Daimon. Revista Internacional de Filosofía, nº 90 (2023), pp. 7-12
ISSN: 1130-0507 (papel) y 1989-4651 (electrónico) http://dx.doi.org/10.6018/daimon.576781

Presentación de la sección sobre Inteligencia artificial, datos y objetividad. ¿El regreso del naturalismo dataísta?

Presentation of the section on Artificial Intelligence, Data, and Objectivity. The return of dataist naturalism?

ARIEL GUERSENZVAIG[*]
DAVID CASACUBERTA[**]

Es función de la filosofía analizar los problemas desde una perspectiva amplia, lo más genérica posible y transdisciplinar, versus a un acercamiento más detallado y específico de las ciencias experimentales. Sin embargo, esa generalidad puede hacernos perder de vista problemas importantes que se mueven en marcos más específicos y, al llevarse a un extremo, acabar produciendo un discurso filosófico vago y sin concreción.

Encontramos actualmente esta situación en buena parte del discurso divulgativo filosófico que se está generando en relación al impacto de la inteligencia artificial (IA) en la sociedad. En este, no se distingue entre las diferentes técnicas y metodologías usadas para desarrollar sistemas computacionales con IA, los diferentes campos de aplicación de esas tecnologías, ni los diferentes tipos de problemas éticos que esas tecnologías pueden presentar. Sin precisar los conceptos clave (ni siquiera la propia noción de «inteligencia» o qué significa «pensar»), se explora, por poner un ejemplo, la complejísima cuestión de si los

* Elisava Facultad de Diseño e Ingeniería de Barcelona, UVIC-UCC <aguersenzvaig@elisava.net>, Profesor Contratado Doctor. Sus principales áreas de investigación son, por un lado, el impacto ético de la inteligencia artificial en la sociedad y, por otro lado, la ética de la actividad profesional del diseño. Pertenece al Grupo de Investigación consolidado HIMTS (Human, Interaction, Materials, Technology, and Society). Es miembro del comité de ética de la investigación de la Universidad de Vic-UCC. Publicaciones recientes: Guersenzvaig, A., & Sangüesa, R. (2022). A critical reflection on the treatment of AI system's 'agency' in the (Spanish) media. *Avances en Interacción Humano-Computadora*, *1*(7), 1-4: Guersenzvaig, A. (2021). *The Goods of Design: Professional Ethics for Designers*. Rowman & Littlefield.

** Universidad Autónoma de Barcelona <david.casacuberta@uab.cat>. Profesor Contratado Doctor en el Departamento de Filosofía de la Universidad Autónoma de Barcelona. Su línea de investigación actual son los impactos sociales y cognitivos de las tecnologías digitales. Actualmente es miembro del Grupo de Trabajo de Ética, Seguridad y Regulación de bioinformática Barcelona, investigador del grupo consolidado GEHUCT (Grupo de Estudios Humanísticos en Ciencia y Tecnología). Publicaciones recientes: Casacuberta, D., Guersenzvaig, A., & Moyano-Fernández, C. (2022). Justificatory explanations in machine learning: for increased transparency through documenting how key concepts drive and underpin design and engineering decisions. *Ai & Society*, 1-15; Casacuberta, D., & Guersenzvaig, A. (2019). Using Dreyfus' legacy to understand justice in algorithm-based processes. AI & SOCIETY, 34(2), 313-319.

sistemas con IA «piensan mejor que nosotros». Se genera así un discurso que resulta ser, o bien de corte apocalíptico (como en el caso de los riesgos existenciales de la IA), o bien de un utilitarismo banal que no va más allá de la evaluación superficial de pros y contras (como cuando se reduce el debate en torno a la IA a la examinación de sus beneficios y perjuicios más directos y, de ser posible, de manera matemática).

Sea como fuere, en estos casos la discusión se centra en pseudoproblemas que no tienen ningún impacto en la actualidad y se dejan de lado problemas reales que necesitan *ya* de soluciones y respuestas filosóficas, quedando totalmente desapercibidos. Esta superficialidad la encontramos, por ejemplo, en la discusión filosófica en torno a los vehículos autónomos. La banalidad se hace patente cuando, de manera espuria, se utiliza como principal instrumento de exploración el experimento mental del *dilema del tranvía*, genialmente formulado por Philippa Foot (1967) en el marco de una reflexión mucho más amplia sobre la doctrina del Doble efecto y, más específicamente, sobre la diferencia entre actuar y dejar que algo ocurra. En el caso de los vehículos autónomos, tal como sucede en el famoso *Moral Machine Experiment* (Awad et al., 2018), se plantean escenarios dicotómicos y a veces implausibles (e.g., «¿El coche autónomo debe embestir y matar a un bebé, o a una persona sin hogar?»). Estos escenarios, en vez de permitir y propiciar una reflexión ética profunda como la desarrollada por Foot, se consideran literalmente y como destino filosófico final. Se eluden así cuestiones más relevantes, y también filosóficamente más ricas, como, por ejemplo, qué tipo de movilidad requiere una sociedad moderna y plural, cuál es la responsabilidad de los fabricantes de vehículos, qué rol deben tener el estado y el gobierno, cómo se navegan las tensiones entre los derechos y libertades de los automovilistas y los de otros usuarios de la vía pública, qué grupos se ven principalmente beneficiados o perjudicados por el uso de automóvil, o, de manera aún más abarcadora, qué modelos de ciudad son más conducentes al bienestar.

En paralelo, este uso espurio del *dilema del tranvía* demuestra una falta completa de imaginación y comprensión filosófica, interpretando de manera literal lo que en realidad es un experimento mental cuyo fin es mostrar la interacción y posible inconsistencia de algunas de nuestras intuiciones éticas, así como la centralidad de la intención de los agentes en la consideración de los daños que puedan ocasionar. Otra ingenuidad filosófica muy común en los escritos dataístas es cuando se habla de dejar la investigación científica o la discusión ética en manos de algoritmos.

En este número especial hemos intentado hacer una contribución para enmendar esta confusión y ayudar a elaborar un discurso filosófico más profundo y específico que trate problemas concretos en lugar de ofrecer un falso discurso totalizador que sea, en realidad, superficial y ambiguo. Por lo que hace a esta sección, pensamos que desde la filosofía no se ha hecho suficiente hincapié en la insostenibilidad de los presupuestos de la filosofía de la ciencia dataísta que cree que se acerca un fin del método científico, en el que algoritmos de aprendizaje automático, en base a grandes volúmenes de datos, podrán hacer predicciones útiles en todos los campos de la investigación científica, y en especial en aquellas ciencias aplicadas para el bienestar humano, como la biomedicina, sin necesidad de tener que construir hipótesis y teorías, produciendo así un conocimiento verdaderamente objetivo. En otras palabras, una ciencia *naturalista*, puramente determinada por hechos empíricos, y libre de valores y teorías previas.

Específicamente, entonces, queremos cuestionar el aparente naturalismo dataísta de estos sistemas algorítmicos y tratar su supuesta neutralidad matemática, presentada como garantía

de un recorrido riguroso e impersonal desde los datos al resultado. El enfoque cuantitativo y estadístico suele asociarse a la capacidad de ofrecer credibilidad y confianza. Durante los últimos dos siglos, las mediciones y análisis estadísticos que afectan todas las áreas de la vida pública y privada han permitido la creación y revisión de teorías y han vertebrado el debate público (Desrosières, 1993; Porter, 2020). Una frase como "esto está respaldado por datos" se convirtió en una frase común para legitimar afirmaciones y decisiones sobre pobreza, educación, empleo y prácticamente cualquier otro aspecto de la vida social. Esta mentalidad, conocida como «dataísmo» (Brooks, 2013), se vio acentuada por la amplia disponibilidad de las computadoras y las bases de datos, así como el surgimiento del *Big Data*, la utilización de grandes volúmenes de datos con fines computacionales.

Un influyente ensayo escrito por Chris Anderson *The End of Theory: The Data Deluge Makes the Scientific Method Obsolete* (Anderson, 2008) epitomiza la visión del dataísmo radical como ningún otro. Anderson argumenta que la disponibilidad de grandes volúmenes de datos (el llamado *Big Data*) y la inteligencia artificial harán que las teorías y las hipótesis *ex-ante* sean redundantes. Esta visión subvierte así los principios generales de la filosofía de la ciencia posterior a 1950 en la que el itinerario del descubrimiento comienza con la formulación de hipótesis tentativas (impulsadas por conjeturas fundamentadas en teorías y datos previos) que se validan empíricamente, generando así nuevos datos que sirven para aceptarlas, modificarlas o rechazarlas.

Con la siempre creciente adopción de sistemas computacionales para la automatización de todo tipo de ámbitos públicos y privados, los datos y los cálculos estadísticos van camino de consolidarse como el estándar de facto del conocimiento aplicado. A pesar de los serios problemas que genera su uso, mediante sistemas de IA, por ejemplo, se otorgan préstamos y créditos (Eticas Foundation, 2021), se gestionan presupuestos gubernamentales (Valle-Cruz et al., 2022), se decide quién debe ser investigado por posible fraude en la percepción de prestaciones por cuidado de menores (Hadwick & Lan, 2021), se gestiona y evalúan empleados (Tewari & Pant, 2020), se vigilan las fronteras (Sánchez-Monedero & Dencik, 2022), o se toman decisiones jurídico-penales (Casacuberta y Guersenzvaig, 2019; Morales Moreno, 2021). No resulta entonces exagerado sugerir que con la estadística evolucionando hacia la «ciencia de datos», la visión dataísta que domina buena parte del desarrollo tecnológico actual comienza, en la práctica, a gobernar el mundo.

La inteligencia artificial originalmente estuvo fundamentalmente vinculada a teorías cognitivistas y a un enfoque «simbólico». Sin embargo, la principal técnica de la inteligencia artificial actual es el llamado «aprendizaje automático» *(machine learning* en inglés*)*, que está basado de manera general en representaciones de conocimiento obtenidas mediante técnicas matemáticas y estadística aplicada en conjunción con el procesamiento computacional de enormes volúmenes de datos. El vínculo entre la IA y el enfoque simbólico sigue existiendo pero este se ha debilitado mucho debido a la preponderancia del aprendizaje automático que se enmarca dentro del enfoque denominado «conexionista».

Recientemente, la última evolución de los sistemas de este tipo como los generadores de texto e imágenes como ChatGPT o Stable Difussion (técnicamente basados en un modelo de aprendizaje profundo llamado «transformador») han cosechado amplia atención. También encontramos multitud de otros sistemas orientados a realizar predicciones, evaluaciones y clasificaciones en base a *Big Data*. Así, por ejemplo, mediante la utilización de imágenes

dermatológicas, un sistema de IA «aprende» a detectar melanomas procesando ingentes cantidades de fotos y detectando patrones y correlaciones estadísticas sin necesidad de modelos conceptuales o teóricos previos acerca de *qué es* un melanoma. Manteniéndonos en el ámbito de la salud y por ilustrar con otro ejemplo, estos grandes volúmenes de datos sirven también para crear «simulacros digitales», es decir modelos computacionales que sirven como representación de personas o grupos de personas (y también animales no humanos o plantas). Estos «gemelos digitales» pueden servir para generar predicciones de la evolución de una enfermedad o de la aplicación de un medicamento a lo largo del tiempo. Según informe*s recient*es, las pruebas *in silico* ya comienzan a reemplazar a algunas pruebas tradicionales en laboratorio (Moingeon et al., 2023).

Los métodos cuantitativos y estadísticos para investigaciones científicas no son nada nuevo. A partir del siglo XIX, basándose en el razonamiento inductivo promovido por filósofos como Bacon y los éxitos empíricos de Kepler, Newton y otros científicos naturales durante el siglo XVII, un grupo de pensadores ejecutó una verdadera revolución epistémica al considerar los patrones estadísticos como intrínsecamente explicativos (Hacking, 1990; Porter, 2020). Figuras como Quetelet y Galton establecieron las mediciones cuantitativas y el razonamiento estadístico como un modo legítimo de investigación incluso en las ciencias sociales. Esto se fortaleció con la aparición de la ciencia positiva, matematizada, que adoptó estos métodos como instrumentos tanto para la generación de conocimiento como para su demostración. Dichos métodos ubicuos y sus premisas de rigurosidad, neutralidad de valores y objetividad han sido frecuente y duramente criticados por varios autores (e.g., Hacking, 1990; Desrosières, 1993; Lewontin, 1993; Porter, 2020).

En línea con lo que comentábamos sobre los gemelos digitales en biomedicina, Cho et al., (2022, p.1) plantean que «los simulacros digitales marcan un hito importante en la trayectoria para abrazar la cultura epistémica de la ciencia de datos y un potencial abandono de los conceptos epistemológicos médicos de causalidad y representación». Vale la pena insistir que la perspectiva dataísta del *fin de la teoría* no es un corpus cohesionado de fuentes académicas o un marco teórico en un sentido estricto; más bien, es una mentalidad, e incluso una ideología (Blakely, 2020), que busca permear la generación de conocimiento, y sus implementaciones prácticas, en prácticamente todos los campos de la actividad humana.

¿Hasta qué punto la visión dataísta de la ciencia se ajusta a la realidad? ¿De qué forma esa visión oculta y distorsiona problemas epistémicos y éticos muy relevantes? Estas preguntas y otras similares son algunas de las que queremos explorar con esta selección de artículos. Así, esta Sección 1ª, *Inteligencia artificial, datos y objetividad. ¿El regreso del naturalismo Dataísta?* arranca con *¿Son las computadoras agentes inteligentes capaces de conocimiento?* de Gustavo Esparza y Daniel Martínez. El artículo explora los fundamentos epistémicos del dataismo analizando las posibilidades y límites de los algoritmos generados por aprendizaje automático a la hora de establecer conocimiento científico. Partiendo de un análisis de los fundamentos filosóficos de la arquitectura de programación en dos sistemas de Inteligencia Artificial (AlphaGo y Hide and Seek) se postula que tales algoritmos pueden llegar a ser recursos de conocimiento especiales para la comprobación de hipótesis, estableciendo así cómo los algoritmos de aprendizaje automático pueden realmente producir innovaciones epistémicas.

El resto de artículos se centra en tratamientos críticos de cuestiones asociadas al aprendizaje automático, en particular al mantenimiento y proliferación de sesgos epistémicos y éticos.

El artículo de Cristian Moyano, *La IA usada en biología de la conservación es una buena estrategia de justicia ambiental?,* aborda la cuestión de los sesgos y la analiza en el contexto específico del conservacionismo, aportando así un enfoque distintivo acerca de un tema ampliamente comentado en las investigaciones éticas acerca de la IA en la última década y en este mismo número. En dicho artículo se reconocen las posibilidades que ofrece la IA en el campo de la conservación de especies biológicas, pero apunta también a los problemas reales que un uso no controlado de estos algoritmos podría provocar al expandir sesgos epistémicos y éticos ya presentes en la actualidad en procesos de conservación. El artículo muestra así la importancia de ir caso por caso y analizar en cada disciplina específica que tipos de sesgos epistémicos y éticos son relevantes y cómo afrontarlos.

El tema de los sesgos se explora también en *Discurso influenciado: aprendizaje automático y discurso de odio* de Federico Javier Jaimes. El artículo analiza cómo la propagación sistémica de sesgos vía algoritmos de aprendizaje automático puede, a partir del concepto de «discurso influenciado», explicar la reproducción social de los discursos de odio al enmarcar teóricamente las formas en que algoritmos de aprendizaje automático expanden y normalizan discursos de odio.

Cerrando una brecha: una reflexión multidisciplinar sobre la discriminación algorítmica, de Pilar Dellunde, Oriol Pujol y Jordi Vitrià, plantea un acercamiento sistemático al concepto de discriminación algorítmica, sin duda uno de los principales y más relevantes problemas epistémicos y éticos a la hora de considerar la aplicación de algoritmos de aprendizaje automático en la esfera humana. El artículo plantea la necesidad de entender un algoritmo como una tecnología intencional y por tanto abierta a sesgos, imposibilitando así esa supuesta metodología radicalmente objetiva y libre de teorizaciones que postula el dataísmo.

Finalmente, *Más allá de los datos: la transformación digital del museo tradicional,* de Alger Sans y Vicent Costa investiga asimismo el problema de los sesgos y explora filosóficamente cómo la introducción de tecnologías digitales e inteligencia artificial en el museo tradicional podría transformar los procesos educativos. Los autores insisten en la insuficiencia de un acercamiento dataísta a la hora de detectar sesgos epistémicos y éticos, los cuales podrían llevar a una situación de injusticia epistémica en los museos, y a exacerbar así las discriminaciones y exclusiones ya existentes en los museos tradicionales

Referencias

Awad, E., Dsouza, S., Kim, R., Schulz, J., Henrich, J., Shariff, A., Bonnefon, J.-F., & Rahwan, I. (2018). The Moral Machine experiment. *Nature*, 563(7729), 59-64. https://doi.org/10.1038/s41586-018-0637-6

Blakeley, J. (2020). *We built reality: How social science infiltrated culture, politics, and power.* Oxford University Press.

Brooks, D. (2013, Feb 4) *The Philosophy of Data*. The New York Times. Consultado 01/07/2023 desde https://www.nytimes.com/2013/02/05/opinion/brooks-the-philosophy-of-data.html

Cho, M. K., & Martinez-Martin, N. (2022). Epistemic Rights and Responsibilities of Digital Simulacra for Biomedicine. *The American journal of bioethics : AJOB*, 1–12. Advance online publication. https://doi.org/10.1080/15265161.2022.2146785

Desrosières, A. (1993). *La politique des grands nombres: Historie de la raison statistique*. Éditions La Découverte.

Eticas Foundation. (2021) *Sesgo de calificación crediticia y reproducción de desigualdad en préstamos para vivienda*. Consultado 01/07/2023 https://eticasfoundation.org/es/sesgo-de-calificacion-crediticia-y-reproduccion-de-desigualdad-en-prestamos-para-vivienda/

Foot, P. (1967). The Problem of Abortion and the Doctrine of Double Effect. *Oxford Review*, 5, 5-15.

Casacuberta, D., Guersenzvaig, A. Using Dreyfus' legacy to understand justice in algorithm-based processes. *AI & Soc* 34, 313–319 (2019). https://doi.org/10.1007/s00146-018-0803-2

Hacking, I. (1990). *The Taming of Chance*. Oxford University Press.

Hadwick, David & Lan, Shimeng. (2021) Lessons to Be Learned from the Dutch Childcare Allowance Scandal: A Comparative Review of Algorithmic Governance by Tax Administrations in the Netherlands, France and Germany. *World tax Journal*. 13(4), 609-645.

Lewontin, R. (1993). *Biology as Ideology: The Doctrine of DNA*. Harper Perennial.

Morales Moreno, A. M. (2021). Algoritmos en el estrado, ¿realmente los aceptamos? Percepciones del uso de la inteligencia artificial en la toma de decisiones jurídico-penales. *Ius et Scientia*, 7 (2), 57-87. https://doi.org/10.12795/IETSCIENTIA.2021.i02.05

Moingeon, P., Chenel, M., Rousseau, C., Voisin, E., & Guedj, M. (2023). Virtual patients, digital twins and causal disease models: Paving the ground for in silico clinical trials. *Drug discovery today*, 28(7), 103605. https://doi.org/10.1016/j.drudis.2023.103605

Porter, T. (2020). *Trust in Numbers: The Pursuit of Objectivity in Science and Public Life,* new edition. Princeton University Press.

Sánchez-Monedero, J., & Dencik, L. (2022). The politics of deceptive borders: 'biomarkers of deceit' and the case of iBorderCtrl. *Information, Communication & Society,* 25(3), 413-430. https://doi.org/10.1080/1369118X.2020.1792530

Tewari, I., & Pant, M. (2020). Artificial Intelligence Reshaping Human Resource Management : A Review. *2020 IEEE International Conference on Advent Trends in Multidisciplinary Research and Innovation* (ICATMRI), Buldhana, India. https://doi.org/10.1109/ICATMRI51801.2020.9398420

Valle-Cruz, D., Fernandez-Cortez, V., & Gil-Garcia, J. R. (2022). From E-budgeting to smart budgeting: Exploring the potential of artificial intelligence in government decision-making for resource allocation. *Government Information Quarterly*, 39(2), 101644. https://doi.org/https://doi.org/10.1016/j.giq.2021.101644