



Los fundamentos epistemológicos de la ciencia de datos: un análisis crítico

Jules Desai¹, David Watson², Vincent Wang³, Mariarosaria Taddeo^{4,5}, Luciano Floridi^{4,6*}

¹ Facultad de Filosofía, Universidad de Oxford, Radcliffe Humanities, Woodstock Road, Oxford OX2 6GG, Reino Unido

² Department of Statistical Science, University College London, 1-19 Torrington Place, Londres WC1E 7HB, Reino Unido

³ Department of Computer Science, University of Oxford, Parks Road OX1 3QD, UK⁴

Oxford Internet Institute, University of Oxford, 1 St Giles', Oxford, OX1 3JS, UK⁵

The Alan Turing Institute, British Library, 96 Euston Rd, London NW1 2DB, UK

⁶ Departamento de Estudios Jurídicos, Universidad de Bolonia, Via Zamboni, 27, 40126 Bolonia, IT

*Correo electrónico del autor de la correspondencia: luciano.floridi@oii.ox.ac.uk

Resumen

La moderna abundancia y prominencia de los datos ha llevado al desarrollo de la "ciencia de datos" como nuevo campo de investigación, junto con un conjunto de reflexiones epistemológicas sobre sus fundamentos, métodos y consecuencias. Este artículo ofrece un análisis sistemático y una revisión crítica de los principales problemas y debates abiertos en la epistemología de la ciencia de datos. Proponemos una división de la epistemología de la ciencia de datos en los cinco ámbitos siguientes: (i) la constitución de la ciencia de datos; (ii) el tipo de investigación que identifica; (iii) los tipos de conocimiento que genera la ciencia de datos; (iv) la naturaleza y el significado epistemológico de los problemas de "caja negra"; y (v) la relación entre la ciencia de datos y la filosofía de la ciencia en general.

Palabras clave

Caja negra; Ciencia basada en datos; Ciencia de datos; Epistemología; Fundacionalismo; Filosofía de la ciencia.

1. Introducción

La ciencia de datos se ha convertido en un campo de investigación maduro recientemente, impulsado por la proliferación de datos e infraestructuras informáticas. Aunque muchos han escrito sobre los problemas filosóficos de la ciencia de datos, rara vez se unifican en una "epistemología de la ciencia de datos" holística (evitamos la expresión más genérica "filosofía de la ciencia de datos", de la que hablaremos más adelante). En su estado actual, esta epistemología es vibrante pero caótica. Por este motivo, en este artículo revisamos la bibliografía pertinente para ofrecer una perspectiva unificada de la disciplina y sus lagunas; evaluamos el estado del debate; ofrecemos un análisis contextual de la importancia, relevancia y valor de diversos temas; e identificamos áreas de interés filosófico descuidadas o poco exploradas. No abordamos las implicaciones de la ciencia de datos en materia de gobernanza, ética, legislación y sociedad (GELSI). Ya reciben una atención considerable, y su análisis quedaría fuera del alcance del presente trabajo, aunque, en última instancia, señalaremos conexiones obvias. Parece claro que la epistemología y la ética de la ciencia de datos (en el sentido inclusivo de GELSI indicado anteriormente) pueden necesitar encontrar un marco unificado. Aun así, este artículo no sería el contexto adecuado para intentar tal unificación. Metodológicamente, determinamos el alcance del análisis epistemológico mediante una búsqueda bibliográfica estructurada, detallada en el Apéndice. Nuestras conclusiones dividen la epistemología de la ciencia de datos en cinco áreas (véase la Figura 1), y el artículo se estructura en consecuencia. La sección 2 se centra en las explicaciones descriptivas y normativas de la composición de la ciencia de datos, es decir, en las explicaciones de lo que hacen y deberían hacer los científicos de datos. La sección 3 analiza las reflexiones sobre el tipo de investigación que es la ciencia de datos. La sección 4 aborda cuestiones sobre la naturaleza y la genealogía del conocimiento que produce la ciencia de datos. La sección 5 se centra en los llamados problemas de la "caja negra", como la interpretabilidad y la explicabilidad. La sección 6 explora la nueva frontera epistémicamente revolucionaria que plantea la ciencia de datos: el denominado paradigma "sin teoría" en la metodología científica. La sección 7 concluye brevemente nuestro análisis.

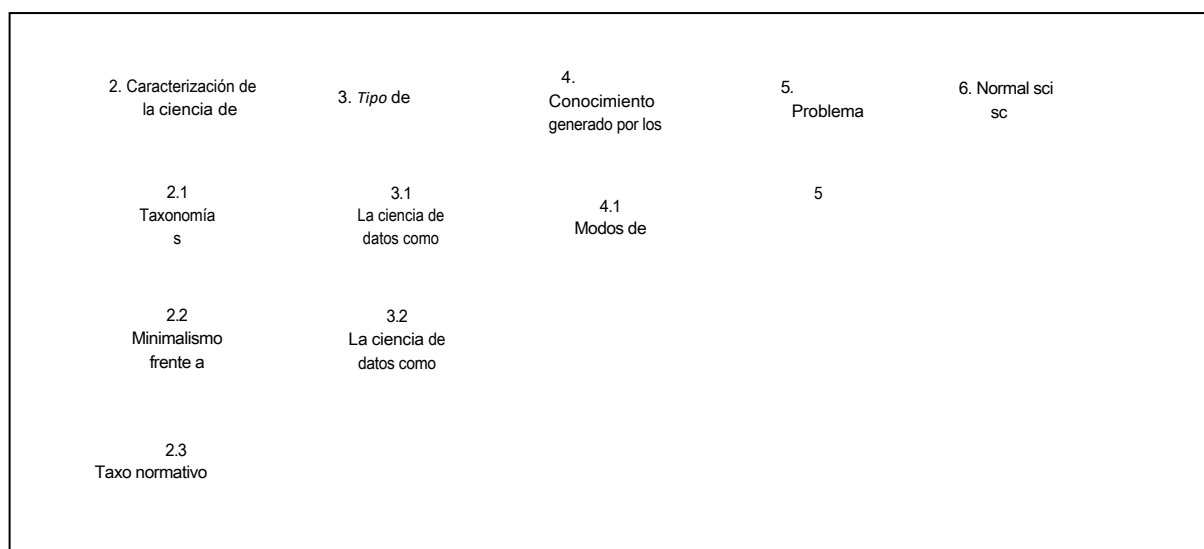


Figura 1. La epistemología de la ciencia de datos: un mapa La epistemología de la ciencia de datos: un mapa.

2. La caracterización de la ciencia de datos

Esta sección revisa las definiciones más relevantes de ciencia de datos propuestas en la literatura, abarcando alternativas descriptivas y normativas. Concluye ofreciendo una nueva propuesta que sintetiza los elementos más valiosos de cada una de ellas.

2.1 Caracterizaciones minimalistas y maximalistas

Los orígenes históricos de la ciencia de datos se remontan al trabajo de los primeros naturalistas. Sin embargo, no fue hasta el estudio sistemático de la probabilidad y la estadística, primero a través de los juegos de azar a finales del Renacimiento (Hacking, 1975) y más tarde mediante análisis sociológicos en la época de la Revolución Industrial (Gigerenzer et al., 2013), cuando empezó a surgir una forma más reconocible de esta práctica, que culminó con la aparición de la genética a finales de la Gran Bretaña victoriana (MacKenzie, 1984). Los incentivos económicos eran primordiales en todo momento, ya fuera a través de mejores estrategias de juego, tablas actuariales más precisas o mejores rendimientos agrícolas. En los albores del siglo XX, la estadística llegó a ser reconocida como una disciplina académica digna de sus propias revistas y departamentos universitarios. Los avances tecnológicos de las décadas posteriores supusieron una ruptura definitiva con la estadística clásica inferencial y basada en la teoría. Nuevos enfoques, como el bootstrapping y las simulaciones de Monte Carlo con cadenas de Markov, sustituyeron los supuestos paramétricos estrictos por la potencia bruta de cálculo. Desde este punto de vista, los algoritmos de aprendizaje automático -que detectan y explotan automáticamente patrones sutiles en grandes conjuntos de datos- no son más que el siguiente paso lógico en una progresión secular hacia formas cada vez más automatizadas de razonamiento empírico.

La cuestión de cuándo precisamente estas primeras incursiones en los modos cuantitativos de análisis cristalizaron en lo que ahora llamamos "ciencia de datos" presupone que la disciplina tiene algún carácter esencial aún no especificado. Aunque nos mostramos escépticos ante cualquier supuesta "solución" al llamado "problema" de la demarcación -en este ámbito, como en la ciencia en general-, observamos dos grandes tendencias en la bibliografía sobre este tema, que denominaremos "minimalistas" y "maximalistas" (más información al respecto más adelante). Los minimalistas buscan condiciones necesarias, lo menos restrictivas posible, pero que sigan creando un espacio único para la ciencia de datos. Los maximalistas buscan condiciones suficientes con ontologías detalladas y taxonomías metodológicas. Los enfoques minimalistas caracterizan los primeros debates sobre la naturaleza de la ciencia de datos. Los análisis contemporáneos tienden a adoptar enfoques maximalistas, identificando en la ciencia de datos un medio para desarrollar un conocimiento causal directamente conectado con el objeto de análisis.

Las concepciones minimalistas no comprometen a la ciencia de datos con ningún método o materia(s), y no hacen ninguna afirmación específica sobre qué tipo de disciplina es la ciencia de datos. Se centran únicamente en los aspectos pedagógicos y su dependencia de la información y los datos. Chambers (1993) y Carmichael y Marron (2018) proporcionan dos ejemplos de relatos minimalistas. Chambers (1993)

presenta una visión "más estadística" de la ciencia de datos, caracterizada como "todo lo *a partir de los datos*" (Chambers, 1993, p. 182, cursiva en el original). Del mismo modo, Carmichael y Marron (2018, p. 117) afirman que la ciencia de datos es "el negocio de aprender de los datos" y que un científico alguien que "utiliza los datos para resolver problemas".

Los enfoques maximalistas son más precisos. Breiman (2001) caracteriza la ciencia de

datos por el tipo de conocimiento que genera. Los estadísticos (entre los que se incluyen los científicos de datos) pueden estar interesados en hacer predicciones correlativas a partir de los datos y extraer información sobre cualquier mecanismo causal natural subyacente asociado. Los conocimientos correlativos/predictivos y causales son

se distinguen y se valoran implícitamente. Se supone que el conocimiento causal se corresponde directamente con los mecanismos "subyacentes" y "naturales" del mundo real, mientras que el conocimiento correlativo/predictivo sólo puede corresponderse con la realidad de forma oblicua, a través de la asociación con la causalidad. Hay tres ejemplos importantes en la literatura que ayudan a aclarar este punto.

Consideremos en primer lugar el relato maximalista de Mallows (2006, p. 322). Se refiere a un fin práctico. Como escribe: "La estadística se ocupa de la relación de los datos cuantitativos con un problema del mundo real, a menudo en presencia de variabilidad e incertidumbre. Intenta hacer preciso y explícito lo que los datos tienen que decir sobre el problema de interés". Mallows hace hincapié en la primacía de la resolución de problemas en lugar de la pedagogía general, un punto que también subrayan Blei y Smyth (2017) a continuación. Un aspecto singular del relato de Mallows es la mención explícita de la variabilidad y la incertidumbre, a las que los métodos científico-estadísticos de datos deben hacer frente. Esto constituye un compromiso implícito con una separación entre el ruidoso mundo real y las construcciones idealizadas familiares a las ciencias naturales y sociales. Además, la estadística se caracteriza como método epistémico fundamental por derecho propio, *como puente entre ambos mundos*.

El segundo ejemplo lo ofrece Donoho (2017, p. 746), que también apoya un enfoque maximalista. Este relato de la ciencia de datos tiene una dimensión sociológica, haciendo referencia al "código de conducta profesional" de la Data Science Association: "'Científico de datos' significa un profesional que utiliza métodos científicos para *liberar* y *crear* significado a partir de datos *brutos* [cursiva nuestra]". Esto subraya una estrecha conexión entre el análisis de datos y la investigación científica, no sólo en la metodología, sino también en sus supuestos y objetivos fundamentales. El uso de "liberar" supone que los datos proceden de procesos susceptibles de estudio y comprensión sistemáticos. Sin embargo, el término "crear" implica una desviación del quehacer científico clásico, sugiriendo la permisibilidad de superponer ontologías artificiales a los datos como medio para alcanzar cualquier fin. De ahí que la ciencia de datos se perfile como, al menos, una cuasi ciencia, si no una ciencia propiamente dicha (más información al respecto en la Sección 3). Cabe señalar que el concepto de datos "brutos" es problemático porque los datos nunca están totalmente desprovistos de interpretación. Como Donoho escribe desde la era de los grandes datos, su suposición de que los "datos brutos" son una base adecuada a partir de la cual destilar y crear significado puede ser consecuencia de la actitud contemporánea de que los datos pueden, son y serán registrados con suficiente profundidad, amplitud y calidad para cualquier dominio problemático.

Por último, Blei y Smyth (2017, p. 8691) dan una caracterización entre el minimalismo y el maximalismo: "la ciencia de datos mezcla el pensamiento estadístico y computacional... Conecta modelos estadísticos y métodos computacionales para resolver problemas específicos de la disciplina." Este punto de vista compromete a la ciencia de datos únicamente con los métodos estadísticos y computacionales, haciendo hincapié en una prioridad más práctica que pedagógica. Sin embargo, esta caracterización no especifica la información -en sentido amplio- como objeto de interés de la ciencia de datos, ni marca disciplinas específicas como padres o pacientes de la ciencia de datos.

2.2 Taxonomías descriptivas

Algunos autores han intentado caracterizar la ciencia de datos proporcionando taxonomías descriptivas y procedimentales de la disciplina. Tres relatos descriptivos, escritos en distintos momentos a lo largo de las últimas seis décadas, ofrecen una perspectiva diacrónica.

Hasta donde sabemos, Tukey (1962) dio la primera taxonomía descriptiva del "análisis

de datos" centrándose en: "procedimientos para analizar datos y técnicas para interpretar los resultados de tales procedimientos; formas de planificar la recogida de datos para que su análisis sea más fácil, preciso o

más precisos; toda la maquinaria y los resultados de la estadística (matemática) que se aplican al analizar los datos" (Tukey, 1962, p. 2). Tukey pretendía ofrecer una descripción transparente de lo que ocurre realmente en el análisis de datos. Como veremos en la sección 3, la opinión ortodoxa en la época en que Tukey escribió este artículo era que el análisis de datos era estadística aplicada y, por tanto, principalmente matemática. Al describir su naturaleza de forma clara y precisa, Tukey transgredió el *statu quo* al separar el concepto de análisis de datos de la estadística aplicada y convertirlo en un campo propio.

Algunos años después de Tukey, Wu (1997) presentó una triple taxonomía descriptiva centrada en la recogida de datos (diseño experimental, encuestas por muestreo); modelización y análisis de datos; y comprensión/solución de problemas y toma de decisiones. Al igual que la de Tukey, esta descripción formaba parte de un proyecto más amplio para orientar la estadística matemática en una dirección científica. El propio Wu apostó por renombrar la "estadística" como "ciencia de los datos" o "ciencia estadística", y destacamos la inclusión del "diseño experimental", manifiestamente científico.

Más recientemente, Donoho (2017) ha proporcionado una extensa taxonomía que cita el programa "Data Science Initiative" de la Universidad de Michigan: "[La Ciencia de Datos] implica la recopilación, gestión, procesamiento, análisis, visualización e interpretación de grandes cantidades de datos heterogéneos asociados con una diversa gama de aplicaciones científicas, traslacionales e interdisciplinarias" (Donoho, 2017, p. 745). Una breve comparación con los relatos de Tukey y Wu pone de relieve la maduración y el crecimiento de la ciencia de datos: los conductos procedimentales han coevolucionado con etapas intermedias entre las entradas y los productos.

No se puede culpar a los relatos anteriores por no haber dado en el blanco, ya que quizá no previeron las crecientes exigencias y posibilidades de la era digital. Sin embargo, todavía podemos identificar un equilibrio entre la especificidad de las restricciones y la contemporaneidad en los relatos descriptivistas de la ciencia de datos, que ha evolucionado junto con la computación. Cualquier relato que aisle la ciencia de datos de su contexto computacional corre el riesgo de quedarse obsoleto, pero cualquier relato que no lo haga debe lidiar con una esfera digital masivamente enmarañada y en evolución, con todos sus mecanismos y requisitos concomitantes. Por lo tanto, en adelante, distinguiremos la ciencia de datos de unos y ceros de la estadística de lápiz y papel por su naturaleza digital y computacionalmente intensiva.

2.3 Taxonomías normativas

Otros pensaban que la concepción del *statu quo* de la ciencia de datos de su época era inadecuada para satisfacer las demandas que la proliferación de datos planteaba a la sociedad. Esto les llevó a desarrollar relatos revisionistas, a menudo proponiendo taxonomías normativas de la ciencia de datos. A continuación se presentan los que pueden considerarse los cuatro enfoques revisionistas más influyentes hasta la fecha: Chambers (1993), Breiman (2001), Cleveland (2001) y Donoho (2017).

Chambers (1993) señala que, en el momento de su análisis, existía una tendencia en la estadística académica hacia lo que él denomina "estadística menor" -estadística matemática filtrada a través de revistas, libros de texto y conferencias- en lugar de dedicarse a las aplicaciones de los datos al mundo real. En este contexto, presenta la siguiente taxonomía tripartita (Chambers 1993, p. 182) de la composición de su "estadística mayor", refiriéndose al concepto antes mencionado como "todo lo relacionado con el aprendizaje a partir de los datos":

1. Preparación de datos (planificación, recogida, organización y validación);

2. Análisis de datos (mediante modelos u otros resúmenes);
3. Presentar datos (por escrito, gráficamente o de otra forma).

La taxonomía de Chambers distingue los procesos y productos de la ciencia de datos de la toma de decisiones y los resultados que se derivan de esos productos. La promoción de la preparación de datos en pie de igualdad con el análisis y la presentación es notablemente clarificante. Los subprocesos de planificación, recopilación, organización y validación anticipan, respectivamente, la obtención, el volumen, la diversidad y la calidad de los datos que requiere la ciencia de datos práctica, por oposición a las preocupaciones abstractas de la "estadística menor". Junto con las descripciones del análisis y la presentación de los datos, se revela una concepción de las limitaciones humanas a la hora de enfrentarse a los datos, y de la ciencia de datos como *e/* esfuerzo epistémico por superar esas limitaciones.

Breiman (2001) se hace eco de la necesidad de que la estadística se acerque al mundo real. Al igual que algunas de las afirmaciones maximalistas de la ciencia de datos analizadas en la sección 2.1, su relato también hace hincapié en que el análisis de datos colabora con disciplinas específicas y, por tanto, actúa sobre ellas, proporcionándoles herramientas analíticas. Para entender la concepción normativa radical de Breiman de la ciencia de datos como desinteresada de la verdad, en favor del conocimiento práctico, es necesario dar un breve rodeo histórico y sociológico. En homenaje a C. P. Snow, Breiman señala que la preferencia por la verdad o la acción caracteriza a dos "culturas" opuestas en estadística: el bando predictivo, que según sus estimaciones en el momento de escribir este artículo en 2001 sólo contaba con un ~2% de los estadísticos académicos y analistas de datos, y el bando de la inferencia, que contenía al resto. Los del primer bando están interesados principalmente en generar etiquetas precisas sobre datos no vistos. Los del segundo se centran en revelar mecanismos y estimar parámetros. En la sección 4 volveremos sobre esta distinción. El revisionismo de Breiman se pone de manifiesto cuando afirma que el énfasis en la inferencia sobre la predicción ha llevado a una obsesión distractora con "teoría irrelevante" y a la extracción de "conclusiones cuestionables", impidiendo así a los estadísticos "trabajar en una amplia gama de interesantes problemas actuales". Hoy en día, los tamaños relativos de las dos culturas están casi invertidos (por ejemplo, Anderson, 2008). La visión de Breiman de una ciencia de datos sin teorías marca una desviación significativa del proyecto epistemológico clásico de "entender el entendimiento".

Cleveland (2001, pp. 22-23) consideraba que los programas de enseñanza de su época eran deficientes y producían profesionales de los datos no preparados para las exigencias de una sociedad cada vez más rica en datos. En este contexto sociológico, propone la siguiente taxonomía:

1. Investigaciones multidisciplinarias (análisis de datos en el contexto de distintas áreas disciplinares)
2. Modelos y métodos para datos (modelos estadísticos, métodos de construcción de modelos, métodos de estimación, etc.)
3. Computación con datos (hardware, software, algoritmos)
4. Pedagogía (planificación curricular, formación escolar/universitaria/empresarial)
5. Evaluación de herramientas (análisis descriptivo y revisor de las herramientas y sus métodos de desarrollo).
6. Teoría (problemas fundamentales y teóricos de la ciencia de datos)

La taxonomía de Cleveland propone una concepción de la ciencia de datos como disciplina científica de pleno derecho, fundamentalmente computacional. Cuatro puntos son especialmente relevantes en este caso. En primer lugar, la inclusión de la "Computación" junto a los "Modelos y métodos para datos" marca la ciencia de datos como fundamentalmente

digital, separándola de la estadística en general. A diferencia de la taxonomía de Chambers, ahora se reconoce explícitamente que los ordenadores son el vehículo que hace posible la ciencia de datos. En segundo lugar, bajo el epígrafe "Pedagogía", se reconoce la necesidad de preservar y

propagar la ciencia de datos como campo académico y comercial. En tercer lugar, la novedosa inclusión de "Evaluación de herramientas" y "Teoría", ausentes en los relatos anteriores, señala una concepción de la ciencia de datos como autorreflexiva y progresista. En cuarto lugar, el hecho de que "Investigaciones multidisciplinares" se sitúe al mismo nivel que los otros cinco taxones indica una relativa despriorización de la aplicación. Se trata de un cambio significativo con respecto a los relatos anteriores, como el de Breiman, que tratan la ciencia de datos como un mero medio para alcanzar un fin.

Más recientemente, Donoho (2017, p. 755) ha revisado exhaustivamente la taxonomía para satisfacer las necesidades actuales. Emulando la terminología de Chambers, la "ciencia de datos mayor" se contrapone a algunas de las taxonomías descriptivas descritas en la sección 2.1, que él denomina "ciencia de datos menor". La ciencia de datos mayor consiste en:

1. Recogida, preparación y exploración de datos
2. Representación y transformación de datos
3. Computación con datos
4. Modelización de datos
5. Visualización y presentación de datos
6. Ciencia sobre ciencia de datos

A diferencia de la taxonomía de Cleveland, la de Donoho se centra únicamente en la ciencia de datos *como* campo y medio de investigación. Dos aspectos de esta taxonomía son epistemológicamente interesantes. En primer lugar, al igual que en Cleveland, la presencia repetida de la sexta categoría metacientífica: la ciencia de datos debe reflexionar y hacer ciencia sobre sí misma para mejorar y desarrollarse. En segundo lugar, la descripción de Donoho es completa desde el punto de vista procedimental, ya que comienza con la exploración y recopilación de datos y recorre todos los pasos analíticos, desde los orígenes hasta los productos finales. Este ambicioso alcance contribuye a la fuerza normativa de la propuesta de Donoho.

Hemos llegado al final de la Sección 2. A la luz de estas consideraciones previas, parece razonable proponer la siguiente definición de ciencia de datos:

La ciencia de los datos es el estudio de los sistemas de información (naturales o artificiales), mediante el razonamiento probabilístico (por ejemplo, inferencia y predicción) implementado con herramientas computacionales (por ejemplo, bases de datos y algoritmos).

Esta definición es lo suficientemente amplia como para abarcar todos los casos de aprendizaje automático -supervisado, no supervisado y de refuerzo-, así como procedimientos más genéricos que suelen incluirse en el ámbito de la estadística, como los gráficos de dispersión para inspeccionar tendencias y el bootstrapping para cuantificar la incertidumbre. Puede excluir o no algunos casos extremos, dependiendo de la interpretación que se haga de los términos que lo componen. Por ejemplo, abarca los sistemas deterministas si se considera que son un subconjunto de los sistemas probabilistas. Abarca los modelos de regresión calculados a mano si se considera que la cognición humana es un tipo de cálculo. Sin embargo, se trata de zonas grises, aunque la primera pueda ser un caso obvio de informática y la segunda un caso obvio de estadística. La ciencia de los datos se extiende por ambas disciplinas, haciendo hincapié en diferentes aspectos.

3. Tipo de consulta

Los críticos pueden alegar que la ciencia de datos no es una disciplina académica, sino un conjunto de herramientas agrupadas mediante funciones pragmáticas. La cuestión es si los datos son el "tipo correcto de cosa" para erigirse en objeto de una disciplina. Si "datos" es un concepto demasiado insustancial o los métodos de la ciencia de datos son demasiado heterogéneos, cualquier intento de forjar una ciencia de datos unificada parece condenado al fracaso.¹ Sin embargo, existe una demanda creciente de ciencia de datos, no sólo en el mundo empresarial, sino también en el académico, como demuestra la proliferación de cursos y programas universitarios, puestos de enseñanza especializados, conferencias dedicadas, revistas y puestos en la industria. Por lo tanto, en aras de la argumentación, supongamos que la ciencia de datos está, como mínimo, en camino de convertirse en una disciplina arraigada y madura. La siguiente pregunta es de qué tipo. La bibliografía ofrece tres respuestas principales: una especie de estadística académica, en la que la estadística es una parte formal y teórica de las matemáticas aplicadas; la estadística, pero convenientemente ampliada para sacarla de las matemáticas aplicadas y convertirla en una protociencia; y una ciencia propiamente dicha. A continuación examinaremos en detalle cada una de estas alternativas.

3.1 Ciencia de datos como estadística

Las dos primeras respuestas consideran que la ciencia de datos es una forma de estadística. Por ejemplo, Donoho (2017, p. 746) ofrece una amplia recopilación de artículos, charlas y blogs cuyos autores sostienen que la ciencia de datos es simplemente estadística con otro nombre. Esta postura también se especifica en función de si se considera que la estadística forma parte de las matemáticas aplicadas o está separada de ellas. En el primer caso, Wu (1997) cita una definición de diccionario de la estadística: "las matemáticas de la recogida, organización e interpretación de datos numéricos". Esta visión estrecha del análisis de datos no tiene muchos defensores contemporáneos. La mayor parte de la bibliografía actual acepta que el análisis de datos forma parte de una estadística ampliada, que ya no se considera estrictamente matemática formal (véase la estadística ampliada de Chambers), o bien concede al análisis de datos el estatus de campo independiente, externo pero relacionado con la estadística, que se considera una parte estrecha de la matemática formal aplicada. Breiman (2001) y Mallows (2006) adoptan esta última postura, reclamando la ampliación de la estadística para que incluya elementos científicos y se relacione con disciplinas del mundo real. Esto no implica que la estadística sea en sí misma una ciencia completa. El análisis de datos, desde este punto de vista, sigue siendo estadística, aunque empieza a trascender la inferencia y las prácticas matemáticas deductivas estrictamente formales.

3.2 La ciencia de datos como ciencia

Otros autores sitúan el análisis de datos como una disciplina científica. Carmichael y Marron (2018, p. 120) afirman que una ciencia de datos manifiestamente científica es una "reacción a la estrecha comprensión de la *estadística menor*" [cursiva nuestra]. Hay dos estrategias principales para apoyar la afirmación de que el análisis de datos es una ciencia.

La primera consiste en formular criterios de demarcación para lo que ya llamamos ciencia (cf. Popper, 1959), y luego demostrar que la ciencia de datos los satisface. Tukey (1962, p.5) hizo este intento, estableciendo tres criterios paradigmáticos de demarcación para la ciencia: "contenido intelectual", "organización en una forma comprensible" y "confianza en la prueba de la experiencia como norma última de validez". Al cotejar su ciencia de datos contemporánea con estos criterios,

¹ Se trata de una cuestión abierta en la filosofía de la información, que no abordaremos aquí, ya que es lo suficientemente profunda como para merecer una investigación propia. No obstante, señalaremos que en Floridi (2010) se puede encontrar un intento sostenido de análisis de los conceptos de datos e información.

Tukey llegó a la conclusión de que todo lo que hace que otras disciplinas sean científicas también se aplica a la ciencia de datos. (Donoho 2017) se centra en una característica científica paradigmática de una materia: la formulación de preguntas empíricamente responsables que se resuelven mediante técnicas científicamente rigurosas. Dado que hay espacio conceptual para un campo de esta naturaleza que opera con datos e información, concluye que hay espacio para una próxima ciencia genuina del análisis de datos.

Sin embargo, esta primera estrategia choca con la heterogeneidad de la ciencia. El debate sobre la demarcación perdió fuerza tras la decisiva crítica de Laudan (1983) al falsacionismo de Popper. En vista de ello, la segunda estrategia alternativa consiste en demostrar similitudes relevantes entre la ciencia de datos y las ciencias paradigmáticas, y que estas similitudes justifican una ampliación del concepto general. Por ejemplo, Wu (1997) cita una serie de similitudes importantes entre su taxonomía descriptiva de la estadística y las ciencias paradigmáticas. Estas similitudes incluyen: el enfoque "empírico-físico" de la estadística, en el que utilizamos la inducción para inferir conocimiento a partir de observaciones y la deducción para inferir implicaciones de las teorías; la primacía del diseño experimental y la recogida de datos; y el uso del razonamiento bayesiano para evaluar modelos y pruebas. Sin embargo, hay aspectos notables en los que la ciencia de datos diverge de las ciencias paradigmáticas. Entre estas diferencias se encuentran el tipo de conocimiento que genera (véase la sección 4), los modos de inferencia lógica que utiliza (véase la sección 4) y el estatus que otorga a las hipótesis (véase la sección 6).

Otra diferencia podría ser que la ciencia de datos se sitúa de algún modo junto a las ciencias normales, proporcionándoles las herramientas y los recursos necesarios para realizar descubrimientos más profundos y específicos de cada disciplina. Si estas diferencias se consideran suficientemente significativas, cabe pensar que la ciencia de datos podría no ser una ciencia en absoluto, o tal vez podría ser una ciencia trascendental. Este es el tema de la siguiente sección.

3.3 La ciencia trascendental

El debate anterior pasa por alto la posibilidad de que la ciencia de datos no sea ni estadística aplicada ni ciencia, sino algo totalmente distinto. Wiggins ha expresado este pensamiento en una comunicación privada con Donoho, afirmando que "la ciencia de datos no es una ciencia... Es una forma de ingeniería, y los que la hacen en este campo la definirán, no los aspirantes a científicos" (Donoho 2017, p. 764). Se podría hacer una afirmación similar sobre la informática, que tiene sus raíces en las matemáticas, pero está lo suficientemente especializada como para constituir su propio campo de investigación. Como se desprende de las taxonomías descriptivas anteriores, existen muchas similitudes relevantes entre la informática y la ciencia de datos. Una posibilidad interesante sería escindir tanto la "ciencia" de los datos como la "informática" en categorías no científicas propias.

La cuestión se agudiza si se considera la ciencia de datos como una *base* para la ciencia empírica. Al igual que Wittgenstein llegó a considerar la filosofía como un conjunto de herramientas y métodos para resolver la confusión en otras áreas como las matemáticas o la psicología, la ciencia de datos puede concebirse como una función trascendental para las ciencias, como la condición para la posibilidad de la investigación empírica como tal. No hay nada fundamentalmente diferente entre, por ejemplo, las taxonomías de Linneo y las ontologías jerárquicas familiares a los gestores de bases de datos. Los diarios de Tycho Brahe son esencialmente un conjunto de datos de alta calidad. Las leyes del movimiento de Newton son un algoritmo, obtenido a partir de datos empíricos y verificado con ellos, para predecir los valores de algunas variables físicas a partir de los valores de otras. No vamos a seguir este

enfoque en este artículo. Lo planteamos más como una sugerencia a explorar que como una tesis a defender.

4 El conocimiento generado por la ciencia de datos

Esta sección examina el conocimiento generado por la ciencia de datos. El análisis se estructura en dos partes relacionadas: el proceso, o *cómo*, (relativo a los modos de inferencia) y el producto, o *qué*, (referido a los productos epistémicos) de la ciencia de datos.

4.1 Modos de inferencia

Los distintos medios de indagación tienen afinidades diferentes con los tres modos típicos de inferencia: deducción, inducción y abducción. La epistemología de la ciencia de datos refleja en qué medida los científicos de datos utilizan estos distintos modos.

Las inferencias deductivas están presentes en la ciencia de datos a través de la gran dependencia del razonamiento matemático y lógico. La teoría de la probabilidad, el cálculo diferencial, el análisis funcional y la informática teórica son disciplinas puramente deductivas muy utilizadas para deducir las propiedades de los algoritmos y diseñar nuevos procedimientos de aprendizaje sin preocuparse demasiado por el comportamiento empírico. Por ejemplo, el algoritmo de retropropagación utilizado para optimizar parámetros en redes neuronales combina elementos de álgebra lineal y cálculo multivariable para converger, de forma demostrable, en un óptimo local de una función objetivo. No se necesitan conjuntos de datos para obtener este resultado.

Las inferencias inductivas también tienen una importancia fundamental. Los datos son una muestra finita del mundo. La ciencia de datos identifica estructuras en los datos y las destila en información que se aplica más allá de los propios datos. Esto se consigue proyectando los patrones y estructuras encontrados en los datos a nuevos contextos, yendo más allá del dominio antecedente. Esta proyección es precisamente la inferencia inductiva. Esto representa una solución derrotable al problema de la inducción de Hume, por el que las pruebas estadísticas pueden proporcionar pruebas más sólidas o más débiles a favor de hipótesis particulares (Mayo, 1996; 2018). Por esta razón, Harman y Kulkarni (2007) sostienen que la teoría del aprendizaje estadístico representa una defensa de la inducción basada en principios y sofisticada. Observaciones similares se pueden encontrar en Frické (2015), quien observa que "Los algoritmos inductivos son un pilar central de la empresa Big Data." Más recientemente, Schurz (2019) ha argumentado que los resultados formales del aprendizaje por refuerzo demuestran la optimalidad de la metainducción, resolviendo así el problema de Hume sobre bases *a priori*.

Se puede distinguir entre dos tipos canónicos de inferencia inductiva, la inducción de *objetos* y la inducción de *reglas*. El primero es la predicción informada de instancias singulares no observadas: hipótesis de la forma "la siguiente instancia observada de X será Y" basadas en datos previos de la co-instanciación de X e Y. Es lo que se conoce como inducción de *objetos*. La inducción de *reglas*, por el contrario, plantea afirmaciones universales del tipo "todos los X son Y", basadas en los mismos datos. La investigación científica de datos incluye ambas. Las instancias predictivas singulares son habituales en cualquier aplicación del aprendizaje supervisado, donde el objetivo es aprender una función de entradas a salidas. Estos son los tipos de inducciones que interesan a la "primera cultura" de la estadística de Breiman (2001). Al mismo tiempo, uno de los propósitos de la ciencia de datos es identificar estructuras y mecanismos subyacentes. El proyecto de inferencia causal, que retomamos en la sección 4.2, está dedicado a estas formas de inducción de reglas.

En cuanto a la inferencia abductiva, Alemany Oliver y Vayre (2015) han destacado la importancia del razonamiento abductivo en los métodos de la ciencia de datos, en particular en la forma en que la ciencia de datos se integra en la práctica científica más amplia (véase la

Sección 6 para un análisis más detallado). Sostienen que las herramientas de la ciencia de datos se interesan, en primer lugar, por la exploración de los datos para determinar su estructura interna y, en segundo lugar, por la identificación de las mejores hipótesis para explicar esta estructura. Esta inferencia de la estructura a una hipótesis explicativa es una inferencia abductiva. La idea de que

que la ciencia es esencialmente abductiva se remonta a Peirce, aunque abundan los partidarios modernos (Harman, 1965; Lipton, 1991; Niiniluoto, 2018). El estatus de la abducción en un contexto de datos intensivos se eleva aún más por la virtud teórica de la unificación explicativa (Kitcher, 1989). En la filosofía de la ciencia, una virtud común de una teoría es su poder explicativo, y algunos autores sostienen que dicho poder es motivo para elegir una de dos teorías empíricamente equivalentes (cf. la discusión de van Fraassen (1980) sobre las virtudes pragmáticas). Una dimensión del poder explicativo es el grado de diversidad y heterogeneidad de los fenómenos que una teoría puede explicar simultáneamente (cf. Kitcher, 1976). Si los métodos de la ciencia de datos permiten identificar patrones en una gama diversa y heterogénea de fenómenos, entonces tal vez desarrollemos una imagen más amplia y matizada del poder explicativo de nuestras teorías. En el caso de las teorías que pueden unificar muchos fenómenos, el razonamiento abductivo les confiere un apoyo más sólido teniendo en cuenta diversas técnicas de la ciencia de datos.

Además de ser un fin en sí misma, la reflexión epistemológica sobre los modos de influencia también arroja luz sobre las conexiones entre la ciencia de datos, las matemáticas y la ciencia. Las similitudes entre estas disciplinas -como su relevancia, poder explicativo, utilidad práctica y grado de éxito- son precisamente lo que se cuestiona cuando buscamos ampliar las categorías de forma coherente. Por ejemplo, las pruebas matemáticas se formulan deductivamente. Pero dada la importancia de las inferencias no deductivas en la ciencia de datos, hay que reconocer una importante disimilitud entre ambas y abstenerse de situar la ciencia de datos estrictamente dentro de las matemáticas aplicadas. Del mismo modo, las ciencias naturales utilizan una mezcla de deducción, inducción y abducción en su práctica diaria; las ciencias más formales recurren con más frecuencia a la deducción y las ciencias más aplicadas, a la abducción. Otras ciencias asignan distinta importancia a los diferentes modos de inferencia. Por ejemplo, la abducción es habitual en las ciencias sociales, políticas y económicas. La ciencia cognitiva es otro ejemplo que se basa en la abducción dada la frecuencia de teorías empíricamente equivalentes e infradeterminadas. Parece que la ciencia de datos, si es una ciencia, está en buena compañía.

4.2 Productos epistémicos

La tricotomía del aprendizaje automático -que abarca algoritmos de aprendizaje supervisado, no supervisado y de refuerzo- ayuda a delimitar el tipo de conocimiento generado por la ciencia de datos y sus técnicas.

Los modelos de aprendizaje supervisado predicen resultados basándose en asociaciones observadas. Automatizan el proceso de razonamiento inductivo a escalas y resoluciones que superan con creces la capacidad de los seres humanos. Sin embargo, los grandes conjuntos de datos y los potentes algoritmos no bastan para superar los retos fundamentales inherentes a este modo de inferencia. Un modelo que funciona bien en un entorno puede fracasar estrepitosamente en otro si los datos dejan de ajustarse a los patrones observados. Por ejemplo, un clasificador entrenado para distinguir vacas de camellos puede tener problemas cuando se le presenta una vaca en el desierto o un camello sobre la hierba, suponiendo que el conjunto de entrenamiento sólo contenga imágenes de ambos animales en sus hábitats naturales. Dado que el fondo era un indicador fiable del resultado en el entrenamiento, se podría perdonar al modelo por suponer que ocurriría lo mismo en el momento de la prueba. Esta falibilidad es inherente a todo razonamiento inductivo, que sin embargo nos ayuda a alcanzar muchos objetivos epistémicos importantes.

El aprendizaje no supervisado es un conjunto más heterogéneo de métodos, unidos a grandes rasgos por su tendencia a inferir estructuras sin variables de resultado

predefinidas. Algunos ejemplos son los algoritmos de agrupación, los autocodificadores y los modelos generativos. En el mejor de los casos, estas herramientas pueden arrojar

La información sobre las propiedades latentes, es decir, cómo las muestras o características reflejan algunos hechos subyacentes sobre el proceso de generación de datos. Por ejemplo, los métodos de agrupación se utilizan habitualmente en la investigación del cáncer para clasificar a los pacientes en subgrupos basados en biomarcadores. La idea es que un hecho esencial (por ejemplo, que el cáncer se manifiesta en subtipos identificables) se refleja en alguna propiedad contingente (por ejemplo, los niveles de expresión génica). El riesgo de sobreajuste -es decir, de "descubrir" alguna estructura en los datos de entrenamiento que no se generaliza a los datos de prueba- es especialmente alto en este escenario, ya que no existe una variable de resultado con la que evaluar los resultados.

En el aprendizaje por refuerzo, uno o varios agentes deben navegar por un entorno con poca orientación más allá de una señal de recompensa potencialmente intermitente. El objetivo es deducir una política que maximice las recompensas y/o minimice los costes. Un buen ejemplo es el problema del bandido de brazos múltiples. Un agente debe elegir entre un conjunto predefinido de posibles acciones -es decir, debe "tirar" de algún "brazo"- sin conocer las recompensas o penalizaciones asociadas a cada una. Por lo tanto, un agente en este escenario tiene que encontrar un difícil equilibrio entre la exploración (tirar aleatoriamente de nuevos brazos) y la explotación (tirar continuamente del brazo con la mayor recompensa hasta el momento). El aprendizaje por refuerzo ha impulsado algunos de los logros más notables de la ciencia de datos en los últimos años, como AlphaGo, un algoritmo que es actualmente el mejor jugador del mundo de Go, ajedrez y otros juegos de mesa clásicos. El producto epistémico de estos algoritmos no son ni asociaciones (como en el aprendizaje supervisado) ni estructuras (como en el aprendizaje no supervisado), sino *políticas*: métodos para tomar decisiones en condiciones de incertidumbre.

Por sí solos, estos métodos no proporcionan necesariamente conocimiento causal. Sin embargo, algunas de las investigaciones más importantes sobre IA de los últimos 20 años se han centrado en el razonamiento causal (Spirtes et al., 2000; Pearl, 2009; Imbens y Rubin, 2015; Peters et al., 2017). Dicha investigación demuestra cómo los supuestos probabilísticos pueden combinarse con datos observacionales y/o de intervención para inferir la estructura causal y los efectos del tratamiento. Sorprendentemente, esta literatura apenas está empezando a ganar tracción en la comunidad de aprendizaje automático. Trabajos recientes en el aprendizaje supervisado han demostrado cómo los principios causales pueden mejorar el rendimiento fuera de la distribución (Arjovsky et al., 2019), mientras que algoritmos complejos como las redes neuronales y los bosques potenciados por gradiente se utilizan cada vez más para inferir los efectos del tratamiento en una amplia gama de entornos (Nie y Wager, 2021). El descubrimiento causal -la tarea de aprender asociaciones causales a partir de datos observacionales- es un problema de aprendizaje no supervisado por excelencia. Ha sido un área activa de investigación desde al menos la década de 1990 y sigue siéndolo en la actualidad (véase Glymour et al. (2019) para una revisión reciente). El aprendizaje por refuerzo -quizás la más obviamente causal de las tres ramas, dada su dependencia de las intervenciones- ha sido objeto de intensa investigación en los últimos años (Bareinboim et al., 2021). Varios autores han demostrado cómo la información causal puede mejorar el rendimiento de estos algoritmos, lo que a su vez ayuda a revelar la estructura causal.

En principio, estos métodos pueden utilizarse para inferir leyes naturales. Schmidt y Lipson (2009) han propuesto lo que parecen ser las leyes de la mecánica clásica obtenidas algorítmicamente. Su método consistía en analizar los datos de movimiento de varios sistemas dinámicos mediante algoritmos que no tenían conocimientos físicos previos de mecánica. Afirman obtener el lagrangiano y el hamiltoniano de esos sistemas dinámicos, junto con varias leyes de conservación. Esto proporciona un dato atractivo para quienes albergan

esperanzas en la posibilidad del descubrimiento autónomo de leyes naturales. El papel de la correlación y la causalidad en la ciencia, así como de la ciencia autónoma y sin teorías, se analizan en la Sección 6.

5. Problemas de caja negra

Las herramientas de la ciencia de datos se han vuelto muy sofisticadas y complejas. Esto se debe en parte a que la ciencia de datos siempre ha respondido a motivaciones prácticas. Cualquier desarrollo que produzca un resultado más satisfactorio (más eficiente, preciso, desplegable, etc.) se adopta en virtud de su utilidad, a menudo sin detenerse a reflexionar sobre cómo debe integrarse en nuestros esquemas conceptuales más amplios. Esto ha llevado a cuestionar la opacidad de estas herramientas. En esta sección evaluamos varios tipos de problemas de caja negra propuestos en la bibliografía.

En primer lugar, puede ser útil hacer una aclaración. Burrell (2016) ha propuesto que hay tres formas en que los algoritmos de la ciencia de datos se vuelven opacos. La primera es su ocultación intencionada para beneficio comercial o personal. La segunda es la opacidad que surge de la alfabetización y competencia tecnológicas como condición necesaria para comprender algoritmos sofisticados. Y el tercero es la complejidad inherente derivada de los procedimientos de optimización algorítmica que superan la capacidad de cognición humana. Los dos primeros son problemas pragmáticos que surgen cuando la ciencia de datos se integra en la sociedad en general (véase Tsamados et. al. (2021) para un trabajo reciente sobre estas cuestiones). No son el tipo de problemas epistemológicos que nos ocupan aquí. Por lo tanto, nos centraremos únicamente en el último problema. Además, ha habido muchas soluciones técnicas o proto-soluciones a varios casos de problemas de caja negra. La naturaleza de estas soluciones no nos concierne porque nos centramos en el nivel filosófico de abstracción por encima de tales investigaciones técnicas. En esta sección, sólo ofrecemos un breve resumen comparativo para ilustrar las (des)similitudes o los casos en los que problemas supuestamente diferentes pueden colapsar en uno solo.

Para empezar, los problemas de las cajas negras son de dos tipos: *conceptuales* y *no conceptuales*. Los problemas conceptuales son los que se refieren a los límites de los conceptos que se emplean al hablar de cajas negras. Por ejemplo, si es posible proyectar conceptos como "explicabilidad" en un contexto de aprendizaje automático de forma coherente y no ambigua. Los problemas no conceptuales, por el contrario, no se refieren a la naturaleza, los límites y las coherencias de los conceptos empleados en sí, sino a los problemas más amplios que se derivan del uso de estos conceptos, como en epistemología. Aquí nos limitaremos únicamente a los problemas no conceptuales en el ámbito de la epistemología. No obstante, conviene tener presente que surgen otros problemas no conceptuales en otros ámbitos, por ejemplo, en la ética o la política.

5.1. Problemas conceptuales

Algunos problemas de caja negra surgen de que nuestros conceptos ordinarios son de alguna manera inadecuados o poco claros cuando se proyectan en contextos de aprendizaje automático. Lipton (2018) ha reconocido esta imprecisión sobre el uso de "interpretación". Observa que "la tarea de interpretación parece subespecificada. Los artículos proporcionan motivaciones diversas y a veces no superpuestas para la interpretabilidad y ofrecen una miríada de nociones de qué atributos hacen que los modelos sean interpretables" (Lipton, 2018, p. 36). Del mismo modo, Doshi-Velez y Kim (2017) han señalado la falta de acuerdo sobre una definición de "interpretabilidad" y sobre cómo debe evaluarse. Identifican dos usos paradigmáticos de la "interpretabilidad" en la literatura: la interpretabilidad en el contexto de una aplicación y la interpretabilidad a través de un proxy cuantitativo. Se echa en falta una definición rigurosa de ambas.

Ha habido algunos intentos de responder a estos problemas conceptuales. Un primer paso

importante es construir una taxonomía clara de cómo conceptos problemáticos como la interpretabilidad son

utilizados, y cuáles son los desiderata y las metodologías de la investigación de la interpretabilidad. Este es el tipo de proyecto al que se dedica Lipton (2018). Un segundo intento es refinar estos conceptos, o al menos realizar el trabajo preliminar para facilitar este refinamiento. Doshi-Velez y Kim (2017) se dedican a este tipo de proyecto, sentando las bases para la posterior definición y evaluación rigurosas de la interpretabilidad. Los autores también han refinado los conceptos de interpretabilidad y sus afines haciendo distinciones precisas dentro de ellos, añadiendo a su estructura. Doshi-Velez y Kim (2017) distinguen entre interpretabilidad local y global para evitar confusiones. La primera se aplica a predicciones individuales y la segunda a todo el límite de decisión o superficie de regresión. Watson y Floridi (2020) hacen una distinción similar entre explicaciones locales (token) y globales (tipo), aunque en un contexto matemático más formal.

Otros trabajos sobre las representaciones desplegadas en los problemas de caja negra se refieren a la relación entre varios términos aproximadamente sinónimos: palabras como "interpretabilidad", "explicabilidad", "comprensibilidad", "opacidad", etcétera. Desde el punto de vista filosófico, es interesante saber si alguno o todos estos términos coinciden total o parcialmente. Algunos comentaristas adoptan un enfoque de grano grueso para tales cognados. Krishnan (2020), por ejemplo, los considera insignificamente diferentes, argumentando que todos estos términos se definen entre sí de una manera circular que hace poco por aclarar conceptos imprecisos. Otros adoptan un enfoque más preciso. Tsamados et al. (2021) hacen hincapié en la diferencia entre explicabilidad e interpretabilidad. La primera se aplica tanto a los expertos como a los no expertos; por ejemplo, el experto científico de datos puede tener que explicar la mecánica de un algoritmo a su cliente no experto. En cambio, la segunda se limita a los expertos (interpretabilidad como interpretabilidad-en-principio). Así pues, en su opinión, la explicabilidad presupone la interpretabilidad, pero no a la inversa.

5.2 Problemas no conceptuales

Los problemas no conceptuales y sus soluciones no abordan las deficiencias de las propias representaciones. En esta sección analizaremos cuatro problemas epistemológicos que han recibido menos atención.

Ratti y López-Rubio (2018) han argumentado que la interpretabilidad es crucial para destilar explicaciones causales a partir de las correlaciones identificadas por técnicas de ciencia de datos, como puede ser el caso en un contexto científico rico en datos. A través del paradigma de los modelos biológicos mecanicistas, observan que para que los biólogos conviertan los modelos correlativos de la ciencia de datos en modelos causales con poder explicativo, los modelos correlativos deben ser interpretables. Esto se debe a una contrapartida general: cuanto más complejo es un modelo, menos explicativo es. Dado que el poder predictivo de los modelos científico-datos está positivamente correlacionado con su complejidad, llegan a la conclusión de que existe un auténtico problema epistemológico de caja negra.

Watson y Floridi (2020) han interpretado la sobreadaptación como un tipo diferente de problema epistemológico de caja negra, como una especie de caso Gettier algorítmico. La sobreadaptación se produce cuando un modelo de aprendizaje automático realiza predicciones correctas en el corpus de entrenamiento, pero no logra predecir correctamente en los datos de prueba. Citan los resultados de Lapushkin et al. (2016), en los que las imágenes de un caballo compartían una marca de agua sutil y distintiva. El clasificador de imágenes resultante asoció fuertemente esa marca de agua con la etiqueta "caballo" y, por lo tanto, no pudo clasificar correctamente los caballos en un conjunto de prueba cuando la marca de agua estaba ausente. Watson y Floridi proponen que esto es similar a los casos de Gettier, en los que uno se

forma una creencia accidentalmente verdadera a través de mecanismos de generación de conocimiento poco fiables. Sostienen que un marco para interpretar las cajas negras reducirá el sobreajuste.

Krishnan (2020) ha señalado el punto epistemológico más amplio de que, en la medida en que los algoritmos de aprendizaje automático puedan tener una dimensión pedagógica (que podamos aprender de los errores que puedan cometer los algoritmos), deben ser interpretables o comprensibles para que podamos aprender algo en absoluto. Lipton (2018, sección 2.4) (citando a Kim et al., 2015 y Huysmans et al., 2011) hace una observación similar sobre la informatividad de los algoritmos. Así pues, una mayor transparencia algorítmica conlleva importantes beneficios epistémicos.

El debate anterior da la impresión de que estos problemas son importantes y merece la pena resolverlos. Sin embargo, no todos los comentaristas están de acuerdo. Hay dos tipos principales de objeciones. Algunos admiten que las cajas negras son opacas, pero niegan que la forma correcta de proceder sea intentar explicar o interpretar su funcionamiento interno. En su lugar, sostienen que las cajas negras deberían sustituirse por cajas no negras igualmente capaces. Otros niegan que las cajas negras sean problemáticas en absoluto. Presentaremos sucesivamente ambos tipos.

Rudin (2019) ha expresado una objeción del primer tipo. Está de acuerdo en que la falta de interpretabilidad de los algoritmos de aprendizaje automático es un problema. Sin embargo, ella toma esto no como una motivación para construir mejores métodos de interpretabilidad *post hoc*, sino como una razón para rechazar los modelos opacos por completo. Rechaza el tópico de que la precisión y la interpretabilidad están inversamente relacionadas. En su opinión, los problemas de caja negra deben disolverse (más que resolverse) mediante modelos globalmente transparentes que tengan un rendimiento comparable al de sus competidores de caja negra.

Zerelli et al. (2019) han expresado una objeción del segundo tipo, argumentando que la opacidad de las cajas negras no es un problema genuino en absoluto. Consideran que el debate sobre la explicabilidad evidencia un pernicioso doble rasero. Señalan que no exigimos explicaciones explícitas y transparentes a los jueces, médicos, generales militares o banqueros. Más bien, la justificación se encuentra simplemente en la fiabilidad pasada: precisión y éxito demostrados y sostenidos. Si imponemos las mismas normas a los algoritmos, el problema de la explicabilidad se disuelve una vez más.

En una línea similar, Krishnan (2020) ha argumentado que nuestra preocupación por la interpretabilidad y sus cognados es innecesariamente exagerada. La imprecisión inherente a estos términos les impide hacer el trabajo que se requiere de ellos: "La interpretabilidad y sus cognados son nociones poco claras... Aún no sabemos qué concepto(s) debe(n) recoger las definiciones técnicas o, de hecho, si existe algún concepto de interpretación o interpretabilidad que deba recogerse técnicamente" (Krishnan, 2020, p. 490). Pero a diferencia de Doshi-Velez y Kim, Krishnan no toma esto como una motivación para afinar tales conceptos para el progreso posterior, ya que preocuparse por ellos nos distrae de nuestras necesidades reales. Krishnan sostiene que la mayoría de las motivaciones *de facto* para tratar la interpretabilidad como un problema epistemológico en primer lugar se deben a otros fines (por ejemplo, sociales, políticos, etc.). Por ejemplo, las auditorías de sesgos algorítmicos utilizan la explicabilidad como medio para evitar consecuencias poco éticas.

Simpatizamos con el proyecto general de Krishnan. Muchos autores asumen acríticamente que los problemas de caja negra son necesariamente importantes, y las preocupaciones epistemológicas sobre conceptos como la interpretabilidad son a menudo en la práctica medios para otros fines. Sin embargo, no estamos de acuerdo en que éstos agoten la utilidad epistemológica de tales conceptos, como atestiguan los ejemplos de la sección 5.2. Podría darse el caso de que preocuparse por los problemas de caja negra sea un uso ineficaz y subóptimo del esfuerzo filosófico (especialmente en el contexto hiperpragmático en

el que se despliegan principalmente los métodos de la ciencia de datos). Sin embargo, los problemas de caja negra *como* objetos de interés epistemológico siguen siendo relevantes al menos para algunas partes de una filosofía completa de la ciencia de datos.

6. La ciencia normal en un paradigma de uso intensivo de datos

Tras haber examinado hasta ahora las cuestiones fundamentales de la filosofía de la ciencia de datos, podemos ampliar la investigación para considerar cómo la ciencia de datos podría dar forma a la ciencia y a la filosofía de la ciencia en general. Kuhn (1970) propuso que la ciencia atraviesa ciclos de normalidad, crisis y, en última instancia, revolución. En la fase normal, los científicos se dedican a resolver rompecabezas gnósticos con las herramientas del paradigma imperante. Sin embargo, recientemente se ha propuesto que la proliferación de datos ha inaugurado una nueva era de ciencia agnóstica. En ella, el conocimiento científico puede generarse, y los métodos matemáticos y científicos de datos desplegarse, sin ningún conocimiento o comprensión previos de los fenómenos o sus interrelaciones. Kitchin (2014) ha recopilado el trabajo de Gray (que se encuentra en Hey et al. (2009)) para dilucidar la naturaleza de este nuevo paradigma y situarlo en la historia de la ciencia (véase la Tabla 1). Esta sección explora el alcance y las implicaciones de la ciencia agnóstica.

Paradigm	Naturaleza	Formulario	W
a	Experimental	Empirismo; descripción del phe	
Primero	Teórica	natural Modelización y géneros	
Segund	Computacional	Simulati	
o	Exploración		

Tabla 1. Paradigmas científicos (tomado de Kitchen (2014, p. 3), recopilado de Hey et al. 2009)

6.1. Agnosticismo sobre la aplicación de las matemáticas

Una identificación del agnosticismo la proporcionan Napoletani et al. (2018), quienes observan que la aplicación de *facto* de las técnicas matemáticas en la ciencia está experimentando una transformación agnóstica. Observan que los métodos clásicos requerían tanto la comprensión previa de los fenómenos como las interconexiones entre los elementos de los conjuntos de datos. Así ocurre, por ejemplo, si se desea modelizar alguna población biológica mediante ecuaciones diferenciales. La naturaleza de los modelos que uno utiliza, qué parámetros incluir, etc., requieren que el científico tenga conocimientos previos y comprensión sobre biología de poblaciones, cálculo multivariante, etc. También necesitan que el científico conozca la estructura básica del conjunto de datos. Las cosas son muy distintas en el análisis de datos contemporáneo. Allí, el científico puede permanecer en gran medida agnóstico o desinformado sobre cualquier teoría científica subyacente y la estructura de sus datos. Con las herramientas de la ciencia de datos contemporánea, los datos brutos pueden analizarse y la estructura explotarse de forma más o menos automática.

Tras observar que ésta parece ser una dirección importante en la práctica científica, Napoletani et al. plantean la cuestión de segundo orden de por qué las matemáticas y los datos tienen una sinergia tan eficaz. Afirman que una respuesta común es apelar a una resignación de tipo wigneriano ante la "eficacia irrazonable" (Wigner, 1960). Desde este punto de vista, los macrodatos tienen una especie de omnipotencia que concede un éxito irrazonable a herramientas científicas de datos dispares y heterogéneas. Sin embargo, Napoletani et al. rechazan esta respuesta, argumentando que la cuestión puede reformularse en la pregunta más general de si el éxito de los métodos matemáticos en una ciencia normal agnóstica se debe a una similitud entre la estructura de esos métodos y la estructura de los fenómenos

en los corpus de datos. Se trata de una cuestión que merece mayor atención en el debate.

6.2. Ciencia sin teoría

Mientras Napoletani et al. observan la creciente posibilidad de emplear técnicas matemáticas de forma agnóstica, otros han entablado un debate más radical sobre si este agnosticismo anuncia el fin de la elección de teorías en la ciencia en su conjunto. Anderson (2008) sostiene que la ciencia clásica basada en teorías se está quedando obsoleta. En su opinión, la densidad y la pluralidad de las correlaciones derivadas del análisis de cantidades extraordinariamente grandes de datos serán más útiles que las generalizaciones causales de la ciencia clásica. Otros autores han hecho observaciones similares (Prensky, 2009; Steadman, 2013). Kitchin (2014) ofrece una caracterización más formal de este punto de vista, que denomina un nuevo tipo de empirismo. La mencionada reconstrucción de la mecánica clásica mediante aprendizaje automático de Schmidt y Lipson (2009) es un ejemplo provocador de ciencia sin teoría en acción. Los críticos objetan que esto es sensacionalista, demasiado optimista e inflado. Kitchin (2014) presenta un ataque cuádruple contra el análisis de Schmidt y Lipson (2009). Su primer argumento es que, por mucho que los grandes corpus de datos intenten agotar la información en un dominio completo, no dejan de estar teñidos por la tecnología utilizada en su generación y manipulación, la ontología de datos en la que existen y la posibilidad de sesgo de muestreo. De hecho, "todos los datos proporcionan visiones oligópticas del mundo" (Kitchin, 2014). En segundo lugar, siguiendo a Leonelli (2012), señala que ni siquiera la destilación agnóstica de la estructura y los patrones a partir de los datos puede producirse *in vacuo* a partir de toda teoría científica. Debido a su profunda incrustación en la sociedad, las teorías científicas y la formación siempre proporcionan el andamiaje en torno a la recopilación y el análisis de datos. En tercer lugar, en la medida en que la ciencia normal es acumulativa, sostiene que los resultados individuales de las investigaciones científico-didácticas siempre requerirán la interpretación y el encuadre por parte de científicos que a su vez estén equipados con conocimientos de teorías científicas. Y en cuarto lugar, si los datos y los resultados de su análisis se interpretan libres de cualquier teoría de fondo, corren el riesgo de resultar infructuosos. Será difícil que contribuyan a una comprensión fundamental de la naturaleza de los fenómenos, ya que "carecen de arraigo en un conocimiento... más amplio" (Kitchin, 2014). Frické (2015) presenta un punto de vista similar contra este tipo extremo de agnosticismo. Argumenta que se necesita un conocimiento teórico previo para decidir qué datos proporcionar a los algoritmos inductivos en la primer lugar. La teoría no puede eliminarse de la ciencia, ni siquiera en un paradigma basado en los datos.

Creemos que estos argumentos pueden complementarse con otras dos razones en contra del agnosticismo total. La primera se refiere a la cuestión crítica en la filosofía de la ciencia de la teorización de la observación, que sostiene que lo que uno observa está influido por sus compromisos teóricos y preteóricos. Esto es doblemente cierto en el caso de la ciencia de datos, donde las observaciones se recogen, etiquetan y procesan de acuerdo con categorías y rutinas de análisis preexistentes. En segundo lugar, es plausible que la afirmación de Anderson de que las correlaciones serán suficientes para el futuro de la ciencia sea una concepción demasiado ingenua de la empresa científica. Recuerda a la insostenible opinión de Francis Bacon de que la Naturaleza hablaría por sí sola si se la interrogara adecuadamente. En efecto, la ciencia agnóstica de datos puede generar una ciencia predictiva sin conocimiento de ninguna ley natural o mecanismo causal subyacente. Pero la predicción no es el único objetivo de la empresa científica; otro es explicar los fenómenos llegando a conocer la estructura causal subyacente del mundo, lo que ayuda a planificar e intervenir.

El agnosticismo total, por tanto, parece demasiado extremo. La tarea consiste entonces en

cómo integrar las prácticas agnósticas de la ciencia de los datos en la metodología científica. Kitchin (2014) propone una explicación más humilde de esta integración. La denomina "ciencia impulsada por los datos", y adopta la forma de un reequilibrio del

tres modos de inferencia analizados en la sección 4.1. Sostiene que la ciencia normal contemporánea tiene una dimensión experimental-deductiva en la que las hipótesis se deducen de hipótesis más fundamentales y luego se ofrecen para su confirmación o refutación mediante experimentos. En cambio, la ciencia en un paradigma basado en datos eleva el estatus de la lógica inductiva en este proceso de formación de hipótesis, con hipótesis experimentales generadas a partir de correlaciones identificadas por métodos científico-dimensionales en lugar de por deducción a partir de hipótesis parentales. Sin embargo, a diferencia del empirismo ingenuo, la ciencia basada en datos de Kitchin no implica la primacía absoluta de la inducción. Las teorías y sus deducciones desempeñan un papel esencial, por ejemplo, a la hora de encuadrar los datos, orientar los procesos científico-dimensionales que deben desplegarse, integrar los resultados en conocimientos más amplios, generar explicaciones causales, etc. Surge así la imagen de una nueva ciencia, que implica un cambio hacia una empresa más inductiva, al tiempo que mantiene muchas similitudes paradigmáticas y realistas con nuestro modelo actual de ciencia normal.

Se han hecho otras observaciones sobre la introducción de métodos científicos de datos en las ciencias sociales. Lazer et al. (2009) destacan la aparición de la "ciencia social computacional", y Miller (2010) observa la proliferación de datos en el contexto de la ciencia regional y urbana. En ambos casos, se reconoce el potencial de los datos para remodelar las prácticas sociocientíficas. Sin embargo, los autores han señalado las diferencias entre las ciencias naturales y las sociales, lo que probablemente significa que el impacto de los datos en ambas categorías será distinto.

Es probable que el futuro de la ciencia intensiva en datos siga basándose en la teoría, aunque a veces se utilicen métodos agnósticos y científicos de datos para ayudar a la generación de teorías. Desde Reichenbach (1938), en la filosofía de la ciencia se ha hecho una distinción popular entre el contexto de *descubrimiento* y el contexto de *justificación*: de dónde viene una teoría es irrelevante para saber si la teoría es sólida. En consecuencia, se ha convertido en ortodoxo considerar las teorías científicas sólo por su propio contenido, independientemente de sus orígenes. Clásicamente, la genealogía de nuestro conocimiento científico nunca ha tenido relevancia epistémica.

Esta distinción puede ponerse en tela de juicio con la posibilidad de una ciencia agnóstica en un paradigma intensivo en datos. Por ahora, la genealogía de ese conocimiento agnóstico que se genera de forma autónoma a partir de los datos es importante: su posición epistémica depende de las herramientas y los algoritmos de la ciencia de datos que lo generaron y de la calidad de los datos antecedentes. Así, la fiabilidad de las inferencias automatizadas depende de la calidad de los datos subyacentes y del algoritmo o algoritmos utilizados para extraer información de ellos. Este tipo de cuestiones sobre la genealogía de las teorías son quizás demasiado a menudo ignoradas por las modernas filosofías de la ciencia que informan a los paradigmas "gnósticos". Una filosofía de la ciencia en un paradigma de uso intensivo de datos puede verse obligada a abordarlas más directamente.

7. Conclusión

En este artículo ofrecemos un análisis sistemático e integrado del panorama actual de la epistemología de la ciencia de datos. Nos hemos centrado en su evaluación crítica y en la identificación y caracterización de algunas de sus lagunas acuciantes o evidentes en las que reside su interés filosófico. Hemos estructurado esta reconstrucción en cinco áreas: relatos descriptivos y normativos de la composición de la ciencia de datos; reflexiones sobre el tipo de indagación que es la ciencia de datos; naturaleza y genealogía del conocimiento que produce la ciencia de datos; problemas de la "caja negra"; y naturaleza y posición de una nueva frontera

dentro de la filosofía de la ciencia que plantea la ciencia de datos. Cada una de estas áreas alberga una serie de cuestiones importantes y debates activos, y cada una de ellas interactúa con las demás. La imagen resultante es una epistemología rica, interconectada y floreciente, que

sin duda seguirá ampliándose a medida que se produzcan avances tanto filosóficos como tecnológicos, y posiblemente influya en otros puntos de vista interconectados sobre la naturaleza de la ciencia y sus fundamentos.

Apéndice: Detalles de la búsqueda bibliográfica

La búsqueda bibliográfica para el presente trabajo se llevó a cabo como se describe en la siguiente Tabla 2:

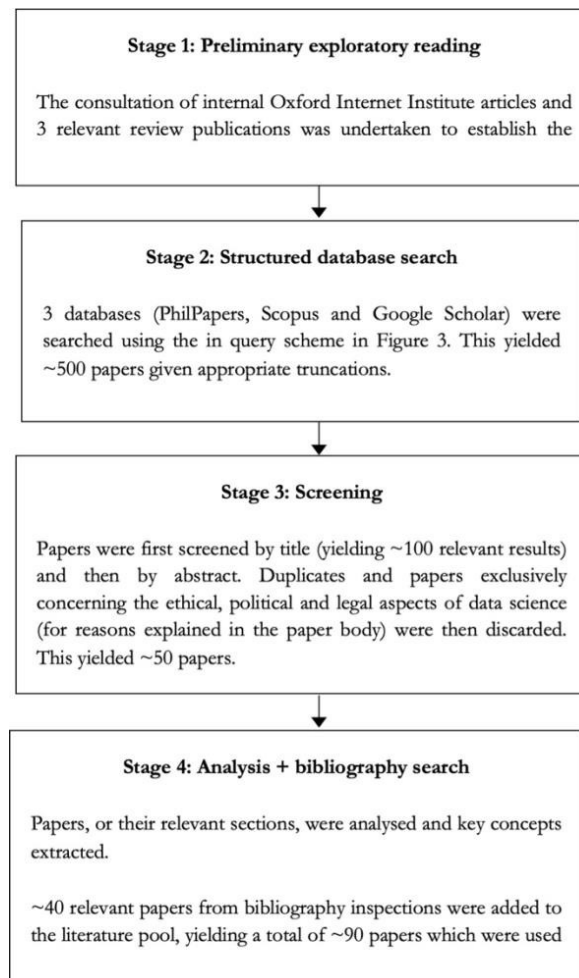


Tabla 2. Esquema de la búsqueda bibliográfica

La bibliografía se limitó exclusivamente a artículos escritos en inglés. El impacto de esta elección en el análisis es probablemente mínimo, ya que no esperamos una variación geográfica, cultural o lingüística considerable en las cuestiones fundacionales de la ciencia de datos, dada la naturaleza global, altamente interdisciplinar y contemporánea de la disciplina. Esto es, tal vez, a diferencia de otras cuestiones filosóficas (por ejemplo, las éticas), que pueden ser más sensibles a tal variación en la genealogía o circunstancia.

Figura 3. Tabla de consultas de búsqueda

Bibliografía

- Alemaný Oliver, M. y Vayre, J.-S. (2015) 'Big data y el futuro de la producción de conocimiento en la investigación de marketing: Ethics, digital traces, and abductive reasoning', *Journal of Marketing Analytics*, 3(1), pp. 5-13. doi: [10.1057/jma.2015.1](https://doi.org/10.1057/jma.2015.1).
- Anderson, C. (2008) 'El fin de la teoría: The Data Deluge Makes the Scientific Method Obsolete', *Wired*. Disponible en: <https://www.wired.com/2008/06/pb-theory/> (fecha de consulta: 14 de diciembre de 2020).
- Arjovsky, M., Bottou, L., Gulrajani, I., & Lopez-Pad, D. (2019) Minimización invariante del riesgo. *arXiv preprint*, 1907.02893.
- Baker, M. (2016) "1.500 científicos levantan la tapa sobre la reproducibilidad", *Nature News*, 533(7604), p. 452. doi: [10.1038/533452a](https://doi.org/10.1038/533452a).
- Bareinboim, E. y Pearl, J. (2016) "Causal inference and the data-fusion problem", *Proceedings of the National Academy of Sciences of the United States of America*, 113(27), pp. 7345-7352. doi: [10.1073/pnas.1510507113](https://doi.org/10.1073/pnas.1510507113).
- Bareinboim, E., Lee, S., & Zhang, J. (2021) An introduction to causal reinforcement learning. Laboratorio CausalAI de Columbia, Informe técnico (R-65).
- Blei, D. M. y Smyth, P. (2017) "Science and data science", *Proceedings of the National Academy of Sciences*, 114(33), pp. 8689-8692. doi: [10.1073/pnas.1702076114](https://doi.org/10.1073/pnas.1702076114).
- Breiman, L. (2001) 'Statistical Modeling: The Two Cultures (with comments and a rejoinder by the author)', *Statistical Science*, 16(3), pp. 199-231. doi: [10.1214/ss/1009213726](https://doi.org/10.1214/ss/1009213726).
- Burrell, J. (2016) 'Cómo "piensa" la máquina: Understanding opacity in machine learning algorithms:', *Big Data & Society*. doi: [10.1177/2053951715622512](https://doi.org/10.1177/2053951715622512).

- Canali, S. (2016) 'Big Data, epistemología y causalidad: Knowledge in and knowledge out in EXPOsOMICS', *Big Data and Society*, 3(2).
- Carabantes, M. (2020) 'Inteligencia artificial de caja negra: An Epistemological and Critical Analysis', *AI and Society*, 35(2), pp. 309-317. doi: [10.1007/s00146-019-00888-w](https://doi.org/10.1007/s00146-019-00888-w).
- Carmichael, I. y Marron, J. S. (2018) 'Data Science vs. Statistics: Two Cultures?', *Japanese Journal of Statistics and Data Science*, 1(1), pp. 117-138. doi: [10.1007/s42081-018-0009-3](https://doi.org/10.1007/s42081-018-0009-3).
- Carroll, J. W. (2020) 'Laws of Nature', en Zalta, E. N. (ed.) *The Stanford Encyclopedia of Philosophy*. Invierno de 2020. Laboratorio de Investigación Metafísica, Universidad de Stanford. Disponible en: <https://plato.stanford.edu/archives/win2020/entries/laws-of-nature/> (consultado: 18 diciembre de 2020).
- Chambers, J. M. (1993) 'Greater or lesser statistics: a choice for future research', *Statistics and Computing*, 3(4), pp. 182-184. doi: [10.1007/BF00141776](https://doi.org/10.1007/BF00141776).
- Cleveland, W. S. (2001) 'Data Science: An Action Plan for Expanding the Technical Areas of the Field of Statistics', *International Statistical Review / Revue Internationale de Statistique*, 69(1), pp. 21-26. doi: [10.2307/1403527](https://doi.org/10.2307/1403527).
- Crawford, K. (2014) 'Critizando el Big Data: Politics, Ethics, Epistemology | Special Section Introduction', p. 10.
- Donoho, D. (2017) '50 Years of Data Science'. doi: [10.1080/10618600.2017.1384734](https://doi.org/10.1080/10618600.2017.1384734).
- Doshi-Velez, F. y Kim, B. (2017) 'Towards A Rigorous Science of Interpretable Machine Learning', *arXiv:1702.08608 [cs, stat]*. Disponible en: <http://arxiv.org/abs/1702.08608> (fecha de consulta: 7 de diciembre de 2020).
- Elragal, A. y Klischewski, R. (2017) "¿Predicción basada en la teoría o en el proceso? Epistemological challenges of big data analytics", *Journal of Big Data*, 4(1), p. 19. doi: [10.1186/s40537-017-0079-2](https://doi.org/10.1186/s40537-017-0079-2).
- Floridi, L. (2010) *Información: A Very Short Introduction*. Oxford University Press.
- Floridi, L. (2012) 'Big Data and Their Epistemological Challenge', *Philosophy & Technology*, 25(4), pp. 435-437. doi: [10.1007/s13347-012-0093-4](https://doi.org/10.1007/s13347-012-0093-4).
- van Fraassen, B. C. (1980) *La imagen científica*. Oxford University Press.
- van Fraassen, B. C. (1989) *Laws and Symmetry, Leyes y simetría*. Oxford University Press. Disponible en: <https://oxford.universitypressscholarship.com/view/10.1093/0198248601.001.0001/acprof-9780198248606> (fecha de consulta: 14 de diciembre de 2020).
- Frické, M. (2015) 'Big data and its epistemology', *Journal of the Association for Information Science and Technology*, 66(4), pp. 651-661. doi: [10.1002/asi.23212](https://doi.org/10.1002/asi.23212).

- Gigerenzer, G., Swijtink, Z., Porter, T., Daston, L., Beatty, J. & Krüger, L. (2013) *El imperio del azar: Cómo la probabilidad cambió la ciencia y la vida cotidiana*. Nueva York: Cambridge University Press.
- Glymour, C., Zhang, K., & Spirtes, P. (2019) Revisión de los métodos de descubrimiento causal basados en modelos gráficos. *Frontiers in Genetics*, 10, 524.
- Hacking, I. (1975) *El surgimiento de la probabilidad: Un estudio filosófico de las primeras ideas sobre probabilidad, inducción e inferencia estadística*. New York: Cambridge University Press.
- Hey, T., Tansley, S. y Tolle, K. (2009) 'The Fourth Paradigm: Data-Intensive Scientific Discovery', p. 287.
- Harman, G. (1965) La inferencia a la mejor explicación. *Philosophical Review*, 74(1), 88-95. Harman, G. & Kulkarni, S. (2007) *Razonamiento fiable: Induction and statistical learning theory*. Cambridge, MA: The MIT Press.
- Hillar, C. y Sommer, F. (2012) Comentario al artículo 'Distilling free-form natural laws from experimental data'.
- Hooker, G. y Hooker, C. (2017) "El aprendizaje automático y el futuro del realismo", *arXiv:1704.04688 [cs, stat]*. Disponible en: <http://arxiv.org/abs/1704.04688> (fecha de consulta: 24 de septiembre de 2020).
- Imbens, G.W. y Rubin, D.B. (2015) *Causal inference for statistics, social, and biomedical sciences: An introduction*. New York, NY, US: Cambridge University Press.
doi:10.1017/CBO9781139025751.
- Kelling, S. et al. (2009) 'Data-intensive Science: A New Paradigm for Biodiversity Studies', *BioScience*, 59(7), pp. 613-620. doi: 10.1525/bio.2009.59.7.12
- Kerridge, I., Mason, P. y Lipworth, W. (2017) "Ethics and Epistemology of Big Data", *Journal of Bioethical Inquiry*, 14(4), pp. 485-488.
- Kim, B. et al. (sin fecha) 'iBCM: Interactive Bayesian Case Model Empowering Humans via Intuitive Interaction', p. 12.
- Kitchin, R. (2014) "Big Data, new epistemologies and paradigm shifts", *Big Data & Society*, 1(1), p. 2053951714528481. doi: 10.1177/2053951714528481.
- Kitcher, P. (1976) "Explanation, Conjunction, and Unification", *The Journal of Philosophy*, 73(8), pp. 207-212. doi:10.2307/2025559
- Kitcher, P. (1989) La unificación explicativa y la estructura causal del mundo. En P. Kitcher & W. Salmon (eds.), *Scientific Explanation*, pp. 410-505. Minneapolis: University of Minnesota Press.

- Krishnan, M. (2020) "Against Interpretability: a Critical Examination of the Interpretability Problem in Machine Learning", *Philosophy & Technology*, 33(3), pp. 487-502. doi: [10.1007/s13347-019-00372-9](https://doi.org/10.1007/s13347-019-00372-9).
- Kuhn, T. S. (1970) *La estructura de las revoluciones científicas*. 2nd Edición. Chicago: University of Chicago Press.
- Laudan, L. (1983) La desaparición del problema de la demarcación. En R.S. Cohen & L. Laudan (Eds.), *Physics, philosophy and psychoanalysis: Ensayos en honor de Adolf Grünbaum* (pp. 111-127). Dordrecht: Springer.
- Lapuschkin, S. et al. (2016) 'Analyzing Classifiers: Fisher Vectors and Deep Neural Networks', en. *Actas de la Conferencia IEEE sobre Visión por Ordenador y Reconocimiento de Patrones*, pp. 2912-2920. Disponible en: https://www.cv-foundation.org/openaccess/content_cvpr_2016/html/Bach_Analyzing_Classifiers_Fisher_CVPR_2016_paper.html (fecha de consulta: 17 de diciembre de 2020).
- Lazer, D. et al. (2014) "La parábola de Google Flu: Traps in Big Data Analysis", *Science*, 343(6176), pp. 1203-1205. doi: [10.1126/science.1248506](https://doi.org/10.1126/science.1248506).
- Leonelli, S. (2014) "¿Qué diferencia hace la cantidad? On the epistemology of Big Data in biology", *Big Data & Society*, 1(1), p. 2053951714534395. doi: [10.1177/2053951714534395](https://doi.org/10.1177/2053951714534395).
- Lipton, P. (1991). *Inference to the best explanation*. London: Routledge.
- Lipton, Z. C. (2018) 'The mythos of model interpretability', *Communications of the ACM*, 61(10), pp. 36-43. doi: [10.1145/3233231](https://doi.org/10.1145/3233231).
- Lowrie, I. (2017) "Racionalidad algorítmica: Epistemology and efficiency in the data sciences", *Big Data & Society*, 4(1), p. 2053951717700925. doi: [10.1177/2053951717700925](https://doi.org/10.1177/2053951717700925).
- MacKenzie, D. (1984) *Statistics in Britain, 1865-1930: La construcción social del conocimiento científico*. Edimburgo: Edinburgh University Press.
- Mallows, C. (2006) 'Tukey's Paper After 40 Years', *Technometrics*, 48, pp. 319-325. doi: [10.1198/004017006000000219](https://doi.org/10.1198/004017006000000219).
- Maruyama, Y. (2021) 'Post-truth AI and big data epistemology: From the genealogy of artificial intelligence to the nature of data science as a new kind of science', *Advances in Intelligent Systems and Computing*, 1181 AISC, pp. 540-549. doi: [10.1007/978-3-030-49342-4_52](https://doi.org/10.1007/978-3-030-49342-4_52).
- Mayo, D. (1996) *El error y el crecimiento del conocimiento experimental*. Chicago: University of Chicago Press.
- Mayo, D. (2018) *La inferencia estadística como prueba severa: Cómo ir más allá de las guerras estadísticas*. Nueva York: Cambridge University Press.

- Mazzocchi, F. (2015) "¿Podrían ser los macrodatos el fin de la teoría en la ciencia?", *EMBO reports*, 16(10), pp. 1250-1255. doi: [10.15252/embr.201541001](https://doi.org/10.15252/embr.201541001).
- Miller, H. J. (2010) 'The Data Avalanche Is Here. Shouldn't We Be Digging?', *Journal of Regional Science*, 50(1), pp. 181-201. doi: <https://doi.org/10.1111/j.1467-9787.2009.00641.x>.
- Mittelstadt, B. D. et al. (2016) 'La ética de los algoritmos: Mapping the debate', *Big Data & Society*, 3(2), p. 2053951716679679. doi: [10.1177/2053951716679679](https://doi.org/10.1177/2053951716679679).
- Napoletani, D., Panza, M. y Struppa, D. (2018) "La estructura agnóstica de los métodos de la ciencia de datos", p. 17.
- Neresini, F. (2017) 'Sobre datos, Big Data e investigación social. Es una verdadera revolución?', en Lauro, N. C. et al. (eds) *Data Science and Social Research*. Cham: Springer International Publishing (Estudios sobre clasificación, análisis de datos y organización del conocimiento), pp. 9-16. doi: [10.1007/978-3-319-55477-8_2](https://doi.org/10.1007/978-3-319-55477-8_2).
- Nie, X. y Wager, S. (2021) 'Quasi-oracle estimation of heterogeneous treatment effects', *Biometrika*, 108(2), pp. 299-319. doi:[10.1093/biomet/asaa076](https://doi.org/10.1093/biomet/asaa076).
- Niiniluoto, I. (2018) *Búsqueda de la verdad por abducción*. Cham, Suiza: Springer.
- Pearl, J. (2009) *Causalidad*. Cambridge: Cambridge University Press. doi: [10.1017/CBO9780511803161](https://doi.org/10.1017/CBO9780511803161).
- Peters, J., Janzing, D., & Schölkopf, B. (2017) *Los elementos de la inferencia causal: Fundamentos y algoritmos de aprendizaje*. Cambridge, MA: The MIT Press.
- Pietsch, W. (sin fecha) "Big Data - La nueva ciencia de la complejidad".
- Popper, K.R. (1959) *La lógica del descubrimiento científico*. Oxford, Inglaterra: Basic Books.
- Portmess, L. y Tower, S. (2015) "Data barns, ambient intelligence and cloud computing: the tacit epistemology and linguistic representation of Big Data", *Ethics and Information Technology*, 17(1), pp. 1-9.
- Powers, T. M. (2017) *Filosofía e informática: Ensayos de epistemología, filosofía de la mente, lógica y ética*. Springer.
- Prensky, M. (2009) "H. Sapiens Digital: From Digital Immigrants and Digital Natives to Digital Wisdom", p. 11.
- Ratti, E. y López-Rubio, E. (2018) 'MECHANISTIC MODELS AND THE EXPLANATORY LIMITS OF MACHINE LEARNING', *Machine Learning*, p. 18.
- Reichenbach, H. (1938) *Experiencia y predicción*. Disponible en: <https://philpapers.org/rec/REIEAP-2> (fecha de consulta: 14 de diciembre de 2020).

- Ribeiro, M. T., Singh, S. y Guestrin, C. (2016) "'Why Should I Trust You?': Explaining the Predictions of Any Classifier", *arXiv:1602.04938 [cs, stat]*. Disponible en: <http://arxiv.org/abs/1602.04938> (fecha de consulta: 24 de septiembre de 2020).
- Rieder, G. y Simon, J. (2016) "Datatrust: Or, the political quest for numerical evidence and the epistemologies of Big Data".
- Rudin, C. (2019) 'Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead', *Nature Machine Intelligence*, 1(5), pp. 206-215. doi: [10.1038/s42256-019-0048-x](https://doi.org/10.1038/s42256-019-0048-x).
- Samek, W., Wiegand, T. y Müller, K.-R. (2017) 'Explainable Artificial Intelligence: Understanding, Visualizing and Interpreting Deep Learning Models', *arXiv:1708.08296 [cs, stat]*. Disponible en: <http://arxiv.org/abs/1708.08296> (fecha de consulta: 24 de septiembre de 2020).
- Schmidt, M. y Lipson, H. (2009) 'Distilling Free-Form Natural Laws from Experimental Data', *Science*, 324(5923), pp. 81-85. doi: [10.1126/science.1165893](https://doi.org/10.1126/science.1165893).
- Schurz, S. (2019) *El problema de Hume resuelto: La optimalidad de la metainducción*. Cambridge, MA: The MIT Press.
- Scott, Z. (2018) *La crisis de reproducibilidad de la ciencia de datos*, *Medium*. Disponible en: <https://towardsdatascience.com/data-sciences-reproducibility-crisis-b87792d88513> (Consultado: 25 de septiembre de 2020).
- Shiffrin, R. M. (2016) 'Drawing causal inference from Big Data', *Proceedings of the National Academy of Sciences*, 113(27), pp. 7308-7309. doi: [10.1073/pnas.1608845113](https://doi.org/10.1073/pnas.1608845113).
- Spirtes, P., Glymour, C., & Scheines, R. (2000) *Causation, prediction, and search*. Cambridge, MA: The MIT Press.
- Steadman, I. (2013) "Big data and the death of the theorist", *Wired UK*, 25 de enero. Disponible en: <https://www.wired.co.uk/article/big-data-end-of-theory> (fecha de consulta: 17 de diciembre de 2020).
- Stupple, A., Singerman, D. y Celi, L. A. (2019) "La crisis de la reproducibilidad en la era de la medicina digital", *npj Digital Medicine*, 2(1), pp. 1-3. doi: [10.1038/s41746-019-0079-z](https://doi.org/10.1038/s41746-019-0079-z).
- Symons, J. y Alvarado, R. (2016) "¿Podemos confiar en Big Data? Applying philosophy of science to software", *Big Data and Society*, 3(2).
- Tsamados, A. et al. (2020) *La ética de los algoritmos: Key Problems and Solutions*. SSRN Scholarly Paper ID 3662302. Rochester, NY: Social Science Research Network. doi:[10.2139/ssrn.3662302](https://doi.org/10.2139/ssrn.3662302).
- Tukey, J. W. (1962) "El futuro del análisis de datos", en. doi: [10.1214/aoms/1177704711](https://doi.org/10.1214/aoms/1177704711).
- Turilli, M. y Floridi, L. (2009) "La ética de la transparencia informativa", *Ethics and Information Technology*, 11(2), pp. 105-112. doi: [10.1007/s10676-009-9187-9](https://doi.org/10.1007/s10676-009-9187-9).

- Waltz, D. y Buchanan, B. G. (2009) "Automating Science", *Science*, 324(5923), pp. 43-44. doi: [10.1126/science.1172781](https://doi.org/10.1126/science.1172781).
- Watson, D. S. y Floridi, L. (2020a) 'The explanation game: a formal framework for interpretable machine learning', *Synthese*. doi: [10.1007/s11229-020-02629-9](https://doi.org/10.1007/s11229-020-02629-9).
- Wheeler, G. (2016) "Machine Epistemology and Big Data", *Routledge Companion to Philosophy of Social Science*. Routledge (2016).
- Wigner, E.P. (1960) 'La irrazonable eficacia de las matemáticas en las ciencias naturales. Richard Courant lecture in mathematical sciences delivered at New York University, May 11, 1959', *Communications on Pure and Applied Mathematics*, 13(1), pp. 1-14. doi: [10.1002/cpa.3160130102](https://doi.org/10.1002/cpa.3160130102).
- Wu, C. F. J. (1997) 'datascience.pdf'.
- Zednik, C. (de próxima publicación) "Solving the Black Box Problem: A Normative Framework for Explainable Artificial Intelligence", *Philosophy and Technology*, pp. 1-24. doi: [10.1007/s13347-019-00382-7](https://doi.org/10.1007/s13347-019-00382-7).
- Zerilli, J. *et al.* (2019) 'Transparencia en la toma de decisiones algorítmica y humana: Is There a Double Standard?', *Philosophy & Technology*, 32(4), pp. 661-683. doi: [10.1007/s13347-018-0330-6](https://doi.org/10.1007/s13347-018-0330-6).