

# Del cliché de la revolución en inteligencia artificial a la incertidumbre de la revolución social

German Sánchez-Torres

Director del grupo de Investigación y Desarrollo en Sistemas y Computación, Facultad de ingeniería, Universidad del Magdalena, Carrera 32 No. 22-08, Santa Marta, Colombia. (correo: gsanchez@unimagdalena.edu.co.)

El inconveniente con la retórica del cliché es que se cimienta en una lógica de creencias que establece que la creencia sobre la realidad es en sí misma la realidad. Con el surgimiento masivo de medios de comunicación, la escasa presencia de rigurosidad científica en éstos y la insipiente importancia de la ética, la percepción generada sobre diferentes aspectos de la realidad, frecuentemente, corre el peligro de caer en el espectro del cliché. Lo significativo de resaltar, para mal, es que con el beneplácito de algunos que hacen ciencia, se termina implantado este tipo de retórica en aspectos propios de la ciencia.

En el campo en particular de la inteligencia artificial, podríamos afirmar que la última década y media, ha sido para la comunidad científica del área, la de mayor desarrollo en términos de los logros alcanzados. Con soluciones que, incluso, los grandes protagonistas no esperaban obtener. Tareas complejas hoy son resueltas con niveles de precisión que superan a las estimaciones con las que el humano las desarrolla. Estos logros no solo están justificados por la proposición de modelos computacionales más sofisticados y robustos estadísticamente, ni necesariamente por el avance normal de las ciencias mismas o fase de Ciencia Normal según T. Kuhn, sino también por el interés de grandes compañías tecnológicas que han puesto a disposición recursos significativos para crear estos desarrollos, concretando la fase de Revolución Científica, según el mismo T. Kuhn (Kuhn, T., 1971).

Si tuviéramos que caracterizar la naturaleza de dichos avances, tendríamos necesariamente que decir que aún predomina el enfoque clásico derivado de los postulados iniciales de A. Turing (Turing, A. M.: 1948)(Turing, A., 1950) basados en imitación y materializados por su prueba clásica que, por supuesto, determina la existencia de inteligencia si se logra una imitación de tal forma que sea indistinguible para un ser humano. En esta caracterización, la

imitación, también se enmarcan todo el enfoque bioinspirado del área, sus diversas técnicas que incluyen desde el Perceptron (McCulloch, W.S., Pitts, 1943) hasta el aprendizaje profundo (Smolensky P., 1986)(Jordan M., 1986)(LeCun Y. *et al.*, 2015), la herramienta más significativa del área hasta estos días. Las críticas a este enfoque se centran en que, si bien ha permitido el desarrollo de soluciones prácticas a problemas complejos, no ha permitido entender la naturaleza de la solución del problema. Como ejemplo, en el caso del reconocimiento de dígitos escritos a mano, las herramientas modernas muestran niveles de precisión superiores al humano. Sin embargo, no es posible extraer en forma de conocimiento, la esencia de la solución. Los invariantes que definen inequívocamente a cada dígito, de tal forma que podamos apropiarlos para incluirlos como heurísticas en modelos de solución de naturaleza diferente. No se ha avanzado tanto en la comprensión de los fenómenos, como sí en la construcción de soluciones orientadas a una habilidad particular.

Pese a los sorprendentes de sus resultados, estas soluciones aún presentan debilidades significativas. Las soluciones logradas son rígidas puesto que no es posible la fácil adaptación a nuevas tareas. Son dependientes de grandes cantidades de datos que incorporan sesgos y suposiciones que el sistema mismo no es capaz de rebatir o despojarse, pero, sobre todo, son carentes de comprensión, la característica más representativa de la inteligencia humana. El enfoque de A. Turing, por demás brillante para su época, indirectamente sugirió el enfoque de la subjetividad de la inteligencia, en el sentido que la existencia y la cantidad de ésta en sistemas artificiales está determinada por el sujeto que evalúa la imitación, el que es engañado. Aunque las ciencias sociales y la medicina han trabajado significativamente en este campo (Spelke y Kinzler, 2007)(McGrew, 2005), las definiciones de inteligencia y los mecanismos para su cuantificación aún no están establecidos de forma tal que permita su fácil incorporación en los postulados de la teoría de aprendizaje automatizado, o por lo menos, no en forma de un mecanismo de aprendizaje generalizado, no enfocado a una habilidad en particular como las soluciones actuales.

Las teorías de aprendizaje automatizado basadas en inferencia estadística presentan como su mayor debilidad que no han podido ser despojadas del costo que implica el influjo impuesto por la ley de los números grandes (Baum y Katz, 1965) y el enfoque de cotas tradicionalmente utilizados, como las cotas de Chernoff y Hoeffding, que sugieren como mecanismo de regularización para obtener generalización, la dependencia a grandes cantidades de datos. El aprendizaje profundo no ha logrado despojarse de esta debilidad. Algunos estudios (Neyshabur, *et al.*, 2017) sugieren que la capacidad de generalización no depende tanto de una correcta parametrización como sí de la cantidad y naturaleza de los datos.

En contraste, quizás como respuesta a las limitaciones descritas de las técnicas recientes, otros mecanismos para regularizar funciones orientadas a la generalización -si el conocimiento o la inteligencia pudieran ser expresados con un conjunto de funciones- es la comprensión del

problema. En esta dirección los aportes de A. Vapnik se distancian del enfoque clásico centrándose en la comprensión de lo que es la inteligencia no en mecanismos que la imiten. Esta concepción de generalización, condensada en su Teoría de Aprendizaje Estadístico (Vapnik, 1995), busca que los sistemas artificiales desarrollen la habilidad de afrontar tareas para las que no fueron explícitamente diseñados. La generalización en sistemas artificiales, o el espectro en el que ésta se ha definido, es la clave para construir sistemas flexibles capaces de resolver tareas infiriendo similitudes con aquellas para las que fueron entrenados, lo que aprendieron. Sin embargo, este enfoque no ha permitido, aún, construir soluciones más eficientes o útiles que las desarrolladas por el enfoque basado en imitación.

Una perspectiva diferente de analizar la evolución de la IA es precisamente analizar la capacidad de generalización de las técnicas. Se inició con sistemas con capacidad de generalización casi inexistente, como los sistemas expertos y, en general, las técnicas que pertenecen al grupo de la Inteligencia Artificial Simbólica, pasando a sistemas más robustos con capacidad de generalización local con habilidades para tareas específicas, como las técnicas de aprendizaje de máquina, el presente. El siguiente paso será un esquema que permita proponer técnicas con una generalización amplia con la cual el sistema sea capaz de extender su habilidad específica a tareas desconocidas para él, pero dentro del mismo dominio. Al final, el nivel ideal que concretaría el objetivo primario de la inteligencia artificial sería algo como una generalización masiva que incluya capacidades cognitivas que permitan aprender automáticamente habilidades para diversas tareas desconocidas, así como tareas en dominios desconocidos para el sistema.

Aun estando lejos de los niveles humanos de inferencia y comprensión, ya la inteligencia artificial ha empezado a transformar muchas de las actividades humanas. Como toda revolución, esta también tendrá efectos económicos y sociales. Sin embargo, la naturaleza de esta revolución magnificará sus efectos por la transversalidad de la tecnología en casi la totalidad de las dimensiones sociales. La automatización inteligente será, sino lo es ya, la primera gran ola de magnitudes significativas. En lo económico, las estimaciones más aceptadas indican pérdidas de millones de empleos que serán sustituidos por mecanismos automatizados inteligentes. Sin embargo, poca relevancia tiene en los titulares el hecho que esos mismos estudios sugieren que se crearán nuevos empleos en cantidades superiores. En realidad, la implicación que hay es reemplazar, en la dirección de especializar, la naturaleza de los puestos de trabajo.

Pero se debe reconocer que no es del todo claro el efecto respecto de lo laboral. Las tendencias actuales en robótica o automatización no nos permiten inferir que algunos trabajos no especializados como albañilería, jardinería, plomería o similares, se vean amenazados con la incursión de esta tecnología. Sí que los trabajos que impliquen actividades repetitivas en

entornos masivos de producción serán susceptibles de ser intervenidos, o por lo menos en la mayoría de estos escenarios. Luego, hacer vaticinios al respecto podría ser solo especulación.

Lo anterior incorpora incertidumbre en lo social, más aún cuando la naturaleza de lo laboral podría generar brechas significativas en aquellos trabajos, que podrían no estar en el foco de interés de las aplicaciones inteligentes, y aquellas que evidentemente serán reemplazadas, generando una mayor especialidad en lo intelectual. Adicionalmente, el hecho de que aún para gran parte del planeta la formación intelectual especializada es todavía un gran reto incrementa la incertidumbre en la cuantificación del efecto. Sin embargo, para la fuerza laboral de mayor afectación, el impacto también estará determinado por las políticas que garanticen acceso a esa preparación. Por supuesto, esto mientras los niveles de generalización en las técnicas aún no sean masivos. El tiempo que tome alcanzar estos niveles de generalización magnifica la incertidumbre de los efectos sociales.

Por fuera del espectro del cliché de los titulares de prensa relativos a la inteligencia artificial sobre escenarios apocalípticos y la existencia pronta de máquinas pensantes, podemos decir que no hay duda, que el futuro de la sociedad está, cuanto menos sesgado, cuanto más, definido por los impactos que tendrá la ubicuidad de la inteligencia artificial. Pero el pobre análisis sobre el estado actual, los retos y el tiempo para lograr dichos retos están aun fuertemente permeados por el cliché y modelados por la incertidumbre.

## Referencias

- Baum L. E. and Katz M., “Convergence rates in the law of large numbers,” *Trans. Am. Math. Soc.*, vol. 120, no. 1, pp. 108–123, Oct. 1965.
- Jordan M. I., *Serial order: A parallel distributed processing approach*. *Advances in psychology*, 121:471–495, 1986
- Kuhn, T. (1971). *La Estructura de las Revoluciones Científicas*, México, Fondo de Cultura Económica.
- LeCun Y., Bengio Y., and Hinton G. Deep learning. *Nature*, 521(7553): 436–444, 2015.
- McCulloch, W.S., Pitts, W. A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics* 5, 115–133 (1943). <https://doi.org/10.1007/BF02478259>
- McGrew K., *The cattell-horn-carroll theory of cognitive abilities: Past, present, and future*. *Contemporary Intellectual Assessment: Theories, Tests, and Issues*, 01 2005.
- Neyshabur, B., Bhojanapalli S., McAllester D., and Srebro N.. 2017. Exploring generalization in deep learning. In *Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS'17)*. Curran Associates Inc., Red Hook, NY, USA, 5949–5958.
- Smolensky P., *Information processing in dynamical systems: Foundations of harmony theory*. Technical report, DTIC Document, 1986

Spelke E. and Kinzler K., Core knowledge. *Developmental science*, pages 89–96, 2007

Turing, A. M.: 1948, 'Intelligent Machinery', in D. Ince (ed.) *Mechanical Intelligence*. North-Holland, Amsterdam, pp. 87–106

Turing, A., 1950, "Computing machinery and intelligence," *Mind* 59, 433–460.

Vapnik V., *The Nature of Statistical Learning Theory*. Springer-Verlag, Berlin, Heidelberg, 1995.