Autonomía kantiana en la toma de decisiones algorítmicas, de los algoritmos *deep learning*: un análisis desde la *Crítica de la razón práctica*

Arnol David Hernández Martínez

Trabajo de Grado para Optar por el Título de

Filósofo

Director/a

Jorge Francisco Maldonado Serrano

Doctor en filosofía

Universidad Industrial de Santander

Facultad de Ciencias Humanas

Escuela de Filosofía

Bucaramanga

2026

Dedicatoria

A mi abuelo, quien puso el primer libro de filosofía y aritmética en mis manos. A mi abuela, quien me llevó a conocer otras tierras y con ello se abrió mi visión del mundo. Y por supuesto, a mis padres por otorgarme libertad desde niño, lo cual potenció mi curiosidad intelectual y horizonte de experiencias.

Agradecimientos

Especial agradecimiento al doctor en filosofía Jorge Francisco Maldonado Serrano, a la Mg. en historia Lina Constanza Díaz Boada, y a la Mg. en filosofía Gretel Gesid Vega por orientarme en la elaboración de este trabajo de grado.

Tabla de Contenido

Pág.

Introducción 8
1. Autonomía kantiana en la crítica de la razón práctica 13
1.1 Introducción del capítulo 13
1.2 De la idea de una crítica de la razón práctica....13
1.3 Ley moral como ratio cognoscendi, libertad como ratio essendi y la posibilidad de una voluntad autónoma....14
1.4 La forma de la ley en el imperativo categórico, dignidad, respeto y postulados 20
1.5 Conclusión 24
2. Deep learning y autonomía funcional 26
2.1 Introducción 26
2.2 Definición de inteligencia artificial, machine learning y deep learning 26
2.3 Redes neuronales artificiales: estructura y funcionamiento....27
2.4 La aparente autonomía en la toma de decisiones algorítmicas y principales limitaciones del deep learning....30
2.5 Conclusión 32
3. Comparación entre autonomía moral kantiana vs autonomía funcional algorítmica 33
3.1 Introducción 33
3.2 La decisión es algo más que un simple outputs 33
3.3 Autonomía vs entrenamiento 34
3.4 Causalidad y libertad....35
3.5 Normatividad y dignidad 37
3.6 Conclusiones....38
4. Conclusiones....38
Referencias Bibliográficas....40

Glosario

Algoritmo: conjunto de pasos definidos que permiten resolver un problema.

Función de pérdida (*loss function*): medida que permite calcular el error entre las predicciones del modelo y los valores reales.

Función matemática: expresión que relaciona variables para modelar un comportamiento o proceso.

Optimización: proceso de minimización del error de la función de pérdida

Pesos (*weights*): valores de multiplicación asignados a las entradas, según la importancia que se establezca para el modelo.

Retropropagación (*backpropagation)*: Algoritmo que calcula el error y luego lo propaga hacia atrás, con esto se ajustan los parámetros para producir mejores resultados.

Resumen

Título: Autonomía kantiana en la toma de decisiones algorítmicas, de los algoritmos *deep learning*: un análisis desde la *Crítica de la razón práctica**

Autor: Arnol David Hernández Martínez**

Palabras Clave: Autonomía, *deep learning*, Kant, razón, libertad.

Descripción: Los avances tecnológicos en inteligencia artificial y *deep learning* han traído agentes artificiales sofisticados, capaces de realizar tareas que tradicionalmente se habían atribuido a la inteligencia humana. Por esto, se ha presentado una ambigüedad terminológica respecto de la autonomía, toma de decisiones y racionalidad; confusiones conceptuales que generan implicaciones éticas problemáticas. En este contexto, el presente trabajo de grado se plantea analizar si es posible atribuir autonomía kantiana a la aparente toma de decisiones de los algoritmos de aprendizaje profundo, los cuales están inspirados en las redes neuronales biológicas de los seres humanos.

En consecuencia, esta investigación se desarrolla desde un enfoque cualitativo que permite la interpretación y comparación de textos filosóficos y técnicos. En el primer capítulo, se reconstruye la moral kantiana a partir del concepto de autonomía y se destaca su carácter racional, autolegislativo, acorde con principios universales y siempre libre de condiciones empíricas. En el segundo, se aborda el funcionamiento del *deep learning* y se muestra principalmente su dependencia de datos, estructuras rígidas previamente fijadas y procesos de optimización. En el tercero, se comparan ambos marcos conceptuales y se responde a la posibilidad de atribuir autonomía en sentido kantiano a estos sistemas.

Como resultado, se define que los algoritmos *deep learning* no son autónomos en sentido kantiano porque carecen de racionalidad, representación de leyes, obligación del deber, actuación conforme a principios universales y libertad; ellos operan en un marco causal e instrumental. Por tanto, esta tesis aporta claridad conceptual y mejor comprensión de las implicaciones éticas de la IA y en particular del DL.

---

* Trabajo de Grado

** Facultad de Ciencias Humanas. Escuela de Filosofía. Director: Jorge Francisco Maldonado Serrano. Doctor en Filosofía UAM.

Abstract

Title: Autonomy in algorithmic decision-making: a Kantian analysis from the critique of practical reason in deep learning*

Author: Arnol David Hernández Martínez**

Key Words: Autonomy, deep learning, Kant, reason, freedom.

Description: Technological advances in artificial intelligence and deep learning have led to the creation of artificial agents capable of performing tasks traditionally attributed to human intelligence. This has resulted in terminological ambiguity regarding autonomy, decision making, and rationality conceptual confusions that generate problematic ethical implications. In this context, this thesis aims to analyze whether it is possible to attribute Kantian autonomy to the apparent decision making of deep learning algorithms, which are inspired by the biological neural networks of human beings.

Consequently, this research is developed from a qualitative approach that allows for the interpretation and comparison of philosophical and technical texts. In the first chapter, Kantian morality is reconstructed based on the concept of autonomy, highlighting its rational, selflegislative character, its adherence to universal principles, and its freedom from empirical conditions. In the second chapter, the functioning of deep learning is addressed, primarily demonstrating its dependence on data, preestablished rigid structures, and optimization processes. In the third section, both conceptual frameworks are compared, and the possibility of attributing Kantian autonomy to these systems is addressed.

As a result, it is determined that deep learning algorithms are not autonomous in the Kantian sense because they lack rationality, the representation of laws, the obligation of duty, action in accordance with universal principles, and freedom; they operate within a causal and instrumental framework. Therefore, this thesis provides conceptual clarity and a better understanding of the ethical implications of AI, and particularly of DL.

*Degree Work.

**Faculty of Humanities. School of Philosophy. Director: Jorge Francisco Maldonado Serrano. PhD in Philosophy, UAM.

Introducción

Para efectos de esta investigación, se aborda la noción de autonomía estipulada por Immanuel Kant en *La Crítica de la Razón Práctica* planteada como norma según la cual la voluntad de la razón establece principios universales sobre una ley moral sin depender de influencias externas, en palabras de Kant "La autonomía de la voluntad es el único principio de todas las leyes morales y de los deberes que les corresponden" (CRPr, A59). En este marco conceptual, se pretende identificar si los algoritmos modernos *deep learning* pueden considerarse autónomos en sentido kantiano.

Ahora bien, en el plano técnico un algoritmo produce un resultado (*output*) según el cálculo programado en el código. En programación convencional la decisión radica en el programador. No obstante, en *Machine Learning*, hay un nuevo paradigma porque la decisión no depende directamente de quién programa, sino del ajuste de funciones matemáticas, según Centeno (2019) para "desarrollar técnicas que permitan a los ordenadores aprender" (p. 1), en este contexto, se ajustan pesos de ecuaciones conforme a la estadística y probabilidad para que la red neuronal produzca un resultado inesperado.

En consecuencia, lo anterior ha incitado debates sobre la autonomía algorítmica y las implicaciones éticas que se derivan. Así las cosas, esta situación plantea una dificultad conceptual fundamental: si la autonomía en sentido kantiano implica autolegislación racional, y capacidad de actuar conforme a principios universales, surge entonces una cuestión filosófica relevante: ¿Puede atribuirse autonomía a los procesos de toma de decisiones de los algoritmos *deep learning* desde la concepción kantiana desarrollada en la *Crítica de la razón práctica*?

A partir de estas circunstancias, el problema de investigación se formula como la tensión entre una concepción filosófica rigurosa de la autonomía y el uso contemporáneo del término en el ámbito tecnológico, y consiste en analizar si los procesos de toma de decisiones de los algoritmos de aprendizaje profundo pueden ser considerados autónomos en sentido kantiano o si, por el contrario, tal atribución es únicamente de carácter funcional. Por cierto, el problema de la agencia y la responsabilidad ética en el área de la inteligencia artificial ha sido abordada por diversos autores desde distintos enfoques, sin que se haya alcanzado claridad sobre su estatuto moral.

Por una parte, Thomas Metzinger (2021) plantea el reto del estudio de una fenomenología sintética así: "risking the creation of artificial consciousness is highly problematic from an ethical perspective, because it may lead to artificial suffering and a consciously experienced sense of self in autonomous, intelligent systems. (p.21), es una propuesta investigativa ante la posibilidad de creación de sistemas autónomos conscientes, pero justo aquí se evidencia la limitación de este autor. De modo que, sus análisis se centran en la evitación de sufrimiento gratuito por parte de las máquinas, pero no establece una respuesta concreta sobre la posibilidad de una conciencia de máquina, más allá de que entiende que aún ellas no tienen conciencia ni sufrimiento.

Con referencia a Nick Bostrom (2003), una de las preocupaciones que se encuentran en sus estudios es acerca de los riesgos que se presentan en una superinteligencia de conformidad con los sesgos propiamente de las máquinas, pero también de los intereses humanos particulares que existen en su programación, por ello menciona: "include the risk of failure to give it the supergoal of philanthropy" (p.5), es válido mencionar que, queda sin resolver finalmente la posibilidad de una ética deontológica en las superinteligencias, si bien se habla de fines filantrópicos, no se hace uso de la filosofía moral kantiana y su carácter universal.

Desde la fenomenología se ha discutido el asunto tecnológico. Por eso, Hubert Dreyfus argumentó que una especie de máquina no puede ser como la inteligencia humana a causa de que carecen de corporalidad, experiencia situada y compresión holística del mundo. En la obra *what computer can't do* (1972) afirma: "we never encounter meaningless bits (…) but only facts which are already interpreted (…) Human experience is only intelligible when organized in terms of a situation in which relevance and significance are already given" (p. 200), aporta a la discusión el asunto cognitivo, sin limitarse al cálculo y manipulación de símbolos. No obstante, no llega a cuestionar la normatividad ni la ética detrás de estos sistemas, por tanto, su crítica se queda en lo fenomenológico y no se extiende a lo trascendental.

En esta línea, la filosofía de la mente también ha estudiado el hecho tecnológico. El autor John Searle (1980) le dio fama a un experimento que se conoce como la *habitación china*: "the machine takes in Chinese stories and questions about them as input (…) and it gives out Chinese answers as outputs. (p. 420), con este determina que la máquina no tiene capacidad de razonamiento, sino que es un artefacto que se sirve de funciones matemáticas a la que le damos una entrada $\boldsymbol{X}$ que se relaciona con un elemento de una salida $\boldsymbol{Y}$. A pesar de todo, el autor no habla de autonomía en sentido kantiano como autolegislación racional.

Luciano Floridi (2013) ha problematizado los espacios digitales donde algoritmos divulgan información. Para este autor, los agentes artificiales carecen de racionalidad, pero tienen poder de actuación e impacto moral, y afirma: In the re-ontologized infosphere, any information agent has an increased power not only to gather and process personal data, but also to control and protect them." (p.236), con esto se puede discutir en quién cae la responsabilidad ética respecto a la implementación de un determinado algoritmo. De todas formas, no explica por qué los algoritmos

no pueden ser autónomos en un sentido fuerte como el kantiano, ni analiza la estructura interna del algoritmo *deep learning*.

Estas propuestas mencionadas presentan tensiones entre sí respecto de lo que implica la autonomía y lo que se entiende por sistemas artificiales. John Searle limita a las máquinas a relaciones sintácticas y Hubert Dreyfus la refuerza porque implementa la experiencia situada humana. Por el contrario, Thomas Metzinger está preocupado por una posibilidad de estado fenomenológico y capacidad de sufrimiento por parte de la máquina. Por el lado de Nick Bostrom el debate se lleva al terreno a riesgos y regulación de estos sistemas. Finalmente, Luciano Floridi alude a que la IA es un agente informacional que media decisiones morales humanas. Aunque existen diferencias en los autores, todos se encuentran en el hecho de problematizar la agencia de máquinas sin hablar en específico del *deep learning* y el fundamento de una autonomía moral universal y racional.

Con todo, se justifica la importancia de este trabajo de grado para la filosofía de la tecnología, debido a que se ha evidenciado que el término de autonomía carece de precisión conceptual en el ámbito tecnológico. En primera instancia, permite clarificar para evitar confusiones sobre el concepto y su uso en el ámbito tecnológico, con lo cual se aporta rigor filosófico. En segunda, se hace necesario delimitar el alcance real que tienen los algoritmos *deep learning* en su agencia, toma de decisiones y responsabilidades éticas; con lo cual se permite una mejor evaluación para evitar atribuciones indebidas que desdibuje los criterios. Así mismo, se trae a la filosofía al corazón de los avances tecnológicos al crearse la conexión entre teoría filosófica clásica y problemas actuales.

En este orden de ideas, conviene enunciar la metodología que se emplea para dar respuesta a la pregunta de esta investigación: enfoque metodológico, tipo de investigación, método y corpus

de análisis. De esta manera, el enfoque es cualitativo porque la investigación se fundamenta en la interpretación de textos filosóficos y técnicos, con esto se logra una muy buena comparativa conceptual desde un lado tecnológico a otro filosófico, sin la necesidad de recurrir a simulaciones algorítmicas *deep* learning. El corpus se centra en la obra *Crítica de la razón prácticia* al desarrollar el concepto de autonomía de forma sistemática, y literatura académica sobre el aprendizaje profundo del *MIT press*. Así mismo, se traen otros artículos a la comparativa para abordar la relación entre la técnica, agencia y moralidad.

Es así como, este trabajo se divide en tres capítulos. En el capítulo uno, se reconstruye la filosofía moral kantiana estipulada en la *Crítica de la razón práctica* a partir de la noción de autonomía. En el segundo, se explica el funcionamiento de los algoritmos *deep learning* en relación con su aparente toma de decisiones, en función de sus estructuras basadas en datos y funciones matemáticas. En el tercero, se establece la comparación entre la autonomía kantiana y la aparente toma de decisiones algorítmicas, con el objetivo de examinar las condicionas para que se le pueda atribuir autonomía en sentido kantiano.

Finalmente, para una cabal comprensión de esta investigación es importante tener en cuenta las abreviaturas usadas. Antes que todo, a lo largo de este trabajo a veces se escribe *deep learning*, aprendizaje profundo o DL; en todo caso, siempre se hace referencia al mismo tipo de algoritmo basado en las redes neuronales artificiales. Así mismo sucede con inteligencia artificial (IA) y con *machine learning* a veces presentado en su traducción al español como aprendizaje de máquina, o en sus siglas en inglés (ML). De igual manera, las obras de Kant no escapan de estas abreviaciones debido a su citación clásica: *Crítica de la razón pura* (CrP), *Fundamentación para una metafísica de las costumbres* (FMC) y *Crítica de la razón práctica* (CRPr).

1. Autonomía kantiana en la *crítica de la razón práctica*

1.1 Introducción del capítulo

El hilo argumentativo de este capítulo presenta la siguiente dirección: en primer lugar, se plantea el problema de la razón práctica de cómo es posible que la razón determine la voluntad y se presenta a la ley moral como un hecho *faktum* de la razón. En segundo, se explica cómo la ley moral es la *ratio cognoscendi* de la libertad y la libertad *ratio essendi* de la ley moral que hace posible una voluntad autónoma. En tercer, se expresa el imperativo categórico en sus tres formas para excluir toda heteronomía como fundamentación moral, y, además, mostrar en ellos la fundamentación de la dignidad y el respeto; así mismo, se señalan los postulados como una exigencia moral de la razón práctica.

1.2 De la idea de una crítica de la razón práctica

En la *Crítica de la Razón Práctica* Kant se formula la pregunta de ¿Cómo es posible que la razón determine la voluntad independientemente de la experiencia? Por todo esto, Kant se pronuncia sobre la idea de una crítica de la razón práctica de la siguiente manera: “aquí la primera cuestión es, pues, si la razón pura basta por sí sola para la determinación de la voluntad, o si sólo cuando está empíricamente condicionada puede ser un fundamento determinante” (CRPr, A30). Por esto, se analiza si a parte del uso teórico de la razón, esta puede tener un fundamento práctico.

En consecuencia, Kant propone que la ley moral se presenta a la conciencia como hecho irreductible y lo justifica así “porque no se le puede deducir de datos precedentes de la razón (…) sino porque ella se nos impone por sí misma como proposición sintética a priori” (CRPr, A56). A saber, no es algo que se pueda demostrar o deducir porque no se puede probar empíricamente, simplemente, se experimenta como obligación en el sujeto racional. De manera que, existe una

obligación en el sujeto a reconocer que la voluntad no se determina por causas externas, sino por esa misma ley moral que reconoce como válida, la cual lo hace sentir obligado a cumplir tal obligación moral absolutamente impuesta por la razón. En este sentido, la experiencia del deber pone de manifiesto la libertad del sujeto y la posibilidad de pensar la autonomía como autolegislación racional.

1.3 Ley moral como *ratio cognoscendi*, libertad como *ratio essendi* y la posibilidad de una voluntad autónoma

Inicialmente se planteó el hecho de la ley moral como algo *faktum* y se esbozó que la experiencia del deber pone de manifiesto la libertad del sujeto racional y la posibilidad de pensar la autonomía como autolegislación racional. A todo esto, conviene dar claridad sobre lo condicionado e incondicionado, con el fin de explicar con detalle lo ya mencionado, a partir de la buena voluntad en relación con la distinción entre el noúmeno y fenómeno. Así, se mostrará cómo la autonomía se fundamenta en una voluntad perteneciente al mundo inteligible, y que permite pensarla como libre e independiente de todo factor empírico o de inclinaciones.

En la filosofía de Immanuel Kant, se piensa lo condicionado y lo incondicionado como dos órdenes de consideración, no como dos realidades independientes. En el asunto de lo empírico, lo que se aparece se encuentra bajo condiciones de la causalidad natural, intrínseca al mundo fenoménico. Pero la razón se ve llevada a pensar algo más que el mundo de los fenómenos, por ello no se detiene en lo condicionado, al contrario, exige pensar lo incondicionado.

En este sentido, se puede hacer el rastro en la filosofía kantiana desde la *Fundamentación de la metafísica de las costumbres*, donde afirma que, “no es posible pensar nada dentro del mundo, ni después de todo tampoco fuera del mismo, que pueda ser tenido por bueno sin restricción alguna,

salvo una buena voluntad" (FMC, A1). Eso sí, con esto no se demuestra la libertad, aunque ya se introduce la idea de una voluntad que no tiene su valor en condiciones externas, sino conforme a una ley moral.

En todo caso, el planteamiento de la obra mencionada encuentra un desarrollo más preciso en la *Crítica de la razón práctica*, donde para su búsqueda de lo condicionado e incondicionado realiza el siguiente análisis lógico:

> La determinación de la causalidad de los seres en el mundo de los sentidos como tal nunca podía ser incondicionada; sin embargo, tiene que haber algo incondicionado para toda serie de condiciones, y, por lo tanto, también una causalidad que se determine totalmente por sí misma. (CRPr, A84)

Con esto, es importante comprender al ser racional desde estas dos perspectivas. Por un lado, en el mundo fenoménico es donde se hace posible la experiencia y con ello el ser humano se ve sometido a la causalidad natural y las inclinaciones. Por otro, en el mundo nouménico es donde se hace pensable la libertad, porque el sujeto puede concebirse como libre y autónomo bajo leyes racionales.

Ahora bien, esta libertad que ahora se presenta como pensable no es aún cognoscible, por esto, es adecuado mostrar de qué modo la razón práctica da conciencia de la libertad en sentido práctico, a saber, mediante la ley moral. Así, Kant propone que:

> la ley moral determina objetiva e inmediatamente la voluntad en el juicio de la razón; pero la libertad, cuya causalidad es determinable sólo mediante la ley, consiste, precisamente, en que limita todas las inclinaciones y, por lo tanto, también la estima de la persona, a la condición de la observancia de su pura ley. (CRPr, A140)

Es importante dejar claro qué significa la determinación de la voluntad antes de relacionar la cita anterior con la *ratio cognoscendi*. Precisamente, tal determinación es una causalidad práctica regida inmediatamente por la ley moral, independiente de cualquier factor empírico. Por consiguiente, la voluntad se determina por la representación de la ley misma, según un principio puramente racional.

Aquí, es pertinente precisar que, la ley moral es la *ratio cognoscendi*; es decir, la razón de conocimiento de la libertad (saben los seres racionales que son libres dentro de una conciencia racional del deber). Lo dicho hasta aquí supone que, la libertad no se conoce de manera directa, específicamente se revela en la conciencia del deber. En efecto, una vez se ha planteado la ley moral como *ratio cognoscendi* de la libertad, ahora es consecuente, mostrar en qué sentido la libertad es *ratio essendi* de la ley moral.

Indiscutiblemente, se puede inferir que la libertad es la *ratio essendi* de la ley moral, porque la única manera de afirmar que esta última se da, es si se presupone que la voluntad no está determinada por la causalidad natural. Dado que, si se toma, por ejemplo: a un sujeto sometido inevitablemente a las leyes del mundo sensorial, donde no tiene capacidad de arbitrio, debido a que todas sus acciones se encuentran determinadas necesariamente por inclinaciones y causas empíricas, se excluye inmediatamente la posibilidad de la obligación moral. En este orden de ideas, la moralidad plantea que el sujeto debe pensarse inmerso en algo distinto a la mera causalidad natural.

Es así como Kant afirma: “La ley moral es en realidad una ley de la causalidad mediante la libertad” (CRPr, A82). Cierto es que, si la ley moral es una ley de la causalidad mediante la libertad, la libertad entonces es condición de posibilidad de la ley moral, en tanto que, sin ella

ningún sujeto se vería obligado, ni la voluntad de este se determinaría conforme a principios racionales.

De lo anterior resulta que, existe una relación de complemento: la ley moral es la razón de conocer la libertad, libertad que a su vez constituye la razón de ser de la ley moral; siendo su condición de posibilidad, por ello, vale la pena introducir la voluntad en esta relación de complemento para comprenderla en su sentido práctico. Si se tiene en cuenta lo anterior, se llega a la noción de autonomía, que permite conectar todo, y explica cómo es posible una voluntad libre. En otras palabras, la autonomía expresa el modo en que la voluntad libre se determina a sí misma conforme a la ley moral, en cuanto manifestación de la relación entre la *ratio essendi* y *ratio cognoscendi*.

Lo más importante, es ver la autonomía como la característica fundamental de la voluntad, en tanto que, la voluntad se da a sí misma la ley. Es relevante dejar claro que, la ley que ella misma se da es desde lo racional, no existe una determinación externa proveniente del mundo de los fenómenos. Este punto, se sustenta desde Kant así: "la autonomía es el principio moral mediante el cual la voluntad determina la acción (...) este hecho está conectado inseparablemente con la conciencia de la libertad de la voluntad" (CRPr, A72), con esto queda sustentado que autonomía y libertad son dimensiones correlativas en un espacio especial de la razón práctica.

Cabe precisar que de lo dicho se infiere una autolegislación, empero, se debe evitar relacionar esto con una capacidad del sujeto de actuar según deseos e inclinaciones propias. En todo momento, el filósofo busca máximas que valgan como ley universal. Por supuesto que, un sujeto autónomo no es quien hace lo que quiere, su accionar debe ser conforme a principios que sus estructuras racionales reconocen como universalmente válidos y necesarios. Implícitamente se encuentra el desinterés personal en la acción moral en la siguiente cita de Kant:

> Un principio que se funda solamente en la condición subjetiva de la receptibilidad de un placer o un displacer (…) puede servir como máxima para el sujeto que la posee, pero no como ley (porque le falta la necesidad objetiva que debe ser conocida a priori), así que tal principio no puede proporcionar nunca una ley práctica" (CRPr, A40)

Lo cual significa que, si el principio depende del placer o displacer pierde su carácter universal, por tanto, no puede constituirse como ley, entonces no cabe duda de que la moral excluye el interés como fundamento de la acción. En definitiva, se hace explícito que el desinterés personal se sigue de la universalidad de la ley, es el seguir la norma por la norma misma y solo mediante este precepto se puede pensar en la moralidad, al conservarse el carácter universal.

Así pues, se vislumbra una aparente tensión, una aparente contradicción entre ser libre y seguir la ley moral. En un primer momento, puede ser paradójico comprender la perspectiva de que la autonomía haga posible la obligación moral. Con todo, esta tensión tiene que ver con la doble perspectiva del ser humano, ya tratada en cuestión, y debe ser analizada desde allí. De este modo, el ser humano racional en su dualidad: en primer lugar, tiene que verse exento de estar enteramente sometido a una causalidad natural en el mundo sensible; en segundo, considerarse obligado por una ley, en cuanto que, perteneciente al mundo nouménico. De ahí que, la obligación moral no contradice la libertad, al contrario, la presupone y manifiesta en sentido práctico, dado que, la voluntad es racional y vinculada al sujeto sin imposiciones externas.

En este sentido, Kant menciona que, en el ser humano "la ley tiene la forma de un imperativo" porque su voluntad, aunque racional no es santa, y podría oponerse a la ley moral. Por ello, la relación entre ley y voluntad se expresa como obligación, esto se justifica en Kant como una "coacción (…) impuesta por la mera razón" (CRPr, A57)

Sin embargo, dicha coacción no se presenta como imposición externa que impida la libertad. Por el contrario, esta coerción se presenta en el interior del sujeto racional, y la libertad se mantiene porque incluso en la obediencia a una ley moral, esta proviene de la razón misma. En efecto, la autonomía no elimina la obligación, sino que, al contrario, la hace posible en la medida en que la autonomía fundamenta tal obligación.

En esta línea, se puede enunciar con precisión que la autonomía se presenta como único principio de la moralidad. En caso contrario, cuando la voluntad se determina por cualquier contenido empírico, se pierde el carácter universal y necesario propio de la moralidad. Por ende, Kant lo expresa concretamente así:

> La autonomía de la voluntad es el único principio de todas las leyes morales y de los deberes que les corresponden; por el contrario, toda heteronomía del arbitrio no sólo no funda obligación alguna, sino que es contraria a este principio y a la moralidad de la voluntad. El único principio de la moralidad consiste en la independencia de toda materia de la ley (es decir, de un objeto deseado) y, al mismo tiempo, en la determinación del arbitrio mediante la mera forma legislativa universal de la cual una máxima debe ser capaz. (CRPr, A59)

Todo esto confirma que, la autonomía como autolegislación universal no es relativismo ni tampoco subjetivismo: por lo que, no se sigue que cada uno crea su moral, sino que la razón, en tanto que universal; legisla universalmente desde esta lógica kantiana. Es así como, la autonomía no es individualismo, y en este orden de ideas, Kant defiende: “la ley de esta autonomía es la ley moral (…) la ley fundamental de una naturaleza suprasensible y de un mundo de entendimiento puro” (CRPr, A75), con lo cual se refuerza el carácter objetivo, universal y necesario; y se muestra

que la autonomía no es algo propio de un sujeto particular, sino de la expresión de la racionalidad misma.

Con esto queda sustentado, que la autonomía es el punto donde se expresa la convergencia entre la ley moral y la libertad: a saber, entre la *ratio cognoscendi y ratio essendi*, de tal manera que la ley moral permite conocer la libertad en sentido práctico, y esta última fundamenta las condiciones de posibilidad de la primera: en la expresión de una voluntad racional autónoma. Así, el sujeto se reconoce como libre y autor de la ley que lo obliga, estableciendo en la autolegislación racional el fundamento último de la moralidad.

1.4 La forma de la ley en el imperativo categórico, dignidad, respeto y postulados

En el acápite anterior se estableció la autonomía como principio de la moralidad, ahora es necesario mostrar la forma bajo la cual la ley moral determina la acción. A continuación, se esclarece el principio que determina la voluntad de manera universal y necesaria. En este sentido, en el imperativo categórico se expresa la forma de la ley moral, aquí se encuentra el criterio de validez de las máximas. Así mismo, una vez expuesto y explicado el imperativo categórico, es posible contraponerlo a la heteronomía, derivar las nociones de dignidad y respeto, y señalar los postulados.

Es por esto por lo que, resulta indispensable hacer la distinción entre imperativos hipotéticos y categóricos. Sobre los primeros, Kant enuncia que: “cuando son condicionados (...) sólo por referencia a un efecto deseado (...) son imperativos hipotéticos” (CRPr, A37), estos dependen de condiciones particulares, por lo que carecen de la necesidad que caracteriza la ley moral. En cambio, los imperativos categóricos tienen la capacidad de determinar la voluntad sin referencia a fines empíricos y por ello son leyes prácticas, porque según Kant: “las leyes prácticas

se refieren sólo a la voluntad, sin considerar lo que se consigue mediante su causalidad" (CRPr, A38). En particular, considera que solo en estos últimos se puede expresar la forma de la ley moral.

Lo dicho hasta aquí supone descartar el imperativo hipotético y profundizar en el categórico. En cuanto al imperativo categórico se expresa de tres maneras, esto no implica una diferencia en contenido, sino en el modo de representar una ley moral. En primer lugar, (FMC, A52) "obra según aquella máxima por la cual puedas querer al mismo tiempo que se convierta en una ley universal". Con esta primera formulación, Kant expresa el criterio formal del imperativo e introduce el concepto de máxima como principio subjetivo de la acción, el cual a su vez está supeditado a una universalización. En concreto, el sujeto debe examinar la regla que guía su conducta para pensarla sin contradicción en su universalización. Acorde con esto, no depende del contenido empírico de la acción, sino de la forma racional de la máxima y su garantía de validez universal y necesaria.

Ahora se hace más que pertinente hacer la comparativa de la autonomía frente a la heteronomía. A este respecto, el punto definitorio radica en que el fundamento de la voluntad depende de un objeto concebido como bueno, y la ley moral se subordina a algo distinto de ella misma. Conviene subrayar que, no es solo que dependa de un objeto, sino que además toma a tal objeto como fundamento del principio práctico. De esta manera, Kant argumenta que:

> Si antes de la ley moral se admitiese algún objeto con el nombre de bueno como fundamento determinante de la voluntad y de él se derivase el principio práctico supremo, entonces éste implicaría siempre heteronomía y excluiría el principio moral (CRPr, A197)

Después de todo, heteronomía significa que la voluntad se determina por algo distinto de la ley misma: póngase por casos 1) la moral teológica la cual está esencialmente arraigada a la idea

de premios y castigos: si el ser racional se comporta bien en este mundo conforme a los planes de un Dios creador, tendrá una recompensa en el paraíso; de portarse mal, será castigado con pasar la eternidad en el infierno diseñado especialmente para su sufrimiento, 2) el hedonismo, una ética hecha con la idea de maximizar el placer. En este caso, las acciones son valoradas según el placer que causan. Entiéndase que, todas estas teorías éticas fundamentan la acción en algo empírico o condicionado, y, por tanto, Kant infiere que ninguna puede fundamentar una moral universal y necesaria, solo la autonomía es garante de tal universalidad y necesidad.

A su vez, en segundo lugar, el imperativo categórico se expresa así: (FMC, A67) "obra de tal modo que uses a la humanidad, tanto en tu persona como en la persona de cualquier otro, siempre al mismo tiempo como fin y nunca simplemente como medio". Con esto se añade que, todo ser racional en cuanto capaz de darse a sí mismo la ley, posee un valor absoluto. En esta secuencia, la humanidad no debe instrumentalizarse, no puede verse como medio para fines particulares, pues, se vulneraría su condición de fin en sí mismo. Como resultado, la validez moral exige que el sujeto debe respetar la condición racional de los demás y de sí mismo, lo que implica reconocer la dignidad del otro, independientemente de toda utilidad, interés o experiencia.

Ahora se puede sostener que, ésta segunda formulación pone de manifiesto la dignidad humana, al comprender al ser racional como fin en sí mismo. Dado que, Kant menciona "la moralidad y la humanidad, en la medida en que ésta es susceptible de aquélla, es lo único que posee dignidad" (FMC, A78). En consecuencia, la dignidad es un valor absoluto y no relativo, que excluye toda instrumentalización. Por esto también, Kant muestra que, "la autonomía es el fundamento de la dignidad de la naturaleza humana y de toda naturaleza racional" (FMC, A79), lo que permite concluir que, en la medida que la voluntad tiene capacidad de darse la ley a sí misma, se da dicho valor absoluto.

En este mismo marco, la ley moral al tiempo que determina objetivamente a la voluntad, produce el efecto subjetivo del respeto. Kant señala que: “La ley moral humilla inevitablemente a todo hombre cuando compara con esa ley la tendencia sensible de su naturaleza. Aquello cuya representación (…) despierta por sí mismo (…) respeto” (CRPr, A132). Este sentimiento proviene del reconocimiento de la ley moral por encima de las inclinaciones del mundo sensible, es decir, no proviene de experiencia o sensación alguna. A saber, esto constituye la manera en que la ley moral resulta efectiva en la subjetividad, y constituye al agente como ser racional autónomo digno.

En tercer y último lugar, el imperativo categórico se expresa así: “Obra según máximas de un miembro que legisla universalmente para un reino de los fines simplemente posible” (FMC, A84). En este punto y con las otras dos formulaciones, se implanta una dimensión sistemática de la moral, en donde un sujeto valida sus máximas no solo desde sí mismo, sino también las piensa como parte de un orden objetivo racional. Por esto, la ley moral implica: universalización, reconocimiento del otro como fin en sí mismo, y un principio que articula una comunidad de seres racionales.

Desde este punto de vista, la noción del reino de los fines proporciona una mayor claridad de la autonomía. Debido a que el individuo se da a sí mismo la ley y además de someterse a esta, se reconoce dentro de una comunidad de seres racionales. En esta lógica kantiana, las máximas concuerdan con un sistema de leyes que los seres racionales podrían validar como propias. Por consiguiente, la autonomía es más que una autolegislación aislada, existe una pertenencia a un orden racional, donde cada sujeto que lo conforma es autor y destinatario de la ley moral.

Finalmente, la autonomía como principio de la moralidad sustenta los postulados de la razón práctica: libertad, inmortalidad del alma y Dios. Kant es puntual al enunciar que estos no son conocimientos teóricos; son “presuposiciones emitidas desde un punto de vista necesariamente

práctico" (CRPr, A238), y a pesar de no expandir el conocimiento especulativo, derivan de la ley moral, en cuanto ésta determina la voluntad.

Pese a no poseer un valor cognoscitivo teórico, los postulados proporcionan realidad objetiva a las ideas de la razón en el uso práctico. Lo anterior implica que, sin importar que ellos mismos no representen objetos de conocimiento, permiten pensar coherentemente las condiciones bajo las cuales puede realizarse efectivamente la ley moral. En efecto, ellos son las condiciones necesarias del bien supremo, donde la ley moral establece la relación entre virtud y felicidad de manera no empírica. Como resultado, la razón práctica admite los postulados como exigencias del obrar moral; son necesarios en la medida en que la realización plena del bien supremo excede las condiciones del mundo sensible.

Se puede condensar lo dicho hasta aquí en que, el imperativo categórico esclarece la forma en que la ley moral determina la voluntad de manera universal y necesaria, desde un criterio de validez no fundamentado en lo empírico. A partir de ello, se mostró que solo la autonomía puede fundamentar la moralidad al alejarse de toda heteronomía, y asegurar la universalidad de la ley práctica. Igualmente, sus tres expresiones se articulan con la dignidad y el respeto, al reconocer a todo agente racional nunca como medio y siempre como un fin en sí mismo. Por último, se señaló la manera en que los postulados son exigencias derivadas de todo este sistema moral. De este modo, la autonomía no solo fundamenta la ley moral, sino que determina el marco completo en el que se piensa y realiza.

1.5 Conclusión

Por lo presentado a lo largo de este primer capítulo, se defiende la tesis de que: la autonomía constituye el concepto central de la moral kantiana, y que esto se expresa en la voluntad racional

determinada por sí misma conforme a la ley moral. En este sentido, analizar el *faktum* de la razón mostró cómo la ley moral se impone en la conciencia práctica, y cómo con esto se hace pensable la libertad. En este orden de ideas, resultó imprescindible indicar la relación entre *ratio cognoscendi* y *ratio essendi* para resaltar que la moralidad solo es posible si se presupone que la voluntad no está sometida a la causalidad natural, y, por el contrario, posee capacidad de autolegislación.

Lo anterior fue la base para pensar que la autonomía se constituye en la capacidad de la razón de darse a sí misma una ley universal, y que toda forma de heteronomía se sale de estos parámetros. Simultáneamente, se esclarecieron las tres expresiones del imperativo categórico y se señaló que éste es la forma de la ley moral, el que permite reconocer a todo ser racional como fin en sí mismo; lo que fundamenta la dignidad y el respeto. Así mismo, se mencionó a los postulados de la razón práctica como la revelación de unas exigencias necesarias derivadas de estas estructuras morales, porque hacen pensable que el bien supremo puede realizarse en el mundo inteligible.

Razones por las cuales, se concluye que: la autonomía es la autolegislación de la voluntad racional conforme a la forma universal de la ley moral, en la cual se articula la libertad y la obligación como expresiones de la racionalidad objetiva práctica. Por consiguiente, en la autonomía se fundamenta la moralidad kantiana, y por ello, es el principio desde el cual se articula toda su moral deontológica, todo el conjunto de su sistema práctico. Con esta definición de autonomía, se tiene la posibilidad de interrogar el funcionamiento de los algoritmos *deep learning* para determinar si pueden ser comprendidas como autónomas en sentido kantiano; al establecer sus alcances, límites y normativa.

2. *Deep learning* y autonomía funcional

2.1 Introducción

Los avances en inteligencia artificial han dado como resultado a máquinas capaces de realizar tareas que tradicionalmente se pensaban propias de la cognición humana: reconocimiento de patrones, lenguaje y la interacción en entornos variables. En particular, el *deep learning* ha tomado gran parte del protagonismo, lo que ha generado inquietudes filosóficas sobre la naturaleza de sus procesos respecto a su capacidad aparente de tomar decisiones, a partir de lo cual se le ha atribuido una posible forma de autonomía.

No obstante, se hace pertinente examinar con rigor técnico el funcionamiento del modelo *deep learning* antes de abordar cuestiones sobre el concepto de su autonomía y capacidad de decidir desde un terreno filosófico. En este sentido, este capítulo reconstruye los fundamentos de esta subárea de la inteligencia artificial hasta proporcionar un análisis de sus redes neuronales artificiales, mecanismos de entrenamiento y procesos de optimización. Con todo esto, se busca esclarecer en qué consiste la toma de decisiones de estos algoritmos, entendida como procesos que pueden interpretarse como mecanismos de predicción condicionada a partir de ajustes estadísticos, y se abre la pregunta sobre si es legítimo hablar de decisión.

2.2 Definición de inteligencia artificial, *machine learning* y *deep learning*

La inteligencia artificial (IA) es un área de conocimiento orientada al diseño de sistemas que realizan tareas: tales como reconocimiento de patrones, resolución de problemas o procesamiento de lenguaje. Actualmente, la IA abarca diversidad de subcampos con enfoque y metodologías variadas, esto la caracteriza como un campo heterogéneo en constante desarrollo. Stuart (2021) sostiene que:

> Abarca en la actualidad una gran variedad de subcampos (…) como el aprendizaje y la percepción, (…) La IA sintetiza y automatiza tareas intelectuales y es, por lo tanto, potencialmente relevante para cualquier ámbito de la actividad intelectual humana. (p. 1)

Dentro de los enfoques más relevantes se encuentra el *machine learning*, en donde no se programa explícitamente cada instrucción, son modelos que "aprenden" a partir de datos de los cuales identifican patrones y ajustan parámetros internos para mejorar el desempeño en tareas puntuales. En este sentido, se configura como un enfoque donde se privilegia la optimización estadística, en lugar de establecer directamente toda la codificación de sus reglas, lo que permite aplicaciones en contextos complejos y variables. Murphy (2012) lo expresa de manera más formal, como el proceso de: "learn a mapping from inputs $\boldsymbol{x}$ to outputs $\boldsymbol{y}$, given a labeled set of input-output pairs $\boldsymbol{D = \{(x_i, y_i)\}_{i=1}^N}$" (p.2)

A continuación, el algoritmo que interesa a esta investigación es el *deep learning* porque representa una subárea del *machine learning* caracterizada por el uso de múltiples capas de procesamiento que simulan redes neuronales humanas de manera artificial como indican Goodfellow et al. (2016): "they are engineered systems inspired by the biological brain" (p.13). En tal sentido, son modelos que mediante la aplicación de funciones matemáticas sucesivas transforman datos de entrada en salidas, en un proceso en el cual interviene una multiplicidad de representaciones.

2.3 Redes neuronales artificiales: estructura y funcionamiento

En el acápite anterior se partió desde la IA, se pasó al ML hasta llegar al DL. Para efectos de esta tesis, conviene profundizar en este último. Por esto, en este acápite se profundizará en cómo se construyen sus redes neuronales artificiales y cómo se articula todo para funcionar. De este

modo, se puede vislumbrar lo que hay de fondo detrás de sus procesos catalogados como toma de decisiones.

Entonces, todo el núcleo fundamental del *deep learning* está constituido por las redes neuronales artificiales. Así las cosas, cada una de estas neuronas recibe valores de entrada, a los que se les asigna unos coeficientes llamados pesos que los van a multiplicar. Goodfellow et al. (2016) sostienen que: "we can think of *w* as a set of weights that determine how each feature affects the prediction" (p. 107). Por tanto, el resultado de una predicción depende enteramente del valor asignado a cada peso en el proceso de entrenamiento.

Seguidamente, una vez dados los valores y los pesos, se implementa una función de activación que introduce no linealidad en el sistema. De esta manera, Goodfellow et al. (2016) señalan: "we must use a nonlinear function to describe the features" (p.172). Debido a esta operación, se superan las transformaciones estrictamente lineales y el sistema puede modelar relaciones complejas de datos.

Sobre este sentido se construyen las neuronas artificiales que se organizan en capas: una capa de entrada, una o varias capas ocultas que gestionan el procesamiento y una capa de salida que arroja el resultado. Este procesamiento descrito se articula entre transformaciones lineales y funciones no lineales, tales como señalan Goodfellow et al. (2016): "affine transformation controlled by learned parameters, followed by a fixed, nonlinear function called an activation function" (p. 172). De ahí que, la característica fundamental del *deep learning* es la multiplicidad de capas que posee, donde cada una transforma la información recibida de manera secuencial y progresiva hasta proporcionar representaciones más abstractas de los datos.

Pues bien, resulta relevante discernir concisamente sobre el funcionamiento de la red: este consiste en propagar toda la información desde la entrada hasta la salida, a tal proceso se le llama *forward propagation*. Aquí, la salida que se ha generado se tiene que comparar con un resultado esperado que estará mediado por una función de pérdida, la cual será capaz de medir el margen de error del modelo. Goodfellow et al. (2016) explican: "The function we want to minimize or maximize is called the objective function or criterion. When we are minimizing it, we may also call it the cost function, loss function, or error function" (p. 82).

A partir de este error, los pesos de la red se ajustarán por medio de un proceso que se conoce con el nombre de retropropagación o *back propagation*, el cual establecerá la modificación de diversos parámetros a través de técnicas proporcionadas por las aplicaciones del cálculo diferencial capaces de minimizar la función de pérdida. Para que el modelo de *deep learning* quede consolidado, todo este proceso descrito se repetirá iterativamente durante el entrenamiento de este.

En síntesis, se evidencia que el funcionamiento de las redes neuronales artificiales consiste en un proceso donde se transforman los datos mediante operaciones matemáticas que articulan la propagación de la información, miden el error y ajustan los parámetros. Este proceso tiene lugar en una fase de entrenamiento donde los parámetros se ajustan iterativamente. Por esto, por más que el algoritmo genere respuestas válidas o resultados correctos no parece existir una comprensión ni deliberación con respecto de los datos, sino un proceso de optimización progresiva de la función de pérdida en función de los datos de entrada, la arquitectura del sistema y el proceso de cálculo.

2.4 La aparente autonomía en la toma de decisiones algorítmicas y principales limitaciones del *deep learning*

Por lo visto hasta aquí, es coherente pasar del problema técnico al conceptual, para posteriormente brindar una aclaración filosófica más precisa sobre la aparente decisión en estos sistemas. En este orden de ideas, conviene determinar en qué medida el término "decidir" está presente en procesos que operan únicamente mediante transformaciones matemáticas. En contextos cotidianos se ha sembrado la sospecha sobre la competencia que posee un sistema para elegir qué contenido mostrar, qué diagnóstico sugerir, y en general, qué acciones ejecutar en situaciones altamente variables. Así, esta atribución da lugar al examen filosófico riguroso para esclarecer qué se entiende por "decidir" con respecto a los seres humanos y a estos procesos estrictamente matemáticos.

Así las cosas, dentro de una perspectiva amplia de la IA y en específico del DL, la decisión se da dentro de un marco que involucra múltiples alternativas posibles que se condicionan por un estado interno del sistema y una entrada definida. Poole y Mackworth (2017) añaden que:

> Inside the black box, an agent has some internal belief state that can encode beliefs about its environment, what it has learned, what it is trying to do, and what it intends to do. An agent updates this internal state based on stimuli. It uses the belief state and stimuli to decide on its actions (p.12)

De esta manera, la selección de estos sistemas se establece a partir de estados internos que se actualizan en función de estímulos, lo que da lugar a acciones condicionadas dentro del sistema. De ahí que, con toda esa capacidad funcional se pudiese creer que eso equivaldría a una suficiencia para decidir. No obstante, no se identifican elementos propios al proceso deliberativo humano, no

se evidencia una reflexión de por medio, o una evaluación que atienda a razones y consideración de principios.

Con respecto a los límites del *deep learning* es crucial destacarlos. Con esto se quiere decir que, pese a los deslumbrantes avances en materia tecnológica de las capacidades de la disciplina de esta subárea de la inteligencia artificial, se presentan límites concernientes al análisis riguroso que tienen que ver con: limitación de datos que le son proporcionados, muchos funcionan como una especie de caja negra y sus procesos son sintácticos más que semánticos**.**

En primer lugar, estos desarrollos algorítmicos son dependientes de los datos con que se les ha entrenado, y no solo eso, sino que también la misma estructura del modelo introduce limitaciones en términos de error y generalización, incluso cuando no se le brinden datos incompletos o viciados. Hastie (2009) en un intento de conocer el error de prueba de un modelo comenta:

> As the model becomes more and more complex, it uses the training data more and is able to adapt to more complicated underlying structures. Hence there is a decrease in bias but an increase in variance. (p.221)

En este caso, se evidencia una problemática de tratamiento sesgado porque si los datos lo están, o incluso si falta información, el modelo simplemente actuará bajo información incompleta que amplifican y reproducen. Después, ineludiblemente se cae en un planteamiento sobre las consecuencias éticas que tales datos incompletos o estructuras que introducen sesgos pueden provocar, esencialmente en terrenos sensibles como aquellos enfocados en la justicia o medicina.

En segundo lugar, gran cantidad de estos modelos funcionan, por hacer un símil; como una "caja negra" donde sus procesos carecen de una interpretación de fácil acceso. Pasquale (2015) lo

describe así: “the term “black box” (…) can mean a system whose workings are mysterious; we can observe its inputs and outputs, but we cannot tell how one becomes the other” (p.3). Conviene subrayar que, aquí se dificulta la comprensión de cómo se toman finalmente las decisiones, ante lo cual los procesos inferenciales no se pueden rastrear con precisión y facilidad, lo que puede viciar procesos donde se requiere de transparencia y delegar responsabilidad.

En tercer, la falta de generalización de sentido fuerte en estos sistemas. Por su parte, Christopher Bishop (2006) señala que: “the performance on the training set is not a good indicator of predictive performance on unseen data due to the problem of over-fitting” (p. 32). En este sentido, a pesar de que pueden identificar patrones simples y complejos, en multitud de casos el desempeño del aprendizaje supervisado y no supervisado puede verse afectado sustancialmente por falta de situaciones no contempladas en los datos de entrenamiento o estructura del modelo.

Según esta caracterización se puede efectuar una distinción entre dos niveles que permiten el análisis: la descripción funcional de estos procesos automáticos que gestan resultados útiles, y el otro asunto filosófico, en que se cuestiona tal funcionalidad de manera tal que se profundiza en la naturaleza de dichos procesos desde la criticidad. Por lo cual, desde la perspectiva funcional se hará legítimo comunicar lo relacionado a las decisiones algorítmicas desde su tecnicidad, mientras que, desde el ámbito filosófico se torna más exigente la descripción lingüística de la palabra “decisión” por toda la problemática terminológica que ello implica.

2.5 Conclusión

Durante este capítulo se reconstruyó el concepto de *deep learning* a partir de sus bases en el área de la inteligencia artificial hasta el análisis de sus redes neuronales artificiales, mecanismos de entrenamiento y optimización. Este itinerario evidenció que sus procesos de funcionamiento

tienen lugar en transformaciones matemáticas orientadas a la minimización de errores para ajustar parámetros iterativamente, todo esto mediado por una dependencia de datos y estructuras previamente definidas.

Por consiguiente, la toma de decisiones en estos sistemas algorítmicos puede entenderse como una serie de predicciones condicionadas, que no involucran necesariamente una comprensión, deliberación o autonomía. Con esto, se abre camino al análisis filosófico para examinar si es legítimo atribuir autonomía y decisión en sentido riguroso a estos sistemas.

3. Comparación entre autonomía moral kantiana vs autonomía funcional algorítmica

3.1 Introducción

En este capítulo se establece la comparación entre la autonomía kantiana y la toma de decisiones de los algoritmos *deep learning*. Para ello, se recapitula lo definido anteriormente para realizar la comparación sistemática entre conceptos kantianos y el funcionamiento técnico de los algoritmos tratados en cuestión. El objetivo es analizar si es legítimo atribuir al DL características propias de la racionalidad práctica kantiana. En este sentido, se divide el capítulo en cuatro acápites que permiten: establecer las diferencias entre decidir y *outputs*, confrontar la autonomía respecto del entrenamiento, problematizar causalidad y libertad para definir en dónde se sitúan estos sistemas, y verificar si se presenta normatividad y dignidad.

3.2 La decisión es algo más que un simple outputs

El output del DL no cumple las condiciones kantianas de decisión. En primer lugar, para Kant la decisión implica la determinación de la voluntad conforme a principios universales. En segundo, en DL aquello parecido a una decisión es un ajuste de pesos y relacionamiento de datos por medio de funciones matemáticas. En este sentido, en los procesos decisionales de estos

algoritmos no existe voluntad ni normatividad, por lo que no puede hablarse de decisión en sentido kantiano.

Es importante entender que seleccionar no es igual que decidir. Si bien los procesos DL tienen alta capacidad de correlacionar datos y optimizar resultados; de ello no se sigue que tengan un juicio, en la medida en que este implica la subordinación de lo que es particular a principios. Además, estos procesos de las máquinas a partir de su correlación pueden optimizar la selección de resultados a un nivel que en entornos variables pueda ejecutar la mejor acción, pero esto no constituye deliberación, dado que no media legislación racional.

En consecuencia, se puede afirmar que ellos operan a nivel sintáctico sin comprensión del contenido, por lo cual no hay decisión. John Searle (1980) refiriéndose a estos procesos computacionales menciona “has a syntax but no semantics” (P. 423), y si una máquina no comprende lo que produce o actúa por un significado, se sitúa fuera del ámbito de la racionalidad práctica kantiana. Es así como queda demostrado que, la toma de decisiones en los algoritmos *deep learning* no es más que un proceso sintáctico carente de condiciones kantianas para la decisión.

3.3 Autonomía vs entrenamiento

El *deep learning* es heterónomo. Por una parte, depende de los datos de entrenamiento, a partir de los cuales ajusta sus parámetros y predicciones. Por otra, depende de su arquitectura en la medida en que sin esta no hay relacionamiento de datos, es a través de ella que se define la forma en que estos se van a relacionar, todo se da en un orden lógico sintáctico, donde se fija la estructura formal de sus operaciones. Por consiguiente, no se presenta el principio moral de autonomía que para Kant es “determinar por sí misma la voluntad, independientemente de todo elemento

empírico" (CRPr, A72). En ninguna instancia el algoritmo puede superponerse a sus estructuras o datos proporcionados, lo que evidencia su dependencia de tales condiciones empíricas.

En este orden, la función de pérdida no es equivalente a la ley moral. En efecto, la *loss function* tiene validez instrumental, depende del problema que se le presente, de los datos y de si puede optimizarse eficientemente. Además, esto último no siempre se logra, como Goodfellow et al (2016) mencionan "Sometimes, the loss function (...) is not one that can be optimized efficiently." (p. 276), lo que refuerza su carácter instrumental y dependiente de condiciones externas. Por el contrario, en la ley moral de Kant "las condiciones sensibles no pueden prevalecer, sino que es totalmente independiente de éstas" (A, 53), aquí además hay una validez *a priori*, universal e independiente de si es fácil de aplicar. Con todo, se denota que no es igual ley moral a función de pérdida, debido a que ésta última se fundamenta en criterios externos.

En este marco, se evidencia que los procesos del *deep learning* se encuentran estructuralmente determinados por condiciones empíricas, en tanto que sus criterios para relacionar datos y optimizar están previamente fijados. Estos algoritmos no pueden determinar los principios que rigen su funcionamiento, se limitan a ejecutar un sistema de reglas dado en el que no existe autolegislación. Por estos motivos, se presenta una dependencia estructural que excluye estos mecanismos del ámbito de la voluntad en sentido kantiano, y con ello, a la autonomía.

3.4 Causalidad y libertad

En la *Crítica de la razón pura*, Kant sostiene que la experiencia de los seres racionales es estructurada por categorías *a priori* del entendimiento. Dentro de estas categorías, se encuentra inmersa la causalidad, sobre la que se pronuncia así: "todas las alteraciones suceden según la ley de la conexión de la causa y el efecto" (CrP, B233), lo que significa que todo acontecimiento

empírico está determinado por un estado precedente conforme a una regla. No obstante, es justamente la característica de la condición de los fenómenos lo que permite pensar una causalidad no determinada empíricamente, es decir, libre. Además, los eventos no se presentan de manera aislada, sino como parte de una sucesión necesaria que posibilita el conocimiento objetivo.

Bajo esta perspectiva el *deep learning* opera completamente dentro del orden causal. Es evidente que los outputs no aparecen de la nada, ni se producen de manera arbitraria, sino que se encuentran determinados por la configuración del sistema y por sus estados previos. Es así como, ellos funcionan conforme a reglas determinadas, en la medida en que dichas reglas estructuran las relaciones causales que generan cada *output*. Por tanto, su funcionamiento se sitúa dentro de una cadena de determinaciones, esto contrasta con la noción kantiana de libertad, porque no se da la ley a sí mismo sino como resultado de condiciones previas, lo que permite cuestionar si puede hablarse de toma de decisiones en sentido literal kantiano.

Ahora bien, este carácter causal no impide que surjan fenómenos como la alucinación. Al respecto, Goodfellow et al (2016) han sostenido que: “they are generally considered to represent the model’s incorrect beliefs about the world (…) “hallucinations” or “fantasy particles.” (p. 608). Esto se entiende en cuanto que, lo modelos en su generación, según Lian Ge (2025) “decompress these representations, often introducing artifacts—hallucinations—where data is sparse or contradictory” (p.2), lo que implica que los datos no se correspondan con el fenómeno. De esta manera, no hay un rompimiento del orden causal, el resultado es producto de las limitaciones intrínsecas al sistema, lo que pone en tensión la confiabilidad en su toma de decisiones.

Sobre la base de lo anterior, al sistema no se le puede adscribir autonomía en su toma de decisiones, puesto que, incluso cuando sus *outputs* son erróneos, dependen de condiciones iniciales y reglas previamente establecidas, por lo cual no existe autodeterminación en sentido práctico

kantiano. Por estos motivos, el *deep learning* se aleja de la noción kantiana al no darse la ley a sí mismo desde la razón, y por ello, no se puede hablar de libertad al tiempo que se confirma que su funcionamiento permanece dentro de un orden estrictamente causal, sin constituir decisión en sentido kantiano de la voluntad al carecer de autodeterminación.

3.5 Normatividad y dignidad

Queda claro que el DL no puede estar obligado por principios por no estar sujeto al deber, su accionar se basa en criterios instrumentales. En este sentido, no hay una coacción racional, ni ley moral; al contrario, solo optimización de la *loss function*. En cambio, la acción moral en Kant está obligada al deber, no se sustenta en un criterio mecánico o técnico, es así como sostiene que: "deber y obligación son las únicas denominaciones que debemos dar a nuestra relación con la ley moral" (CRPr, A147). Con esto dicho, se excluye al *deep learning* de la relación con ley moral, en cuanto que, carece de deber y obligación.

En esta línea, el DL no reconoce fines en sí mismos. En la medida en que el significado no pertenece al sistema, sino que depende del programador y/o usuario, su operación se limita por estructuras formales que ejecutan sin que esto de lugar a una apropiación de significado del contenido manipulado. En los términos de Searle (1980) "that the programmer and the interpreter of the computer output use the symbols to stand for objects in the world is totally beyond the scope of the computer" (p. 423), en estas circunstancias, no puede atribuir valor ni distinguir entre medios y fines. Precisamente, sin reconocimiento de valor ni distinción entre medios y fines, no hay dignidad.

Por lo tanto, en las limitaciones estructurales del *deep learning* se radica la imposibilidad de que operen en el ámbito de la acción moral. En este sentido, estos algoritmos pertenecen al

orden estrictamente técnico por sus relaciones causales y criterios instrumentales, mientras que la acción moral en Kant pertenece a un orden práctico, que se caracteriza por normatividad, autonomía y principios universales. Por esto, queda determinado que, el *deep learning* es una naturaleza distinta de la decisión moral kantiana, donde no hay lugar para el deber, la obligación y la dignidad.

3.6 Conclusiones

A partir de todo lo dicho a lo largo de este capítulo, el *deep learning* no puede ser comprendido como un agente en sentido kantiano porque carece de autonomía, racionalidad, libertad, autonomía y ley moral. En todo momento, sus operaciones se explican desde factores técnicos: estructuras que permiten la causalidad de la que no pueden salir, optimización para fines instrumentales y datos proporcionados; con esto, no se puede atribuir decisión porque no hay una determinación racional de la voluntad conforme a principios universales. Así mismo, al carecer de autolegislación racional no hacen parte del ámbito moral y la dignidad. En definitiva, el *deep learning* está inscrito en el orden técnico sin dominio de racionalidad práctica kantiana, por lo cual no se puede equiparar sus procesos con la toma de decisiones en sentido kantiano.

4. Conclusiones

Con este trabajo queda justificado que no se puede atribuir autonomía a los procesos de toma de decisiones de los algoritmos *deep learning* desde la concepción kantiana desarrollada en la *Crítica de la razón práctica*. Así las cosas, se definió que la autonomía kantiana implica seres racionales capaces de representarse leyes y actuar conforme a ellas con independencia de condiciones empíricas, todo desde una normatividad y principios universales; condiciones que no cumplen los algoritmos DL. Por su parte, estos sistemas operan desde estructuras previamente

fijadas por el programador, dependencia de datos, optimización con base en el ajuste de pesos y funciones matemáticas; lo que los sitúa en un orden causal e instrumental.

Los aportes del desarrollo de este trabajo acorde con la justificación presentada son: se clarificó el significado de la palabra autonomía en sentido filosófico kantiano para evitar su uso ambiguo en el ámbito tecnológico. Además, se estableció que los algoritmos *deep learning* no poseen agencia racional en su toma de decisiones y no son responsables éticamente, para evitar situaciones en donde se desdibuje los criterios de responsabilidad. Con esto, se trajo la filosofía al corazón de los avances tecnológicos para aportar rigor filosófico conceptual.

Referencias Bibliográficas

Bishop, C. M. (2006). *Pattern recognition and machine learning*. Springer.

Bostrom, N. (2003). *Ethical Issues in Advanced Artificial Intelligence*. En *Cognitive, emotive and ethical aspects of decision making in humans and in artificial intelligence* (pp. 12–17). International Institute of Advanced Studies in Systems Research and Cybernetics

Centeno Franco, A. (2019). *Deep learning* (Trabajo de fin de grado, Doble Grado en Matemáticas y Estadística, Universidad de Sevilla, Sevilla, España). Universidad de Sevilla.

Dreyfus, H. L. (1972). *What Computers Can't Do: A Critique of Artificial Reason*. Harper & Row.

Floridi, L. (2013). The Ethics of Information. Oxford University Press.

Ge, L. (2025). *Spectral imaginings and sympoietic creativity: AI hallucinations and the ethics of posthuman creativity*. *Big Data & Society*, *12*(4).

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: Data mining, inference, and prediction* (2nd ed.). Springer.

Kant, I. (2007). *Crítica de la razón pura* (M. Caimi, Trad.). Fondo de Cultura Económica.

Kant, I. (2011). *Crítica de la razón práctica*. Biblioteca Immanuel Kant: México.

Kant, I. (2012). *Fundamentación para una metafísica de las costumbres*. Madrid: Alianza editorial.

Metzinger, T. (2021). *Artificial Suffering: An Argument for a Global Moratorium on Synthetic Phenomenology*. Journal of Artificial Intelligence and Consciousness, *8*(1), 1–24.

Murphy, K. P. (2012). *Machine learning: A probabilistic perspective*. MIT Press.

Searle, J. (1980). *Minds, brains, and programs. Behavioral and Brain Sciences*, 3(3), 417–424.

Pasquale, F. (2015). *The black box society: The secret algorithms that control money and information*. Harvard University Press.

Poole, D. L., & Mackworth, A. K. (2017). *Artificial intelligence: Foundations of computational agents* (2nd ed.). Cambridge University Press.

Russell, S. J., & Norvig, P. (2021). *Artificial intelligence: A modern approach* (4th ed.). Pearson.