MUNTHE, (2020). "A Normative Approach to Artificial Moral Agency", *Minds & Machines* **30**, 195–218. <https://doi.org/10.1007/s11023-020-09525-8>.
- BELLMAN, R. E. (1978). *An Introduction to Artificial Intelligence: Can Computers Think?*, Estados Unidos de América: Boy & Fraser Publishing Company.
- BERLIN, L. (2012). "Medicolegal: Malpractice and ethical issues in radiology: The incidentaloma", *AJR Am J Roentgenol*, Jan;200(1):W91. Doi: 10.2214/AJR.12.8894.
- BLAKEMORE, S. J. y FRITH, U. (2011). *Cómo aprende el cerebro. Las claves para la Educación*, Barcelona, España: Ariel.
- BOGUSZEWSKI, C.L., DE CASTRO MUSOLINO N.R. y L. KASUKI. (2019). "Management of pituitary incidentaloma", *Best Pract Res Clin Endocrinol Metab*, 33(2):101268. Doi: 10.1016/j.beem.2019.04.002.

- BOSTROM, N. (2009). “Ethical Issues in Advanced Artificial Intelligence”, En: S. SCHNEIDER, ed., *Science Fiction and Philosophy, From Time Travel to Superintelligence*, Reino Unido: Wiley-Blackwell.
- BUCHANAN, A. (2011). *Better than Human: The Promise and Perils of Enhancing Ourselves*, Nueva York, Estados Unidos de América: Oxford University Press.
- CARTON, L., CABÉ, N., MÉNARD, O., DEHEUL, S., CAOUS, A.S., DEVOS, D., COTTENCIN, O. y BORDET, R. (2018). “Pharmaceutical cognitive doping in students: A chimeric way to get-a-head?”, *Thérapie*, 73(4):331-339.
- CHAN, S. y J. HARRIS. (2018). “¿Regeneración o mejoramiento de la cognición?: cuestionamientos éticos”. En: CHAN, S., IBARRA PALAFOX, F. y MdJ MEDINA ARELLANO, *Bioética y Bioderecho. Reflexiones Clásicas y Nuevos Desafíos*, México: Instituto de Investigaciones Jurídicas, UNAM. pp. 19-30. Disponible en: <https://archivos.juridicas.unam.mx/www/bjv/libros/10/4733/1.pdf>.
- CHANDLER, A. J. (2018). “Neurolaw and Neuroethics”, *Cambridge Quarterly of Healthcare Ethics*, Vol. 27, pp. 590–598.
- CHAKRABORTY, P. (2017). “Applied AI is Neither Weak Nor Narrow: Isn't It Time To Change the AI Nomenclature?”, *Analytics India Magazine*. Disponible en línea, en: <https://analyticsindiamag.com/applied-ai-neither-weak-narrow-isnt-time-change-ai-nomenclature/>.
- CHARNIAK, E, y D. MCDERMOTT. (1985). *Introduction to Artificial Intelligence*, Estados Unidos de América: Addison-Wesley.
- CRAWFORD, K. (2021). *Atlas of AI, Power, Politics, and the Planetary Costs of Artificial Intelligence*, Reino Unido: Yale University Press.
- COPELAND, B.J. (2020). “Artificial Intelligence”, En: *Encyclopedia Britannica*, 11 de agosto de 2020. Disponible en: <https://www.britannica.com/technology/artificial-intelligence>.
- DELISI, M. y M.G. VAUGHN. (2015). “The vindication of Lamarck? Epigenetics at the intersection of law and mental health”, *Behav Sci Law*, 33(5):607-28.

- DE MEIRA PENNA, J. O. (1985). "A psychological approach to ethics", *J Relig Health*, Sep;24(3):184-96. Doi: 10.1007/BF01597312.
- DI PIETRO, N.C. y J. ILLES. (2013). "Disclosing incidental findings in brain research: the rights of minors in decision-making", *J Magn Reson Imaging*, 38(5):1009-13. Doi: 10.1002/jmri.24230.
- DURANTE, M. (2017). *Ethics, Law and the Politics of Information, A Guide to the Philosophy of Luciano Floridi*, vol. 18, The International Library of Ethics, Law and Technology, Holanda: Springer.
- EICKENHORST, P., VITZTHUM, K., KLAPP, B.F., GRONEBERG, D. y S. MACHE. (2012). "Neuroenhancement among German university students: motives, expectations, and relationship with psychoactive lifestyle drugs", *J Psychoactive Drugs*, 44(5):418-27.
- ERWIN, C. (2015). "Ethical issues raised by epigenetic testing for alcohol, tobacco, and cannabis", *Behav Sci Law*, 33(5):662-674.
- FARAH, J. M. (2012). "Neuroethics: The Ethical, Legal, and Societal Impact of Neuroscience." *The Annual Review of Psychology*, Vol. 63, pp. 571–591.
- FAUST, H.S. (2008). "Should we select for genetic moral enhancement? A thought experiment using the Moralkinder (MK+) haplotype", *Theoretical Medicine and Bioethics*, 29(6):397-416.
- FRANKE, A.G., GRÄNSMARK, P., AGRICOLA, A., SCHÜHLE, K., ROMMEL, T., SEBASTIAN, A., BALLÓ, H.E., GORBULEV, S., GERDES, C., FRANK, B., RUCKES, C., TÜSCHER, O. y K. LIEB. (2017). "Methylphenidate, modafinil, and caffeine for cognitive enhancement in chess: A double-blind, randomised controlled trial", *Eur Neuropsychopharmacol*, 27(3):248-260.
- FRANKISH, K. y W. M. RAMSEY (eds.), (2014). *The Cambridge Handbook of Artificial Intelligence*, Reino Unido: Cambridge University Press.
- GARCÍA-MOLINA, A. (2012). "Phineas Cage y el enigma del córtex prefrontal," *Elsevier Neurología*, Vol. 27(6), pp. 370—375.

- GAYLIN, W. (1984). "In defense of the dignity of being human", *Hastings Cent Rep*, Aug;14(4):18-22.
- GEELHOED, G.W. (1982). "Drug EM. Management of the adrenal "incidentaloma"", *Surgery*, 92(5):866-74.
- GIURGEA, C. (1972). "Vers une pharmacologie de l'activité integrative du cerveau. Tentative du concept nootrope en psychopharmacologie", *Actualités Pharmacologiques*, 25:115-156.
- GREENE, J. y J. HAITD. (2002). "How (and where) does moral judgment work?", *Trends in Cognitive Sciences*, Vol. 6, pp. 517-523.
- GRANT, J.E., REDDEN, S.A., LUST, K. y S.R. CHAMBERLAIN. (2018). "Nonmedical use of stimulants is associated with riskier sexual practices and other forms of impulsivity", *J Addict Med*, 12(6):474-480.
- GUNTER, T.D. (2015). "Behavioral genetics and the forensic mental health provider: an overview", *Behav Sci Law*, 33(5):598-606.
- GUNTER, T.D. y A.R. FELTHOUS. (2015). "Epigenetics and the Law: Introduction to this issue", *Behav Sci Law*, 33(5):595-597.
- GUNTER, T.D. y R.A. PHILIBERT, R.A. (2015) "Smoking, methylation at AHRR, and recidivism risk in a community correction sample of individuals at high risk for recidivism", *Behav Sci Law*, 33(5):691-700.
- HAITD, J. (2001). "The emotional dog and its rational tail: a social intuitionist approach to moral judgment", *Psychological Review*, Vol. 108, pp. 814-834.
- HAN, S. (2019). "The Connection between Charles Darwin's evolutionary theory of 'heredity of behaviors' and the 19th century neuroscience: The influence of neuroscience on Darwin's overcoming of Lamarck's theory of evolution", *Uisahak*, Apr;28(1):291-350. Doi: 10.13081/kjmh.2019.28.291.
- HAUGELAND, J. (1985). *Artificial Intelligence: The Very Idea*, Estados Unidos de América: MIT Press.

HINCAPIÉ SANCHEZ, J., y MdJ MEDINA ARELLANO. (2019). *Bioética: Teorías y Principios. Enseñanza Transversal en Bioética y Bioderecho: Cuadernos Digitales de Casos*. Disponible en: <https://biblio.juridicas.unam.mx/bjv/detalle-libro/6006-1-bioetica-teorias-y-principios-ensenanza-transversal-en-bioetica-y-bioderecho-cuadernillos-digitales-de-casos>.

IBARRA, F. y A. RAVELO, (2021). “A marchas forzadas: la declaratoria general de inconstitucionalidad en materia de cannabis”, *Nexos. El juego de la Suprema Corte*. Disponible en: <https://eljuegodelacorte.nexos.com.mx/author/frida-ibarra/>.

JOHNSON, D. (2006). “Computer Systems: Moral Entities but not Moral Agents”, *Ethics and Information Technology*, núm. 8.

KAPLAN, J. (2016). *Artificial Intelligence, What Everyone Needs to Know*, Estados Unidos de América: Oxford University Press.

KAPLAN, A. y M. HAENLEIN. (2019). “Siri, siri, in my hand: Who’s the fairest in the land? On the interpretations, illustrations, and implications of artificial intelligence”, *Business Horizons*, núm. 62.

KNIGHT, W. (2019). “La inminente y poderosa simbiosis de los ordenadores cuánticos y la IA”, *MIT Technology Review*, 04 de abril de 2019. Disponible en línea: <https://www.technologyreview.es/s/11028/la-inminente-y-poderosa-simbiosis-de-los-ordenadores-cuanticos-y-la-ia>.

KURZWEIL, R. (1990). *The Age of Intelligent Machines*, Estados Unidos de América: MIT Press.

LINARES, J. (2018). “Un enfoque bioético: Argumentos para la despenalización y posible legalización de la marihuana en México”. En: CHAN, S., IBARRA PALAFOX, F. y MdJ MEDINA ARELLANO, *Bioética y Bioderecho. Reflexiones Clásicas y Nuevos Desafíos*, México: Instituto de Investigaciones Jurídicas, UNAM. pp. 19-30. Disponible en: <https://archivos.juridicas.unam.mx/www/bjv/libros/10/4733/1.pdf>.

MAIER, L.J., LIECHTI, M.E., HERZIG, F. y M.P. SCHAUB. (2013). “To dope or not to dope: neuroenhancement with prescription drugs and drugs of abuse among Swiss

- university students”, *PLoS One*, Nov 13;8(11):e77967. Doi: 10.1371/journal.pone.0077967.
- MAINZER, K. (2020). *Artificial Intelligence — When do machines take over?*, Suiza: Springer.
- McCABE, D.P. y CASTEL, A.D. (2008). “Seeing is believing: the effect of brain images on judgments of scientific reasoning”, *Cognition*, Apr;107(1):343-52.
- MONASTERIO ASTOBIZA, A. (2017). “Ética algorítmica: implicaciones éticas de una sociedad cada vez más gobernada por algoritmos”, *Revista Dilemata*, núm. 24 año 9.
- MORSE J. S. (2018). *Neuroethics: Neurolaw*. Estados Unidos de América: Oxford Handbook Online. Disponible en: <https://www.oxfordhandbooks.com/view/10.1093/oxfordhb/9780199935314.001.0001/oxfordhb-9780199935314-e-45>.
- MORSE, S.J. y ROSKIES, A.L. (2013). *A primer on criminal law and neuroscience. A contribution of the law and neuroscience project*, supported by the MacArthur Foundation, Nueva York, Estados Unidos de América: Oxford University Press.
- NATH, R. y V. SAHU, (2017). “The problem of machine ethics in artificial intelligence”, *AI & Society*, núm. 35.
- NELSON, C.A. (2008). “Incidental findings in magnetic resonance imaging (MRI) brain research”, *J Law Med Ethics*, 36(2):315-9, 213. Doi: 10.1111/j.1748-720X.2008.00275.x.
- NILSSON, N. J. (1998). *Artificial Intelligence: A New Synthesis*, Estados Unidos de América: Morgan Kaufmann.
- PACHOLCZYK, A. (2011). “Moral Enhancement: What Is It and Do We Want It?”, *Law, Innovation and Technology*, 3(2).

- PARDO, M.S. y PATTERSON, D. (2013). *Minds, brains, and law. The conceptual foundations of law and neuroscience*, Nueva York, Estados Unidos de América: Oxford University Press.
- PARENS, E. (2004). "Genetic differences and human identities: on why talking about behavioral genetics is important and difficult", *The Hastings Center Report*, 34:S1–36.
- PARTRIDGE, B.J., BELL, S.K., LUCKE, J.C., YEATES. S. y W.D. HALL. (2011) "Smart drugs "as common as coffee": media hype about neuroenhancement", *PLoS One*, 6(11): e28416. Doi: 10.1371/journal.pone.0028416.
- PERSSON, I. y J. SAVULESCU. (2012). *Unfit for the future: The need for moral enhancement*. Reino Unido: OUP Oxford.
- PETOFT, A. (2015). "Neurolaw: A brief introduction", *Iranian Journal of Neurology*, 14(1), pp. 53-58.
- PHILIBERT, R. y C. ERWIN. (2015). "A review of epigenetic markers of tobacco and alcohol consumption", *Behav Sci Law*, 33(5):675-690.
- PICOZZA, E. (2016). *Neurolaw. An introduction*. Suiza: Springer International Publishing Switzerland and Giappichelli Editore.
- POOLE, D., A. K. MACKWORTH y R. GOEBEL. (1998). *Computational Intelligence: A logical approach*, Estados Unidos de América: Oxford University Press.
- RACINE, E., WALDMAN, S, ROSENBERG, J. e ILLES, J. (2010). "Contemporary neuroscience in the media", *Soc Sci Med*, Aug;71(4):725-33.
- RAMO-FERNÁNDEZ, L., SCHNEIDER, A., WILKER, S. e I.T. KOLASSA. (2015). "Epigenetic alterations associated with war trauma and childhood maltreatment", *Behav Sci Law*, 33(5):701-721.
- RENSMA, R. (2013). "Analytical psychology and the ghost of Lamarck: did Jung believe in the inheritance of acquired characteristics?", *J Anal Psychol*, Apr;58(2):258-277. Doi: 10.1111/1468-5922.12008



- REMBOLD, S. (2014). “‘Human enhancement’? It’s all about ‘body modification’! Why we should replace the term ‘human enhancement’ with ‘body modification’”, *Nanoethics*, 8:307-315.
- RICH, E., y K. KNIGHT. (1991). *Artificial Intelligence*, Estados Unidos de América: McGraw-Hill.
- RIPPON, G. (2019). *The Gendered Brain: The new neuroscience that shatters the myth of the female brain*. Londres, Reino Unido: Random House.
- ROSKIES, A. (2002). “Neuroethics for the New Millenium”, *Neuron*, Vol. 35, pp. 21–23.
- RUSELL, S. J. y P. NORVIG. (2009). *Artificial Intelligence, A Modern Approach*, 3ª. Ed., Estados Unidos de América: Pearson.
- SANDEL, M. J. (2007). *The Case Against Perfection: Ethics in the Age of Genetic Engineering*, Cambridge, MA., Estados Unidos de América: Harvard University Press.
- SCOTT, N.A., MURPHY, T.H. y J. ILLES. (2012). “Incidental findings in neuroimaging research: a framework for anticipating the next frontier”, *J Empir Res Hum Res Ethics*, 7(1):53-7. Doi: 10.1525/jer.2012.7.1.53.
- SHEN, X. F. (2016). “The Overlooked History of Neurolaw”, *Fordham Law Review*, p. 667.
- SEARLE, J. (2003). *Minds, Brain and Sciece*, Estados Unidos de América: Harvard University Press.
- (1980). “Minds, brain, and programs”, *The Behavioral and Brain Sciences*, núm. 3.
- SPRANGER, M. T. (2012). *International Neurolaw. A comparative analysis*, Londres, Reino Unido: Springer.
- STIP, E., MIRON, J.P., NOLIN, M., LETOURNEAU, G., BERNAZZANI, O., CHAMELIAN, L., BOILEAU, B., GUPTA, M., LUCK, D. y O. LUNGU. (2019).



- “Incidentaloma discoveries in the course of neuroimaging research”, *Can J Neurol Sci.*, 46(3):275-279. Doi: 10.1017/cjn.2018.397.
- SZUCS, B., y J. P. A. IOANNIDIS, (2017). “Empirical assessment of published effect sizes and power in the recent cognitive neuroscience and psychology literature”, *Plos Biology*. Disponible en: <https://doi.org/10.1371/journal.pbio.2000797>.
- TAMATEA, A.J. (2015). “‘Biologizing’ psychopathy: Ethical, legal, and research implications at the interface of epigenetics and chronic antisocial conduct”, *Behav Sci Law*, 33(5):629-643.
- TAYLOR, J. S., HARP, J. A., y ELLIOTT, T. (1991). “Neuropsychologists and neurolawyers”, *Neuropsychology*, 5(4), pp. 293–305.
- TERRY, N. (2015). “Developments in genetic and epigenetic data protection in behavioral and mental health spaces”, *Behav Sci Law*, 33(5):653-661.
- THOMAS, C.R. (2015). “Epigenetics and child psychiatry: ethical and legal issues”, *Behav Sci Law*, 33(5):644-652.
- TOLLEFSBOL, T.O. (2019). *Transgenerational epigenetics*. 2nd ed. Cambridge, MA, Estado Unidos de América: Academic Press.
- TOLLON, F. (2019). “Moral Agents or Mindless Machines? A Critical Appraisal of Agency in Artificial Systems”, *Hungarian Philosophical Review*, vol. 4 núm. 63.
- UNDERWOOD, E. (2012). “Neuroethics. When a brain scan bears bad news”, *Science*, 338(6106):455. Doi: 10.1126/science.338.6106.455.
- VAN GESTEL, S y VAN BROECKHOVEN, C.V. (2003). “Genetics of personality: Are we making progress?”, *Molecular Psychiatry*, Vol. 8, pp. 840–852
- VÉLIZ, C. (2021). “Moral zombies: why algorithms are not moral agents”, *AI & Society*, 36, pp. 487-497.
- VINCENT, N.A. (2013). *Neuroscience and legal responsibility*, Nueva York, Estados Unidos de América: Oxford University Press.

WINSTON, P. H. (1992). *Artificial Intelligence*, Estados Unidos de América: Addison-Wesley.

## 2. Jurisprudencia y sentencias relevantes

SUPREMA CORTE DE JUSTICIA DE LA NACIÓN (SCJN), México. Amparo en Revisión 237/2014 (Caso SMART).

—— Amparo en Revisión 623/2017.

—— Amparo en Revisión 1115/2017.

—— Amparo en Revisión 1163/2017.

—— Amparos en Revisión 547/2018 y 548/2018.

—— Tesis “PROHIBICIÓN ABSOLUTA DEL CONSUMO LÚDICO DE MARIHUANA. NO ES UNA MEDIDA PROPORCIONAL PARA PROTEGER LA SALUD Y EL ORDEN PÚBLICO”. Número de Registro: 2019382. Localización: [J]; 10a. Época; 1a. Sala; Semanario Judicial de la Federación; 1a./J. 9/2019 (10a.); Publicación: Viernes 22 de Febrero de 2019 10:24 h.

—— Tesis “PROHIBICIÓN ABSOLUTA DEL CONSUMO LÚDICO DE MARIHUANA. ÉSTA PERSIGUE FINALIDADES CONSTITUCIONALMENTE VÁLIDAS.” Número de Registro: 2019381 Localización: [J]; 10a. Época; 1a. Sala; Semanario Judicial de la Federación; 1a./J. 7/2019 (10a.); Publicación: Viernes 22 de Febrero de 2019 10:24h.

—— Tesis “INCONSTITUCIONALIDAD DE LA PROHIBICIÓN ABSOLUTA AL CONSUMO LÚDICO O RECREATIVO DE MARIHUANA PREVISTA POR LA LEY GENERAL DE SALUD.” Número de Registro: 2019365. Localización: [J]; 10a. Época; 1a. Sala; Semanario Judicial de la Federación; 1a./J. 10/2019 (10a.); Publicación: viernes 22 de Febrero de 2019 10:24h.

—— Tesis “DERECHOS DE TERCEROS Y ORDEN PÚBLICO. CONSTITUYEN LÍMITES EXTERNOS DEL DERECHO AL LIBRE DESARROLLO DE LA PERSONALIDAD.” Número de Registro: 2019359. Localización: [J]; 10a. Época; 1a.

Sala; Semanario Judicial de la Federación; 1a./J. 6/2019 (10a.); Publicación: Viernes 22 de Febrero de 2019 10:24h.

—— Tesis “DERECHO A LA PROTECCIÓN DE LA SALUD. DIMENSIONES INDIVIDUAL Y SOCIAL.” Número de Registro: 2019358. Localización: [J]; 10a. Época; 1a. Sala; Semanario Judicial de la Federación; 1a./J. 8/2019 (10a.); Publicación: Viernes 22 de Febrero de 2019 10:24h.

—— Tesis “DERECHO AL LIBRE DESARROLLO DE LA PERSONALIDAD. SU DIMENSIÓN EXTERNA E INTERNA.” Número de Registro: 2019357. Localización: [J]; 10a. Época; 1a. Sala; Semanario Judicial de la Federación; 1a./J. 4/2019 (10a.); Publicación: Viernes 22 de Febrero de 2019 10:24h.

—— Tesis “DERECHO AL LIBRE DESARROLLO DE LA PERSONALIDAD. LA PROHIBICIÓN PARA EL AUTOCONSUMO DE MARIHUANA CONTENIDA EN LA LEY GENERAL DE SALUD INCIDE PRIMA FACIE EN EL CONTENIDO DE DICHO DERECHO FUNDAMENTAL.” Número de Registro: 2019356. Localización: [J]; 10a. Época; 1a. Sala; Semanario Judicial de la Federación; 1a./J. 3/2019 (10a.); Publicación: Viernes 22 de Febrero de 2019 10:24h.

—— Tesis “DERECHO AL LIBRE DESARROLLO DE LA PERSONALIDAD. BRINDA PROTECCIÓN A UN ÁREA RESIDUAL DE LIBERTAD QUE NO SE ENCUENTRA CUBIERTA POR LAS OTRAS LIBERTADES PÚBLICAS.” Número de Registro: 2019355. Localización: [J]; 10a. Época; 1a. Sala; Semanario Judicial de la Federación; 1a./J. 5/2019 (10a.); Publicación: Viernes 22 de Febrero de 2019 10:24h.

### 3. Legislación

Constitución Política de los Estados Unidos Mexicanos (1917, última reforma 28-05-2021). Disponible en:

[http://www.diputados.gob.mx/LeyesBiblio/pdf\\_mov/Constitucion\\_Politica.pdf](http://www.diputados.gob.mx/LeyesBiblio/pdf_mov/Constitucion_Politica.pdf).

Diario Oficial de la Federación, (15 julio 2021) Declaratoria General de Inconstitucionalidad 1/2018. Disponible en: [https://www.dof.gob.mx/nota\\_detalle.php?codigo=5623991&fecha=15/07/2021](https://www.dof.gob.mx/nota_detalle.php?codigo=5623991&fecha=15/07/2021).

Ley General de Salud (1984, última reforma 01-06-2021). Disponible en: [http://www.diputados.gob.mx/LeyesBiblio/pdf\\_mov/Ley\\_General\\_de\\_Salud.pdf](http://www.diputados.gob.mx/LeyesBiblio/pdf_mov/Ley_General_de_Salud.pdf).

#### 4. Páginas Web

WADA, (2021). *World Anti-doping Code International Standard. Prohibited List*. Disponible en: [https://www.wada-ama.org/sites/default/files/resources/files/2021list\\_en.pdf](https://www.wada-ama.org/sites/default/files/resources/files/2021list_en.pdf).

Dana Foundation, (2002). *Neuroethics: Mapping the Field*. Disponible en: <https://dana.org/article/neuroethics-mapping-the-field/>.