

XV Jornadas de Sociología. Carrera de Sociología, Facultad de Ciencias Sociales, UBA.

Eje 1. Filosofía, Teoría, Epistemología, Metodología

Mesa 7. Ciencias sociales computacionales

Tensiones y concepciones epistemológicas en torno al big data y la inteligencia artificial

Gastón Becerra (UBA/ CONICET) gastonbecerra@sociales.uba.ar

Una manera de ordenar y sistematizar las controversias epistemológicas que observamos en torno al big data y la inteligencia consiste en proponer una serie de dualidades o tensiones que, generalmente, las vinculan con discusiones más amplias del campo de la epistemología y la filosofía de la ciencia. Así, por caso, en torno a la naturaleza del dato y la objetividad, se discute sobre representación/relación, dato crudo/interpretado, o hecho/valor. Por su parte, sobre la manera de recortar objetos y objetivos, se discute sobre conducta/sujeto, o sobre explicación/funcionamiento. En relación a los fundamentos metodológicos y epistemológicos se discute sobre dato/teoría, correlación/causalidad, o información/conocimiento. Además, tendiendo un puente entre campos, podríamos incluir la teoría social que aporta tensiones como humano/post-humano, inteligencia/comunicación, o transparencia/explicabilidad, que redundan en otras que le son propias como individuo/sociedad, descripción/crítica. En este trabajo buscamos mapear estas tensiones distinguiendo, además, maneras de abordarlas, ya sea porque se busca reducir la tensión considerando sólo uno de polos, o remarcar el vínculo indisoluble de sus polaridades.

1. Introducción

No debería sorprendernos la magnitud de la discusión epistemológica en torno al big data y la inteligencia artificial. Esto se debe, entre otras razones, a que podemos entenderlas como tecnologías en las que lo epistémico es el rasgo distintivo de sus contextos de aplicación y usos pretendidos, de los materiales con los que se trabaja, y el tipo de operaciones que despliegan (Alvarado, 2023). También porque su desarrollo ha venido acompañado de una retórica –una mitología aclaran boyd y Crawford (2012)– que pretende cuestionar lo que entendemos por conocimiento y cuáles son los criterios con los que lo evaluamos, llegando a sostener que ofrecerían formas superiores de inteligencia, mayor objetividad, o mayor capacidad de acción, o diagnosticar crisis y revoluciones paradigmáticas. Finalmente,

también se asiste a un reordenamiento de la división social del trabajo cognitivo, asumiendo una mayor confianza y dependencia en las corporaciones que impulsan estos desarrollos.

Atender a esta discusión es uno de los objetivos centrales de la sociología interesada por el big data, los algoritmos, la inteligencia artificial, o por la manera en que estas se integran en las ciencias sociales computacionales. Para aportar a ella, en este trabajo exploramos una manera de ordenar y sistematizar la cuestión proponiendo “tensiones” y “concepciones”¹. Con las primeras nos referimos a la manera en que se ha estructurado la discusión epistemológica reciente, haciendo uso de dicotomías o dualidades tales como teoría/dato, causalidad/correlación, hecho/valor, entre muchas otras. Una mirada transversal a ellas da cuenta de la multidimensionalidad de la discusión epistemológica, que abarca problemas como la naturaleza del dato, las características epistémicas de los análisis algorítmicos y sus fundamentos, la manera en que se recorta el objeto y la unidad del análisis, y el importante problema de los objetivos, la orientación y la justificación de la investigación con big data y la implementación de inteligencia artificial. Con las segundas, nos referimos a una manera particular de encarar el análisis en las tensiones, ya no buscando posiciones, sino más bien observando cómo se problematiza el dualismo. Particularmente, argumentaremos que hay dos concepciones generales que cruzan distintos dualismos, una de ellas buscando preservar sólo uno de los términos de la dualidad implicada, y la otra buscando su integración para dar con una mirada más compleja. La primera de las concepciones la rastreamos al empirismo, mientras que la segunda la ejemplificamos con los *Critical data studies*.

2. Tensiones y concepciones

Para comenzar queremos sistematizar algunos antecedentes que han optado por proponer pares de conceptos dicotómicos como marco para posicionar y comparar las distintas voces en las disputa epistemológica del big data y la inteligencia artificial. Así, por caso, Kitchin (2014) en el número inaugural de *Big Data & Society* analiza las nuevas expresiones empiristas que hablan de una revolución del big data con pares de conceptos epistemológicos que interesan a una ciencia social computacional en ciernes, tales como correlación/causación, explicación/predicción, subjetivo/objetivo, muestreo/exhaustividad (pretendida), conocimiento universal/disciplinar. Resnyansky (2019) apunta a la misma

¹ Estas nociones se han desarrollado en colaboración con Pedro Giordano y José Antonio Castorina, a quienes se deben eximir de las imprecisiones en el presente trabajo. En las secciones siguientes se referencian los antecedentes más recientes de estas colaboraciones. Quien los consulte podrá encontrar en ellos intentos previos, reconceptualizaciones, recontextualizaciones, líneas abiertas de trabajo, discusiones con otras propuestas conceptuales alternativas, y otras problematizaciones que aquí no llegamos a incluir.

discusión sobre una revolución del conocimiento, señalando que la misma se sostiene sobre dicotomías tales como causación/correlación, razonamiento inductivo/deductivo, preeminencia de hipótesis/datos en el diseño de la investigación, o que su alcance discurre entre aplicación universal/disciplinaria y la ubicuidad/visibilidad. Wagner-Pacifi, Mohr, y Breiger (2015) buscan ordenar las preguntas sobre la ontología, metodología y usos del big data a través de 3 binarismos: vida/dato tendiente a discernir la primera de las dimensiones, inducción/deducción para estimar el modelo de ciencia a seguir, y mente/máquina para referir al problema de la interpretabilidad y la inteligibilidad de los resultados de los análisis algorítmicos. Esta misma cuestión es tratada por Leonelli (2020) en su entrada en la *Stanford Encyclopedia of Philosophy* con el par inteligencia humana/artificial, además de enfrentar las concepciones representacionistas/relacionistas en cuanto a la naturaleza del dato, recurrir a la clásica distinción entre hecho/valores de la filosofía de la ciencia, y discutir si el objetivo es generar información/conocimiento. También, para delimitar y caracterizar el campo de las ciencias de datos, Floridi y otros (Desai et al., 2022) distinguen entre definiciones minimalistas/maximalistas, de acuerdo a si se atiende o no ciertos supuestos ontológicos y ciertas técnicas metodológicas, una ciencia de datos más grande/chica con estadísticas menores/mayores, de acuerdo al alcance de las tareas consideradas, y la caracterización del campo entre un rama de la estadística/una disciplina científica propiamente dicha. Finalmente, en *Feminismo de datos* D'Ignazio y Klein (2020) argumentan a favor de problematizar el fenómeno del poder en los datos, cruzando la distinción entre ética/justicia, corriendo el foco entre sesgo/opresión, desplazando el objetivo entre imparcialidad/equidad, la reflexión entre responsabilidad/co-liberación, y eventualmente entre comprender los algoritmos/comprender la historia, la cultura y el contexto. Las autoras también advierten que los sistemas de conteo y clasificaciones usuales de los datos suelen incurrir en “falsos binarismos” –tales como razón/emoción, naturaleza/cultura, cuerpo/mundo, hechos/valores o tradicional/moderno– que esconden valores sociales e imponen jerarquías.

En todos estos antecedentes, las dicotomías se muestran como marcos analíticos útiles para abrir una controversia tendiente a relacionar, distinguir y/o comparar posiciones. Luego, este análisis se realiza de distintas maneras: se puede proceder como si las dicotomías ofrecieran categorías en oposición exhaustiva y excluyente –bajo la forma esto/otro–, o planteando entre ambos polos una tensión, un continuo de posiciones posibles en las que uno puede ubicarse –bajo la forma más cerca/lejos de un polo que del otro–. Con cualquiera de estas dos maneras se podría cruzar dicotomías o tensiones, dando lugar a un espacio de posiciones con dos o más ejes. De hecho, esto es lo que proponen Russel y Norvig (2010) en su clásico *Artificial intelligence: a modern approach* al contraponer pensamiento/acciones y humanas/racionales, distinguiendo definiciones de I.A. que discurren por su semejanza al pensamiento humano,

la conducta o acción humana, el pensamiento racional, y la acción racional. Más allá del caso particular, el problema con este tipo de marco analíticos es que se ven muy limitados a controversias en la que predominan muy pocas tensiones o dicotomías. De lo contrario, el espacio de posiciones se vuelve demasiado complejo y se gana poco orden.

Un camino analítico distinto podría ser, en lugar de identificar posiciones entre dicotomías, abstraer la manera en que se problematizan los polos. Así, por ejemplo, frente a la tensión correlación/causalidad no se trataría de ver si un discurso epistemológico está más cerca del primer polo que del segundo, sino de preguntarse si acaso considera que entre ambos elementos se da: (1) un reduccionismo en el que predomine sólo un término, (2) tomar ambos elementos como indisociables y reflexionar sobre cómo integrarlas en un planteo común. Por cierto, es posible una tercera alternativa (3), si es que se pretende restituir el dualismo de los términos, creyendo que son irreductibles. Esta tercera concepción no parece estar presente en la discusión que nos interesa, o al menos, no con una presencia clara en la literatura.

La ventaja comparativa de esta manera de conducir el análisis es que permite abstraerse de la tensión en particular, y desde allí vincular y comparar el tratamiento de otras dicotomías en un espacio multidimensional. En otros trabajos hemos denominado a esta manera de problematizar las tensiones como una “concepción”, y hemos propuesto que algunos programas teóricos de la sociología se pueden caracterizar por tener una concepción generalizada (Giordano & Becerra, *en prensa*). Es cierto que transferir este análisis del terreno de la meta-teoría sociológica al campo de las discusiones epistemológicas no es sencillo, porque aquí la discusión ya no es entre programas teóricos desarrollados por un autor o una tradición –cuya coherencia en la concepción probablemente sea producto de una concepción ontológica de lo social o un marco epistémico– sino que, en las discusiones que nos interesan en este trabajo, no hay interlocutores claros, sino más bien, mitologías, *hypes*, provocaciones por parte de propulsores muy diversos, y espacios discursivos que no necesariamente valoran la no-contradicción –e.g., el sentido común–. Luego, en otro trabajo previo (Becerra & Castorina, 2023) propusimos que en la discusión epistemológica del big data conviven dos marcos epistémicos –una noción análoga a la que aquí tratamos como concepción–: uno tendiente a la escisión y otro tendiente a la relación entre elementos. Esta ponencia se basa mayormente en este último trabajo, pero hemos querido resaltar la presencia de tensiones y concepciones en la literatura, y hacer más explícitos los discursos involucrados: por un lado, uno que sistemáticamente plantea una escisión entre los elementos, y que tienden a suponer un reduccionismo de un polo sobre otro, que identificamos con algunos supuestos y reclamos empiristas; por el otro, la mirada crítica y sociológica del big data, la inteligencia artificial y los algoritmos en general, que tiene su

expresión más clara en los *Critical data studies* (Iliadis & Russo, 2016; Kitchin & Lauriault, 2014; Verständig, 2021).

3. Discusiones epistemológicas

En este apartado mapeamos estas tensiones, dejando para la conclusión una síntesis que explicita las concepciones mencionadas.

La naturaleza del dato y el problema de la objetividad

Uno de los problemas centrales corresponde a cómo entendemos la naturaleza de los datos, una discusión que generalmente pone en cuestionamiento la pretensión de objetividad. Aquí la dicotomía aparece contraponiendo una caracterización de los datos como captura y aprehensión de algunos aspectos de la realidad, a través de un proceso constructivo que les da forma y que condiciona el alcance de su interpretación, frente a otra que entiende que los datos son una representación de la realidad, que pueden “hablar en su nombre”, y que son buenos datos si es que son “datos crudos”. Esta última metáfora corresponde a Geoffrey C. Bowker, quien la contrapone a los “datos cocidos”, retomando la distinción con la que Levi-Strauss revisita el vínculo entre naturaleza/sociedad. El planteo ha sido recogido por varios trabajos muy influyentes en las ciencias sociales, como el de boyd y Crawford (2012), o el de Lisa Gittelman (2013), quien directamente lo utilizó como título de su obra *Raw Data Is an Oxymoron*.

Una de las mejores conceptualizaciones al respecto de esta problemática es la que desarrolla Sabina Leonelli cuya dicotomía distingue entre una mirada representacional y otra relacional del dato:

La visión representacional interpreta los datos como representaciones confiables de la realidad que se producen a través de la interacción entre los humanos y el mundo. Las interacciones que generan datos pueden tener lugar en cualquier entorno social, independientemente de los fines de la investigación. ... Desde el punto de vista relacional, los datos son objetos que se tratan como evidencia potencial o real para afirmaciones científicas de manera que pueden, al menos en principio, ser examinados y contabilizados. El significado asignado a los datos depende de su procedencia, sus características físicas y lo que estas características representan, y las motivaciones e instrumentos utilizados para visualizarlos y defender interpretaciones específicas (Leonelli, 2020)

Distintas pretensiones epistemológicas se fundan en ambas miradas. La visión representacional escinde los momentos de recolección y análisis, en un sentido análogo al que el empirismo lógico había planteado al orientar la distinción hecho/valor de cara al ideal de la objetividad, proponiendo hechos independientes de la subjetividad y de cualquier interpretación teórica. En contraparte, la relacional entiende que la distancia entre estos momentos no puede sino poner en riesgo la validez del uso y la interpretación de los datos, y que una deconstrucción de la historia del *dataset* sería esencial para reconocer los límites a considerar en la interpretación. Luego, la distinción hecho/valor es discutida en la visión relacional: hay presencia de valores (sociales) tanto en la interpretación del dato como en su recolección (construcción), ya que, de hecho, el proceso de recolección, extracción, o apropiación de los datos se realiza –en gran parte– en función de su valor (económico).

El objeto y el objetivo de estudio

Otra cuestión importante corresponde a la delimitación del objeto del estudio, y con ello también al objetivo hacia el que se implementa investigación con big data, o se aplican técnicas algorítmicas y de inteligencia artificial. Una posible tensión que sintetiza este problema se puede observar entre sujeto/conducta, y lo plantea Zuboff (2019) cuando sugiere que el “capitalismo de vigilancia” asume un instrumentalismo que recorta los sujetos sociales a meras conductas para volverlas un *commodity* en un mercado de comportamientos:

La perspectiva instrumentarista se forjó en el dominio intelectual del “conductismo radical”. Gracias a sus capacidades de “Gran otro”, este reduce la experiencia humana a la conducta observable y medible, a la vez que permanece indiferente a los sentidos de la experiencia. Es profundo, infinito y, siguiendo sus orígenes conductuales, radicalmente indiferente a nuestros sentidos y motivos (Zuboff, 2019, p. 20).

Una mirada alternativa buscaría contextualizar los sujetos y las prácticas, y a recuperar el sentido de las acciones, para delimitar un sujeto más complejo. La dicotomía en la unidad de referencia se continúa en el direccionamiento de la investigación: como varios autores señalan, en estos desarrollos –y más aún en la sociología computacional– conviven lógicas provenientes de espacios del saber –e.g., ciencias/ingenierías– con objetivos divergentes, tales como explicar/controlar (Lin, 2015; Wallach, 2018).

Los fundamentos metodológicos y epistemológicos

Tal vez la controversia epistemológica más extendida a partir del impacto del big data y la inteligencia artificial, y del terreno intermedio de problemas que se abren con el aprendizaje automático, corresponda al reclamo de que la construcción de teorías y explicaciones

causales se habría vuelto un objetivo obsoleto. Quienes promueven esta visión sostienen que generar conocimiento inductivamente a partir de correlaciones sería no sólo suficiente, sino incluso más eficiente, ya que, con la abundancia de datos, el proceso es más barato y rápido, y sería más fácil desde allí llegar a desarrollar predicciones estadísticas. Esta posición se encuentra discutida hartamente, y no hay duda del consenso a favor de diseños mixtos que sitúen las oportunidades del big data y la inteligencia artificial en un proceso de investigación que proponga una dialéctica virtuosa entre teoría y dato; y de enmarcar dichos diseños en la reflexividad y una vigilancia epistemológica que alcance a problematizar la relación entre la academia y las corporaciones (Becerra & Castorina, 2023). Kitchin lo expone en estos términos:

Existe una alternativa potencialmente fructífera [...] que adopta y amplía las epistemologías [...]. Estos enfoques emplean técnicas cuantitativas, estadística inferencial, modelado y simulación siendo conscientes y abiertos a sus deficiencias epistemológicas, y recurriendo a la teoría social crítica para enmarcar cómo se lleva a cabo la investigación, cómo se da sentido a los hallazgos y cómo se emplea el conocimiento resultante. [...] Pero se debe] reconocer que la investigación no es neutral y que se enmarca de maneras sutiles y explícitas en aspiraciones e intenciones de sus investigadores y de quienes financian, lo que se traduce de varias maneras en políticas, instrumentos y acciones (Kitchin, 2014, p. 9).

En términos de dicotomías, esta controversia se ha tratado suponiendo los pares dato/teoría y correlación/causalidad. Todos estos elementos confluyen eventualmente en la aspiración final: información/conocimiento (Frické, 2009), cuyo contexto de emergencia parece ser una cierta desvalorización (social) del conocimiento (científico) en tiempos de post-verdad. Análogamente, otros autores problematizan la distinción entre conocimiento experto/general (Hansen & Quinon, 2023).

4. Posibles concepciones y palabras finales

A lo largo de las tensiones mencionadas la literatura señala la presencia de dos discursos. Por un lado, uno que recoge reclamos empiristas, que se destaca en la discusión de las tensiones ligadas a los fundamentos metodológicos es donde esta presencia es más evidente. Leonelli (2020), Kitchin (2014), Frické (2009), Resnyansky (2019), Desai et. al. (2022) lo identifican discutiendo la manera de entender a los modelos, en la tensión correlación/causalidad, y en la disputa entre dato/teoría. También en las tensiones que tratan con la naturaleza del dato: Leonelli (2020) lo ubica detrás de la concepción representacional.

En el mismo problema, es claro que el tratamiento de los dogmas empiristas en la tensión hecho/valor condicionan estos discursos desde un entendimiento escisionista de la objetividad. Luego, en relación a las tensiones vinculadas al objeto y al objetivo, D'Ignazio y Klein (2020) lo ubican en la mirada que no escapa a las condiciones inmediatas de privilegio y no cuesta la matriz del poder en la que se insertan las clasificaciones. Zuboff (2019) parece identificar otro discurso, al referirse al marco epistémico del conductismo radical, pero no sería arriesgado hipotetizar una posición empirista en el recorte de la unidad de análisis a las conductas (visibles).

El segundo discurso es el que ejemplificamos con los *Critical data studies*, cuya concepción tiende a tratar las dualidades como indisociables y a repensarlas en una interacción constitutiva. Así, toman predominancia las nociones que tienden puentes entre dicotomías, como, por ejemplo, la noción de “contexto” en las discusiones en torno a la naturaleza del dato, a la objetividad, y la delimitación del objeto, o la de “reflexividad” que supone la vigilancia de los vínculos entre los elementos. Por mencionar a los autores de nuestro corpus: es contextual la visión relacional de los datos (Leonelli, 2020), la reflexión sobre sus condiciones de elaboración y la evaluación de los impactos de la introducción de la tecnología en ensamblados sociales (Beer, 2017); Kitchin (2014, p. 1), de hecho, explícitamente reclama una “epistemología situada, reflexiva y contextualmente vinculada” para aprovechar la introducción de la tecnología.

Para finalizar quisiéramos hacer una última mención a la forma en que mayormente se ha dado a conocer la discusión epistemológica que nos interesa: la que habla de una revolución del conocimiento a partir del big data y la inteligencia artificial. Al respecto, coincidimos con Resnyanik (2019) en que el discurso de la supuesta revolución se sostiene en una apuesta por escindir y discontinuar elementos que las ciencias sociales hace décadas que problematizan, y para los que se han ensayado diversos caminos de abordaje que no recaen en reduccionismos, o al menos, no parece ser esa la concepción dominante.

Referencias

- Alvarado, R. (2023). AI as an Epistemic Technology. *Science and Engineering Ethics*, 29(5), 32. <https://doi.org/10.1007/s11948-023-00451-3>
- Becerra, G., & Castorina, J. A. (2023). Hacia un análisis de los marcos epistémicos del big data. *Cinta de Moebio*, 76, 50–63. <https://doi.org/10.4067/S0717-554X2023000100050>
- Beer, D. (2017). The social power of algorithms. *Information Communication and Society*,

- 20(1), 1–13. <https://doi.org/10.1080/1369118X.2016.1216147>
- boyd, D., & Crawford, K. (2012). Critical Questions for Big Data. *Information, Communication & Society*, 15(5), 662–679.
<https://doi.org/https://doi.org/10.1080/1369118X.2012.678878>
- D'Ignazio, C., & Klein, L. (2020). *Data feminism*. The MIT Press.
- Desai, J., Watson, D., Wang, V., Taddeo, M., & Floridi, L. (2022). The epistemological foundations of data science: a critical review. *Synthese*, 200(6), 1–27.
<https://doi.org/10.1007/s11229-022-03933-2>
- Frické, M. (2009). The knowledge pyramid: A critique of the DIKW hierarchy. *Journal of Information Science*, 35(2), 131–142. <https://doi.org/10.1177/0165551508094050>
- Giordano, P., & Becerra, G. (n.d.). Integración y dualismo: concepciones generales para tratar tensiones en las teorías de Pierre Bourdieu y Niklas Luhmann. *Revista Mexicana de Ciencias Políticas y Sociales*.
- Gitelman, L. (2013). *“Raw Data” Is an Oxymoron*. The MIT Press.
<https://doi.org/10.1080/1369118X.2014.920042>
- Hansen, J. U., & Quinon, P. (2023). The importance of expert knowledge in big data and machine learning. *Synthese*, 201(2), 1–21. <https://doi.org/10.1007/s11229-023-04041-5>
- Iliadis, A., & Russo, F. (2016). Critical data studies: An introduction. *Big Data & Society*, 3(2), 205395171667423. <https://doi.org/10.1177/2053951716674238>
- Kitchin, R. (2014). Big Data, new epistemologies and paradigm shifts. *Big Data & Society*, 1(1). <http://journals.sagepub.com/doi/10.1177/2053951714528481>
- Kitchin, R., & Lauriault, T. P. (2014). Towards critical data studies : Charting and unpacking data assemblages and their work. *Geoweb and Big Data*, 1–19.
http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2474112
- Leonelli, S. (2020). Scientific Research and Big Data. *Stanford Encyclopedia of Philosophy*.
- Lin, J. (2015). On Building Better Mousetraps and Understanding the Human Condition: Reflections on Big Data in the Social Sciences. *Annals of the American Academy of Political and Social Science*, 659(1), 33–47. <https://doi.org/10.1177/0002716215569174>
- Resnyansky, L. (2019). Conceptual frameworks for social and cultural Big Data analytics: Answering the epistemological challenge. *Big Data & Society*, 6(1), 205395171882381.

<https://doi.org/10.1177/2053951718823815>

Russell, S., & Norvig, P. (2010). *Artificial intelligence: a modern approach*. Pearson Education.

Verständig, D. (2021). Critical Data Studies and Data Science in Higher Education. *Seminar.Net*, 17(2). <https://doi.org/10.7577/seminar.4397>

Wagner-Pacifici, R., Mohr, J. W., & Breiger, R. L. (2015). Ontologies, methodologies, and new uses of Big Data in the social and cultural sciences. *Big Data & Society*, 2(2), 205395171561381. <https://doi.org/10.1177/2053951715613810>

Wallach, H. (2018). Viewpoint: Computational social science ? computer science + social data. *Communications of the ACM*, 61(3), 42–44. <https://doi.org/10.1145/3132698>

Zuboff, S. (2019). Surveillance Capitalism and the Challenge of Collective Action. *New Labor Forum*, 28(1), 10–29. <https://doi.org/10.1177/1095796018819461>