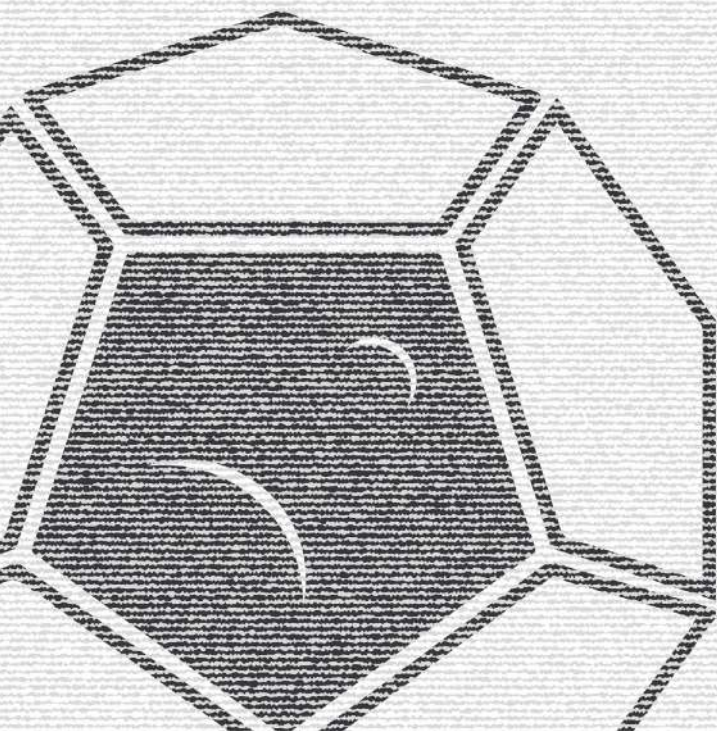


Maximiliano Bozzoli
Luis Salvatico
David Merlo
(Eds.)

Epistemología e Historia de la Astronomía

Volumen I



Epistemología e Historia de la Astronomía

Volumen I

Maximiliano Bozzoli

Luis Salvatico

David Merlo

(Eds.)

**Colecciones
del CIFYH**



Epistemología e historia de la Astronomía / Maximiliano Bozzoli ... [et al.]; compilación de Luis Salvatico; David C. Merlo. - 1a ed. - Córdoba : Universidad Nacional de Córdoba. Facultad de Filosofía y Humanidades, 2023.

Libro digital, PDF

Archivo Digital: descarga y online

ISBN 978-950-33-1721-1

1. Astronomía. 2. Historia. 3. Epistemología. I. Bozzoli, Maximiliano. II. Salvatico, Luis, comp. III. Merlo, David C., comp.
CDD 520.3

Publicado por

Área de Publicaciones de la Facultad de Filosofía y Humanidades - UNC

Córdoba - Argentina

1º Edición


Área de
Publicaciones

Diseño de portadas: Manuel Coll y María Bella

Diagramación: María Bella

Imagen portada: “JEHA (*Jornadas de Epistemología e Historia de la Astronomía*)” (2021), de Maximiliano Bozzoli

2023



Esta obra está bajo una Licencia Creative Commons
Atribución-NoComercial-SinDerivadas 4.0 Internacional.



Del universo de datos a los datos del universo.

Notas epistemológicas sobre el uso de inteligencia artificial en astronomía

Andrés A. Ilcic*

Resumen

El objetivo de este trabajo es presentar una lectura epistemológica de las prácticas que hacen uso de herramientas de inteligencia artificial en distintas ramas de la astronomía. La primera parte es una breve introducción a una posible historia de la astronomía como una ciencia basada en datos, señalando al concepto de dato —y al de metadato asociado— como un pilar fundamental sobre el que se puede apoyar una noción de garantía epistémica. Esto hace que sea central estudiar el rol que tanto modelos como técnicas tienen en la generación, el procesamiento y la interpretación de los datos. Se defiende que en lugar de la automatización del descubrimiento, lo que las técnicas contemporáneas permiten es un cambio cuantitativo significativo en la cantidad de dimensiones que se pueden considerar en un espacio de parámetros sobre el cual se podrán explorar las similitudes con datos provenientes de otras fuentes.

Palabras clave: *grandes datos, inteligencia artificial, aprendizaje automático, epistemología, robustez.*

Abstract

The aim of this paper is to present an epistemological overview of the practices that make use of artificial intelligence tools in different branches of astronomy. The first part is a brief introduction to a possible history of astronomy as a data-driven science, pointing to the concept of data—and its associated metadata—as the touchstone on which a notion of epistemic warrant can be based. This makes it central to study the role that both

* Centro de Investigaciones de la Facultad de Filosofía y Humanidades (UNC) – (CONICET).

models and techniques play in the generation, processing and interpretation of data. It is argued that rather than automation of discovery, what contemporary techniques allow is a significant quantitative change in the number of dimensions that can be considered at once in a parameter space over which similarities with data from other sources can be explored.

Keywords: *big data, artificial intelligence, machine learning, epistemology, robustness.*

Introducción

Dado el rol central que tiene la noción de “dato” en la astronomía, ligado al de observación, creo que bien puede tomarse como una disciplina paradigmática para llevar a cabo un análisis de la forma en la que dicho concepto se ha ido complejizando y ganando matices con el paso del tiempo, tanto en sus aspectos teórico-instrumentales y, como consecuencia de esta evolución, los epistemológicos. A su vez, “dato” es un término que también resuena en torno a la inteligencia artificial, especialmente dada la reciente disminución de los costos asociados a su almacenamiento y procesamiento; factores que contribuyeron al éxito de las arquitecturas de redes neuronales artificiales como técnica de exploración y manipulación de conjuntos enormes de datos (dando lugar a la era de los “*big data*”). De las múltiples dimensiones desde las que se pueden explorar las facetas epistemológicas que se observan cuando se presta atención al impacto que estas herramientas tecnológicas tienen en la práctica científica, en este trabajo me propongo examinar brevemente la manera en la que las afirmaciones acerca de la naturaleza de los fenómenos bajo estudio en la astronomía históricamente han dependido de avances técnicos (dentro de los que incluyo a las técnicas matemáticas) como soporte crucial de dichas afirmaciones. El énfasis de la reconstrucción que ofrezco aquí está en señalar los cambios cualitativos en las clases de afirmaciones y sus justificaciones que se pueden llevar a cabo acerca de la naturaleza a partir de las nuevas técnicas de manipulación de datos habilitadas por los avances en aquellas subdisciplinas englobadas en la actualidad bajo el rótulo de inteligencia artificial.

1. Grandes datos antes de los *big data*

El constante crecimiento (cuasi-exponencial) de la capacidad de cómputo que la humanidad ha experimentado especialmente desde la creación de los microprocesadores ha llevado, en conjunción con una substancial disminución de los costos de almacenamiento, a una superabundancia de datos. Esto se debe a que es cada vez más sencillo añadir sensores a dispositivos electrónicos que permiten generar un registro de su entorno o de sus condiciones de uso por parte de los agentes que interactúan con ellos. En diferentes contextos, este fenómeno reciente se dio a llamar “la era de los grandes datos” (los *big data*, en inglés), con constantes referencias a un “diluvio de datos”, especialmente en sectores de negocios y servicios de *software*. Esta “explosión” de datos en distintos ámbitos es heredera directa de las tecnologías de producción y procesamiento de datos en contextos científicos, aunque bien es necesario destacar que la investigación directa en estas técnicas también se ha visto impulsadas en los últimos años por las oportunidades comerciales que las industrias tecnológicas han logrado captar al nivel de los consumidores. Las ciencias computacionales, entre las que destacaría a la astronomía por el rol crucial que tiene en ella el concepto mismo de dato, se han visto beneficiadas profundamente por la retroalimentación entre los avances tecnológicos y la experticia entre los distintos sectores involucrados en la exploración de las fronteras computacionales e ingenieriles.

Es precisamente en esta *frontera* entre los algoritmos, los datos y la infraestructura computacional que emerge la llamada “ciencia de datos”; que antes bien creo que más bien puede caracterizarse como una disciplina que hace a una “zona de intercambio” de especialidades bajo el aspecto de conjunto de técnicas de resolución de problemas generales. Parte de esta generalidad o “universalidad” proviene del hecho de que en muchos casos las técnicas de procesamiento de datos pueden ser aplicadas a una gran parte del flujo de trabajo de manera “agnóstica”, en tanto suelen no depender de la semántica de dichos datos. Ahora bien, esto no nos deja frente a una “ciencia” agnóstica ni automática, como ha sido sugerida en algunas ocasiones (Napoletani et al., 2011). El rol de los expertos de dominio es más importante que nunca; debiendo involucrarse en las distintas etapas del “ciclo de vida” de los datos, especialmente en el desarrollo de técnicas de visualización científica que faciliten la exploración “final” de los datos

preseleccionados como significativos por el análisis exploratorio; solo así puede lograrse que dichos resultados sean interpretados e incorporados a los ciclos teóricos y experimentales. Uso el término “ciclo” para cubrir al menos parcialmente lo que en algunos contextos, habitualmente ligados a la computación científica y de alto desempeño, se conoce como “paradigma” (Hey et al., 2009), dado que la semántica habitual del término en filosofía de la ciencia sugiere una interpretación kuhneana en la que los paradigmas, vistos como “etapas epistémicas”, sugieren una confrontación radical hasta en los referentes fenoménicos. Al contrario, creo que la interpretación filosófica de la creación de nuevos ciclos o escenarios con metodologías propias de trabajo que, en lugar de reemplazar las formas anteriores de llevar a cabo las tareas científicas, habilita nuevas fronteras de exploración y obliga a complejizar la manera en la que se entiende la naturaleza y la justificación de los alcances de las afirmaciones de conocimiento acerca de los fenómenos, tanto naturales como artificiales.

Si bien en la literatura científica se suelen encontrar interpretaciones del giro tecnológico o que apuestan por un cambio radical en la manera de hacer ciencia, al estilo de “La investigación astronómica está pasando de estar basada en hipótesis a estarlo en datos [*data-driven*] y de manera intensiva [*data-intensive*] (Zhang y Zhao, 2015), es importante tomar con cuidado dichas afirmaciones. Un acercamiento epistemológico a las prácticas sugiere una mirada más cauta:

El intento de establecer una distinción tajante entre la ciencia basada en hipótesis y la ciencia intensiva en datos es inadecuado; estos modos de investigación no son, de hecho, ortogonales y a menudo se entremezclan en la práctica científica concreta. (Elliott et al., 2016, p. 881)

Ahora bien, sí es claro que la manera de entender qué es lo que constituye conocimiento científico se ve ampliamente afectado y complejizado a medida que el soporte tecnocientífico de las disciplinas se vuelve cada vez más extenso, profundo y ubicuo. Además, se trata de una transformación que no solo afecta a las llamadas ciencias naturales: las ciencias sociales pueden verse beneficiadas y afectadas de modos muy singulares en términos de los métodos que la abundancia de datos y las mejoras para su procesamiento pueden traer. Basta solo recordar cómo los orígenes técnicos del procesamiento de datos estuvieron ligados a hacer posible en tiempo y forma el relevamiento poblacional del censo de los Estados

Unidos de 1890: la máquina tabuladora de Herman Hollerith. Además de proveer la manera en la que se interactuaría con las computadoras durante varias décadas por venir, las tarjetas perforadas de Hollerith (basadas en las tempranas implementaciones de Jacquard y Babbage acerca de cómo configurar valores de entrada para una máquina) permitían representar (o *reducir*) un individuo a una matriz de características cuyo soporte tecnológico era el medio para realizar inferencias generales acerca de la población. Siguiendo una línea parcialmente sugerida por, entre otros, Paul Humphreys (1995, 2004) con respecto al rol de la tecnología en el análisis epistemológico de una ciencia que recurre a ella, toda clase de soporte tecnocientífico que vuelva factibles *en la práctica* algunas inferencias que antes no eran tratables debe verse como capaz de transformar tanto una ciencia al punto de incluso cambiar los criterios sobre los que se establecen las credenciales epistémicas. Si bien hay numerosas formas de aproximarse y construir una epistemología de las simulaciones computacionales en ciencia (Durán, 2018; Morrison, 2015; Weisberg, 2013; Winsberg, 2010), creo que los avances en las distintas disciplinas requieren de una exploración pormenorizada de la manera en la que en cada comunidad se construye y juzga al conocimiento, en un sentido similar al propuesto por Ian Hacking (1992) originalmente como “autojustificación” para las ciencias de laboratorio. En esta dirección, los análisis deben concentrarse en las metodologías de investigación y aplicación en cada dominio y casos particulares (García, 2015; Oberkamp y Roy, 2010).

1.1. Ciberinfraestructura

Un aspecto crucial para comprender las metodologías de trabajo, por tanto, es incorporar en los análisis lo que podríamos llamar los aspectos de la materialidad, en un sentido más próximo a lo que en estudios sociales y filosóficos de la técnica se conoce como “cultura material” (Preston, 2013). Sin duda, esto es algo que va de la mano del giro que puso de manifiesto la importancia de las prácticas experimentales (Hacking, 1983; Radder, 1988/2012) y simulacionales (Galison, 1996) para una adecuada conceptualización filosófica de las prácticas científicas.

En particular, la preocupación por la infraestructura tecnológica debe incorporar los avances en materia de las interfaces que se generan entre los distintos grupos de usuarios y los usuarios con cada uno de los “pro-

ductos de datos”, como suelen llamarse a las colecciones curadas publicadas por un grupo de investigación asociado a un experimento o proyecto para que pueda ser empleado por la comunidad científica general.

Ya a principios de este milenio, un reporte para la *National Science Foundation* (NSF) remarcaba la importancia de pensar políticas de ciberinfraestructura científica a largo plazo y de profundo impacto en las comunidades, no limitada a la mera creación de centros de computación de alto desempeño:

La visión que está surgiendo es la de utilizar las ciberinfraestructuras para construir entornos digitales más ubicuos y exhaustivos que se conviertan en interactivos y funcionalmente completos para las comunidades de investigación en términos de personas, datos, información, herramientas e instrumentos y que operen a niveles sin precedentes de capacidad de cálculo, almacenamiento y transferencia de datos. Cada vez más, los nuevos tipos de organizaciones científicas y entornos de apoyo a la ciencia son esenciales, no opcionales, para las aspiraciones de las comunidades de investigación y para ampliar la participación en esas comunidades. Pueden servir a individuos, equipos y organizaciones de manera que revolucionen lo que pueden hacer, cómo lo hacen y quiénes participan. Esta visión también tiene profundas implicaciones más amplias para la educación, el comercio y el bien social. (Atkins et al., 2003, p. 2)

Habitualmente un criterio historiográfico que puede tomarse para remarcar un punto de transición entre una “preocupación” en una comunidad científica por una actitud que lleva a la consolidación de un campo de investigación particular (atendiendo a que lo que para un dominio de investigación puede ser una mera técnica o un recurso tecnológico bien puede, y suele, ser el objeto de estudio propio de otra comunidad) es el surgimiento de conferencias dedicadas al encuentro de especialistas del tema. Puede decirse que es en la década de 1980 cuando estos intereses se condensan en cambios en las metodologías de trabajo, pudiéndose observar un giro hacia la interdisciplinariedad, especialmente a través de la catálisis ofrecida por la *Conferencia de Gestión de Bases de Datos Científicos y Estadísticos* (la *Scientific and Statistical Database Management Conference* o SSDBM), que tiene lugar desde 1981, incluyendo a especialistas de áreas de la informática y de la matemática, como también de las disciplinas que las encuentran como aplicaciones. El campo de la visualización científica propiamente así llamado puede decirse que surge en 1987, con

un *workshop* dedicado al tema (McCormick et al., 1987) en la conferencia SIGGRAPH (*Special Interest Group on Computer Graphics and Interactive Techniques*), cuya inauguración se remonta a 1974. Cabe destacar que las conferencias de esta clase cada vez más han tenido un peculiar rol como espacios de creación de comunidades, especialmente entre la academia y la industria, algo que también debe ser atendido por filósofos, historiadores y sociólogos de la ciencia y la tecnología. La visualización científica es hoy en día un aspecto crucial para el avance de las disciplinas que trabajan con grandes volúmenes de datos (Ilcic y Reynoso, 2015), como lo es también la creación de interfaces por medio de las cuales los investigadores puedan llevar a cabo una obtención rápida y posterior integración de datos de diferentes fuentes.

En varios sentidos, se puede afirmar que la astronomía fue pionera en reconocer la importancia de fijar criterios de estándares tecnológicos para habilitar la colaboración científica de distintas comunidades; quizás uno de los casos más importantes en esta dirección de estándares de infraestructura sea la propuesta original para la creación de los observatorios virtuales (OV) (Szalay y Gray, 2001). Parte del éxito científico de los OV se debe a la manera en la que permitieron la reconfiguración de las comunidades de investigación, pudiendo cada vez más integrar grupos de trabajo de otras disciplinas y de distintas regiones del planeta:

El OV es un entorno de investigación emergente, abierto, basado en la web y distribuido para la astronomía con conjuntos de datos masivos y complejos. Reúne archivos y servicios de datos, así como herramientas de exploración y análisis de datos. Se trata de un entorno tecnológico, pero impulsado por la ciencia, que ofrece excelentes oportunidades de colaboración entre astrónomos y profesionales de la informática y la estadística. También es un ejemplo de un nuevo tipo de organización científica, intrínsecamente distribuida, intrínsecamente multidisciplinar, con un espectro inusualmente amplio de colaboradores y usuarios. (Djorgovski y Williams, 2005, p. 2)

De menor escala, pero no por eso menos importante, otro claro ejemplo es el formato de imágenes FITS, en el que se puede apreciar la importancia de los metadatos que describen las condiciones de observación. De hecho, si bien puede usarse para imágenes, el formato es tan flexible que es mejor describirlo como una base de datos que incorpora en su definición una *interfaz de uso* pensada para soportar los desafíos de la clase de

investigación en la cual las afirmaciones de conocimiento de un objeto o evento astronómico dependen de la reconstrucción posterior de una gran serie de observaciones fragmentadas en múltiples dominios. Algo semejante puede decirse acerca del *bicode* o *refcode*, estándar que emergió tras la necesidad de crear identificadores únicos a las referencias bibliográficas en astronomía e incorporado a los distintos sistemas de datos, en particular al Astrophysics Data System (ADS).

2. Grandes datos, astroinformática e inteligencia artificial

Los observatorios dedicados a realizar censos del cielo (*sky surveys*, en inglés) son uno de los más claros indicadores del desafío tecnocientífico asociado a la labor científica que involucra grandes bases de datos. Por ejemplo, se espera que el Observatorio Vera Rubin, actualmente en construcción en Chile por un esfuerzo internacional —y conocido hasta hace poco como LSST (por *The Large Synoptic Survey Telescope*)—, produzca en total unos 500 petabytes de datos. El *Square Kilometre Array*, que podría dar su primera luz en 2027, generaría un total planificado de aproximadamente 4 exabytes.

Solo las dificultades asociadas al almacenamiento y al transporte de un volumen tan significativo de datos obligan a montar equipos con gran capacidad de procesamiento para preseleccionar los datos que finalmente se almacenan y sobre los cuales se llevan a cabo las tareas posteriores de descubrimiento, además de proveer medios para el análisis de datos en tiempo real, que muchas veces son críticos para la investigación de eventos astronómicos de corta duración o aquellos novedosos en tanto no pudieron ser predichos de manera de planificar observaciones o misiones programadas. Aquí ya se puede observar la necesidad de comprender cómo las decisiones metodológicas implicadas en el diseño de un instrumento de captura de señales deben partir de un consenso por parte de la comunidad científica involucrada para determinar, a través de lo que puede verse como un juicio de valor, qué es lo que debe priorizarse con los recursos (técnicos, económicos y epistémicos) en un momento determinado.

El reciente auge en la aplicación de inteligencia artificial (IA) en contextos científicos como técnica de descubrimiento es heredero de múltiples líneas de trabajo, repletas tanto de éxitos como fracasos, que abarcan muchas décadas y un sinnúmero de disciplinas. La aplicación exitosa

de arquitecturas de IA basada en redes neuronales a datos astronómicos puede situarse a principios del siglo XXI, siendo (Tagliaferri et al., 2003) un claro ejemplo. Sin embargo, es recién con la aparición de las llamadas arquitecturas de aprendizaje profundo o *deep learning* (LeCun et al., 2015) que se puede apreciar un cambio más radical en cuanto al tipo de estrategias de indagación de conjuntos de datos que se habilita. Ahora bien, la transformación no está tanto en términos de la automatización del descubrimiento *per se*, sino en la creación de nuevas interfaces por medio de las cuales es posible dar sentidos a los datos y manipular o transformar *tipos* de datos, todo ello con una facilidad que hace apenas unos años era impensada.

Si bien para muchos fenómenos la cantidad de datos disponibles y ya clasificados por expertos (o mediante una tarea simplificada en la que se pide a voluntarios humanos colaboración en la tarea por medio de aplicaciones web), en otros casos las técnicas de manipulación de imágenes facilitadas por las tecnologías de aprendizaje automático profundo permiten generar datos sintéticos sobre los que luego se pueden entrenar otros algoritmos que podrán ser utilizados para encontrar patrones significativos en datos nuevos, entre los que se destaca la detección de exoplanetas por imágenes directas, una tarea muy difícil incluso para expertos por la debilidad de la señal con respecto al ruido (ver fig. 1).

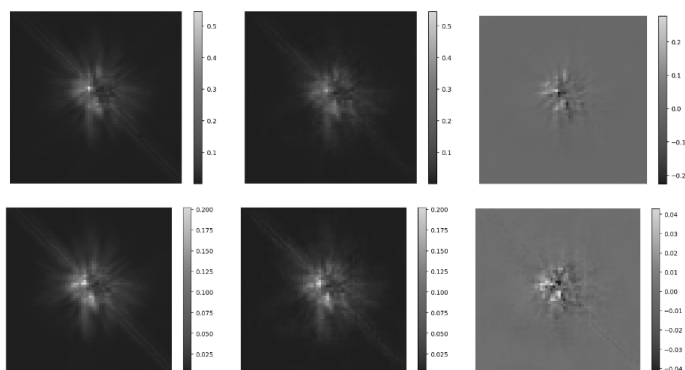


Figura 1. A la izquierda se observa el patrón de moteado que surge al filtrar la luz de la estrella a la que orbita un exoplaneta observado de manera directa. La columna del medio muestra la reconstrucción artificial de una imagen semejante por medio de una red adversaria generativa. La columna final es un mapa de las diferencias o del “grado de similitud” entre las anteriores a partir del valor de cada píxel (Yip et al., 2020).

Hay una gran cantidad de literatura científica que puede consultarse como una introducción al estado actual de las técnicas de grandes datos en astronomía (e.g. Kong et al., 2020; Sachdeva et al., 2022; Skoda y Adam, 2020; Zhang y Zhao, 2015). La pregunta a la que busco tornar brevemente ahora es la de qué significa esta nueva clase de técnicas de exploración del vasto universo de datos para nuestras afirmaciones de conocimiento del universo.

3. De los datos a los fenómenos (y de regreso)

Si bien los filósofos han prestado bastante atención al uso de técnicas estadísticas en las ciencias e incluso algunos toman ciertas interpretaciones (epistémicas, habitualmente bayesianas) como una forma de dar cuenta o explicar la racionalidad misma de la ciencia y de las dinámicas de cambio teórico, muchas de las técnicas que se aplican en la actualidad tienen como objetivo transformar o reducir en dimensiones la gran complejidad de los datos para facilitar la tarea de encontrar en ellos patrones relevantes y no, como en las técnicas más clásicas, intentar poner a prueba una serie de hipótesis acerca de las distribuciones de probabilidad que podrían describir

el proceso de generación subyacente a los datos. Por esta razón, se trata todavía de un campo muy fértil para la exploración filosófica, especialmente aquellas que presten atención a las prácticas científicas y, en ellas, al uso de modelos.

Desde mi punto de vista, creo que el principal cambio que las nuevas técnicas de IA permiten está en la escala en la que se puede llevar a cabo una exploración sistemática de un espacio hiperdimensional de parámetros posibles con el objetivo de encontrar en ellos relaciones no triviales o patrones que puedan volverse significativos. Los algoritmos para hacer agrupamientos [*clustering*], clasificación, análisis de correlación e identificación de valores atípicos [*outliers*] dependen de lo que se asuma como objetos o procesos típicos en el dominio de trabajo. Detectar una anomalía en un conjunto de datos es una tarea primordial, puesto que puede indicar tanto un error en alguna de las etapas computacionales del procesamiento de datos o en el funcionamiento mismo de los instrumentos involucrados en la observación, mientras que al mismo tiempo puede revelar indicios de la existencia de un nuevo objeto de interés científico.

Quizás lo más importante para una lectura epistemológica de estas prácticas es que, cuando se las mira de manera generalizada, lo que se encuentra es la necesidad de establecer una métrica de similitud (ya sea una distancia o una densidad) en lo que podríamos llamar un “espacio de características” sobre el cual se pueden llevar a cabo comparaciones con datos provenientes de diferentes fuentes, como lo es, en el caso de la astronomía, la posibilidad de combinar observaciones de distintos rangos del espectro electromagnético, además de aquellas que puedan llegar a detectarse a partir de otra clase de interacción fundamental conocida, como las ondas gravitacionales. De esta manera, la IA como técnica permite un apreciable aumento en la cantidad de características que pueden evaluarse al mismo tiempo y la posibilidad de validar los resultados con simulaciones que tomen en cuenta las idiosincrasias de los instrumentos de detección; es solo a través de esta reconstrucción que se puede proceder a una proyección ontológica de los datos a un fenómeno que puede tomarse como causa u origen de una señal. En la proyección, por tanto, entra tanto el conocimiento que se pueda llegar a tener de la señal como del ruido sobre el que la señal se detecta. En esta dirección, estoy de acuerdo con James Woodward, quien en un importante artículo remarcaba que,

por lo general, una variedad de factores causales, muchos de ellos relacionados con los detalles de un dispositivo experimental o de detección concreto, desempeñarán un papel en la producción de un dato determinado o en la evaluación de su importancia como prueba. (Woodward, 1989, p. 395)

Quizás la forma más clara de apreciar esta tesis acerca de la exploración de un espacio de parámetros en busca de patrones significativos es apreciando un caso extremadamente simple cuya fertilidad epistémica sigue siendo sorprendente: un diagrama de Hertzsprung-Russell (fig. 2).

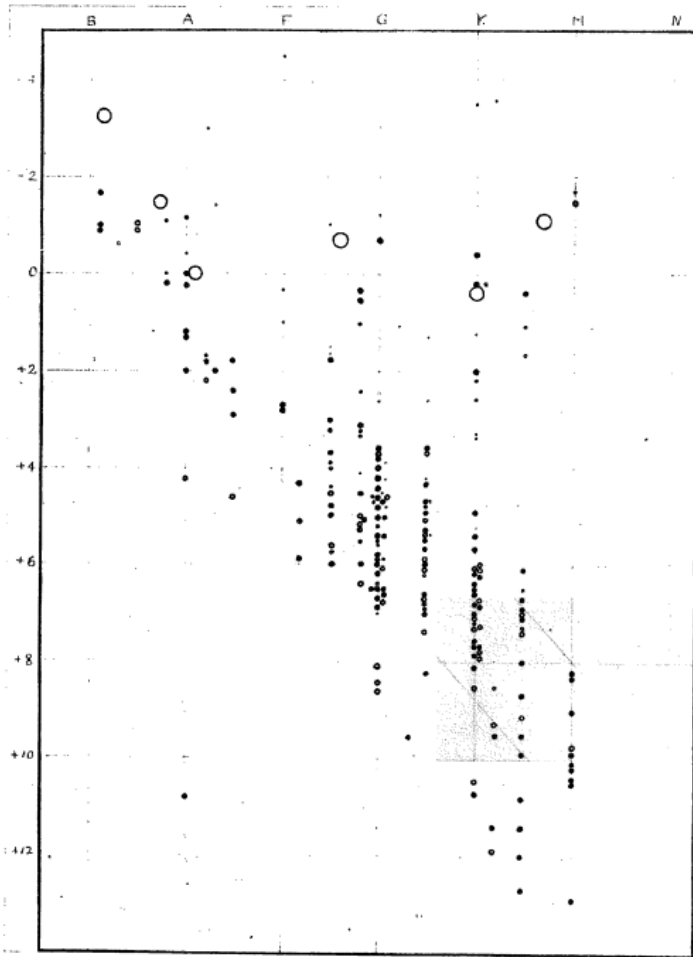


Figura 2. Diagrama de Hertzsprung-Russell en la publicación original de H.N. Russell (1914, p. 285). Sobre la clasificación estelar introducida por Antonia C. Mauri, una de las “Computadoras de Harvard”, Hertzsprung señaló en 1908 la relación por el tipo de agrupamiento en torno a la luminosidad. En base a esto, el primer diagrama propiamente dicho fue publicado por su colega Hans Rosenberg dos años después (Rosenberg, 1910).

En este gráfico de dispersión se puede apreciar la relación no trivial entre una serie de parámetros, como pueden ser la luminosidad de una serie de estrellas con respecto a su clasificación espectral o temperatura efectiva, lo cual ya sugiere un entramado entre una serie de observaciones plasmadas *como* datos y un conocimiento de fondo acerca de las observaciones y de los fenómenos que permiten la creación del diagrama mismo; que ahora puede ser utilizado para nuevas inferencias acerca de lo que están representando. Claramente, el hecho crucial está en que la relación entre estos parámetros no es trivial, es decir, no cubren todo el plano y permiten, por tanto, la postulación de hipótesis acerca de los eventuales fenómenos que están detrás de esta particular distribución. Mediante las técnicas de manipulación de datos contemporáneas, lo que se habilita es una exploración de muchos más factores o parámetros al mismo tiempo, cuyo sentido físico puede juzgarse solo en términos de la similitud con resultados de simulaciones y experimentos de otras áreas de la ciencia. A esto hemos dado a llamar una “articulación de modelos” (Ilcic y Reynoso, 2018), aunque claramente hay que atender a que toda articulación es siempre muy provisoria y simplemente un paso más en la búsqueda de nuevas y más profundas relaciones entre los factores que se pueda correlacionar.

Conclusiones

Si bien es posible que en el futuro las técnicas de aprendizaje maquínico puedan automatizar cada vez más algunos aspectos del proceso de descubrimiento de conocimiento científico en bases de datos, de momento su principal potencial está en su capacidad de funcionar como herramientas sobre las cuales se pueden crear *interfaces* o *espacios de interacción* para que los expertos de distintas disciplinas estudien los datos y hagan inferencias acerca de los procesos que podrían haberlos generado. Es precisamente en esta *inferencia explicativa* a la causa más probable que pueda postularse como origen de la señal que se realiza una proyección ontológica para intentar justificar el paso de las afirmaciones sobre datos a los eventos o procesos de la naturaleza.

El énfasis en los datos cambia el tipo de prácticas asociadas a la validación de instrumentos y simulaciones, presentando tanto numerosos desafíos metodológico-teóricos e ingenieriles, como una infinitud de posibilidades para la exploración sistemática de un cada vez mayor espacio

de parámetros que recorrer en busca de relaciones o parámetros significativos sobre los cuales llevar a cabo una proyección ontológica acerca de los tipos de procesos naturales que podrían subyacer, desde las mejores teorías disponibles; eso sí, siempre atentos a las anomalías que, necesariamente, ocurrirán. Lo que esta perspectiva sobre los aspectos materiales deja entrever es que, allí en donde se encuentran teorías, observaciones y simulaciones, se genera una interfaz solo tras numerosas decisiones de diseño a base de las restricciones impuestas tanto por la materialidad técnica como por el conocimiento disponible.

Agradecimientos

Quiero expresar mis agradecimientos a los organizadores y a los asistentes de las 1ras Jornadas de Epistemología e Historia de la Astronomía por sus valiosos trabajos y conversaciones supralunares. Los cuidadosos comentarios hechos por un/a revisor/a anónimo de la versión original de este trabajo fueron de vital importancia para mejorar el ajuste de parámetros necesario para reducir el nivel de ruido en la presentación escrita de estas ideas, las anomalías siguen siendo mi culpa y material para trabajos futuros. Parte de esta investigación se llevó a cabo en el marco de los proyectos FONCyT-PICT-2016-1524, Consolidar de SECyT-UNC (2018-2022), y una beca doctoral interna de CONICET, convocatoria 2015.

Bibliografía

- Atkins, D. E., Droegemeier, K. K., Feldman, S. I., Garcia-Molina, H., Klein, M. L., Messerschmitt, D. G., Messina, P., Ostriker, J. P., y Wright, M. H. (2003). *Revolutionizing science and engineering through cyberinfrastructure: Report of the National Science Foundation blue-ribbon advisory panel on cyberinfrastructure* (pp. 203-226). National Science Foundation.
- Bozzoli, M. (2015). Abducción, evidencia y observación en astronomía. En V. Rodríguez, M. Velasco, & P. García (Eds.), *Epistemología y prácticas científicas* (pp. 73-98). Imprenta de la FFyH, UNC.

- Djorgovski, S. G., y Williams, R. (2005). *Virtual observatory: From concept to implementation* (arXiv:astro-ph/0504006). arXiv. <https://doi.org/10.48550/arXiv.astro-ph/0504006>
- Durán, J. M. (2018). *Computer simulations in science and engineering: Concepts - practices - perspectives*. Springer International Publishing. // www.springer.com/gp/book/9783319908809
- Elliott, K. C., Cheruvilil, K. S., Montgomery, G. M., y Soranno, P. A. (2016). Conceptions of good science in our data-rich world. *BioScience*, 66(10), 880-889. <https://doi.org/10.1093/biosci/biw115>
- Galison, P. (1996). Computer simulations and the trading zone. En P. Galison & D. J. Stump (Eds.), *The disunity of science: Boundaries, contexts, and power* (pp. 118-157). Stanford University Press.
- García, P. (2015). Simulaciones computacionales y modelos: consideraciones metodológicas. En V. Rodríguez, M. Velasco, & P. García (Eds.), *Epistemología y prácticas científicas* (pp. 117-134). Imprenta de la FFyH, UNC.
- Hacking, I. (1983). *Representing and intervening: introductory topics in the philosophy of natural science*. Cambridge University Press.
- Hacking, I. (1992). The self-vindication of the laboratory sciences. En A. Pickering (Ed.), *Science as practice and culture* (29-64). University of Chicago Press.
- Hey, T., Tansley, S., y Tolle, K. (Eds.). (2009). *The fourth paradigm: Data-intensive scientific discovery*. Microsoft Research.
- Humphreys, P. (1995). Computational science and scientific method. *Minds and Machines*, 5(4), 499-512. <https://doi.org/dttx9w>
- Humphreys, P. (2004). *Extending ourselves: computational science, empiricism, and scientific method*. Oxford University Press.

- Ilcic, A. A., y Reynoso, J. (2015). Si de entender se trata: el rol de la visualización en el procesamiento de datos. En V. Rodríguez, P. García, & M. Velasco (Eds.), *Epistemología y prácticas científicas* (pp. 171--185). Imprenta de la FFyH, UNC.
- Ilcic, A. A., y Reynoso, J. (2018). Hacia una articulación de modelos: el caso de big data. En S. Chibeni, L. Zaterka, J. Ahumada, & D. Letzen (Eds.), *Filosofía e Historia de la Ciencia en el Cono Sur* (pp. 181-192). Editorial de la Universidad Nacional de Córdoba. <http://www.afhic.com/wp-content/uploads/2019/01/hacia-una-articulaci%C3%B3n-de-modelos.pdf>
- Kong, L., Huang, T., Zhu, Y., y Yu, S. (2020). *Big data in astronomy: Scientific data processing for advanced radio telescopes*. Elsevier.
- LeCun, Y., Bengio, Y., y Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436-444. <https://doi.org/bmqp>
- McCormick, B. H., DeFanti, T. A., y Brown, M. D. (1987). Visualization in scientific computing: a synopsis. *IEEE Computer Graphics and Applications*, 7(7), 61-70. <https://doi.org/10.1109/MCG.1987.277014>
- Morrison, M. (2015). *Reconstructing reality: Models, mathematics, and simulations*. Oxford University Press.
- Napoletani, D., Panza, M., y Struppa, D. C. (2011). Agnostic science. Towards a philosophy of data analysis. *Foundations of Science*, 16(1), 1-20. <https://doi.org/cw6c9n>
- Oberkampff, W. L., y Roy, C. J. (2010). *Verification and validation in scientific computing*. Cambridge University Press.
- Pasquinelli, M. y Joler, V. (2021). El Nooscopio de manifiesto: La inteligencia artificial como instrumento de extractivismo del conocimiento (J. Blanco, A. Berti, & A. A. Ilcic, Trans.). *laFuga*, 25.

- <https://lafuga.cl/el-nooscopio-demanifiesto/1053> (Obra original de 2020)
- Preston, B. (2013). *A philosophy of material culture: action, function, and mind*. Routledge.
- Radder, H. (2012). *The material realization of science: from habermas to experimentation and referential realism*. Springer. (Obra original de 1988)
- Rosenberg, H. (1910). *Über den Zusammenhang von Helligkeit und Spektraltypus in den Plejaden*. <https://doi.org/10.1002/asna.19101860503>
- Russell, H. N. (1914). Relations between the spectra and other characteristics of the stars. *Popular Astronomy*, 22, 275-294. <https://ui.adsabs.harvard.edu/abs/1914PA.....22..275R>
- Sachdeva, S., Watanobe, Y., y Bhalla, S. (2022). *Big-data-analytics in astronomy, science, and engineering: 9th International Conference on Big Data Analytics, BDA 2021, Virtual Event, December 7-9, 2021, Proceedings*. Springer Nature.
- Skoda, P., y Adam, F. (2020). *Knowledge discovery in big data from astronomy and earth observation: Astrogeoinformatics*. Elsevier.
- Szalay, A. y Gray, J. (2001). The world-wide telescope. *Science*, 293(5537), 2037-2040. <https://doi.org/10.1126/science.293.5537.2037>
- Tagliaferri, R., Longo, G., Andreon, S., Capozziello, S., Donalek, C., y Giordano, G. (2003). Neural networks for photometric redshifts evaluation. En B. Apolloni, M. Marinaro, & R. Tagliaferri (Eds.), *Neural Nets* (pp. 226-234). Springer. https://doi.org/10.1007/978-3-540-45216-4_26

- Turing, A. M. (2004). Lecture on the Automatic Computing Engine. En B. J. Copeland (Ed.), *The essential Turing* (pp. 378-394). Clarendon Press. (Obra original de 1947)
- Weisberg, M. (2013). *Simulation and similarity: using models to understand the world*. Oxford University Press.
- Winsberg, E. B. (2010). *Science in the age of computer simulation*. The University of Chicago Press.
- Woodward, J. (1989). Data and phenomena. *Synthese*, 79(3), 393-472. <https://doi.org/b8mr9h>
- Yip, K. H., Nikolaou, N., Coronica, P., Tsiaras, A., Edwards, B., Changeat, Q., Morvan, M., Biller, B., Hinkley, S., Salmond, J., Archer, M., Sumption, P., Choquet, E., Soummer, R., Pueyo, L., y Waldmann, I. P. (2020). *Pushing the limits of exoplanet discovery via direct imaging with deep learning* (Vol. 11908, pp. 322-338). https://doi.org/10.1007/978-3-030-46133-1_20
- Zhang, Y., y Zhao, Y. (2015). Astronomy in the big data era. *Data Science Journal*, 14(0), 11. <https://doi.org/10.5334/dsj-2015-011>

