

¿Dónde queda la verdad en la IA?

Recientemente, hemos constatado cómo la IA proporciona a sus usuarios una serie de experiencias y soluciones que cada día son más y más refinadas. Este auge parece indicar que muy pronto emplearemos la IA como herramienta para identificar y disolver problemas de nuestra cotidianeidad.

A priori, podríamos convenir que el límite de “realizabilidad máxima” de la IA fuerte consistirá en igualar la solidez propia de una conciencia plenamente desarrollada –en esta nueva tecnología, según la lógica matemática, el conocimiento de los realizadores nos informa sobre la idoneidad de la formulación algorítmica empleada-. Por tanto, dicho “límite máximo” está relacionado, de manera directa, con la percepción recibida por un usuario externo. Se alcanzará cuando dicho usuario admita la imposibilidad de distinguir entre los impulsos e información que recibirá en su interrelación con la IA y los estímulos que habitualmente le llegan de la interrelación con cualquier otro ser humano. Este proceso de disolución de las barreras experienciales del usuario, esto es, la posible implementación de técnicas de IA que supongan la minoración de las habilidades cognoscitivas que nos permite distinguir entre lo virtual y lo factual, nos ha llevado a preguntarnos: ¿qué papel otorga la IA a la verdad y la causalidad?

Respecto a la verdad

Un análisis lógico-filosófico debe servir, al menos, para alertarnos sobre la tentación de preguntar, de manera ingenua, dónde queda la cualidad de ser verdadero en las respuestas que nos llegan desde la IA. Más bien, nos debería conducir a entender que la pregunta con la que iniciamos el estudio de la verdad sería más próxima a: ¿qué es aquello que confiere validez a las proposiciones emitidas dentro de un sistema de IA?

La primera cuestión a despejar al iniciar el análisis lógico de dicho sistema artificial será conocer cuáles son los criterios que nos indican el modo en que se validan o invalidan ciertas proposiciones. Es importante percatarnos de que la “verdad” o “falsedad” de ciertas proposiciones tratadas por sistemas de IA no tienen por qué aportar algún tipo de vinculación positiva o negativa respecto de la realidad fáctica. Por tanto, podemos evolucionar respecto a la pregunta inicial y considerar que la verdad de la proposición emitida dentro de un sistema de IA no consistirá en su posible relación con hechos u observaciones de la realidad, sino que está diseñada para que otras proposiciones, convergentes dentro del mismo sistema, se acoplen coherentemente.

¿Qué es aquello que confiere validez a las proposiciones emitidas dentro de un sistema de IA?

Siendo honestos, debemos reconocer que lo máximo a lo que podremos aspirar con un análisis lógico será descubrir cuáles son los criterios que dicho sistema emplea de manera interna para validar las transiciones entre proposiciones. Efectivamente, no sería poco conocer qué parámetros y qué modelo de datos son

empleados para validar dichas transiciones.

El monismo, intrínseco a nuestra propuesta filosófica, nos ayuda a percatarnos de lo estéril que resultaría contemplar la posibilidad de un “supuesto sujeto” que confirme o desmienta la validez de las proposiciones internamente generadas por el sistema formal. La posibilidad de disponer de un “sujeto externo” que actúa como verificador imparcial quedará invalidada en cualquier tipo de análisis lógico mínimamente riguroso.

Para llevar a cabo nuestra propuesta, en primer lugar afirmamos que cualquier proposición o respuesta ofrecida por un sistema de IA debe ser absolutamente diferenciada de la propia sensación del usuario que, de manera natural, siempre será ajena al propio sistema formal y que, además, queda fuera del rango de estudio de un análisis lógico. La cuestión de la verdad –adecuación entre lo afirmado en una proposición y el estado de cosas que expresa– es absolutamente independiente de las percepciones que tendrán los usuarios en su relación con tecnologías de IA.

Nuestro análisis lógico-filosófico evitará identificar proposiciones “lógicamente validas y coherentes” con proposiciones que nos informan de la naturaleza verdadera de un evento. Actualmente, la coherencia y solidez del constructo lógico-matemático –el engranaje que sustenta al tipo de programas que genéricamente denominamos bajo el nombre de IA– es de tal calado, que permite generar distintos tipos de respuestas en los usuarios externos que interactúan con determinados entornos que ofrecen experiencias de IA.

Entenderemos que el usuario externo se hace perceptor de los estímulos y respuestas que aporta la IA cuando se sumerge y navega en programas específicos, admitiendo, con anterioridad, un valor de utilidad en las respuestas que le son proporcionadas. Con meridiana claridad, comprobamos que este usuario externo no interviene, en absoluto, ni en la definición pre-sistemática ni en la estructura de axiomas que dota de coherencia al sistema ni en las definiciones que delimitan el alcance de las respuestas que se ofrecen¹.

De manera genérica, los sistemas de IA nos ofrecen la enorme ventaja de aglutinar, ordenar y estructurar enormes cantidades de datos e, incluso, perfeccionar el modo de hacerlo –de manera que cada vez se adecuarán mejor a nuestras necesidades concretas–. Según nuestro análisis, queda fuera del rango del actual modelo de sistemas de IA la posibilidad de ofrecer respuestas que garanticen la verdad.

Respecto a la causalidad

El estudio de la causalidad que emprendemos se ciñe a un entorno plenamente definido de modo formal, allí donde toda proposición que participa del armazón del sistema de IA será tratada como proposición analítica. Decimos que las proposiciones constitutivas de estos sistemas son analíticas; en el sentido de que no pueden ser modificadas o contradichas por proposiciones procedentes de la experiencia o de la intuición del usuario. Dichas proposiciones analíticas, que participan coherentemente en el sistema, son implementadas con independencia de su posible contenido factual.

Debemos recalcar que, al referirnos a los elementos fundamentales que constituyen cualquier programa de IA, es fundamental advertir que todos ellos son ajenos a la experiencia acumulada por cada usuario. La causalidad clásica, respecto a los procesos contingentes, se asienta en la verificabilidad de las proposiciones putativas. De esta manera, la experiencia de cada individuo –en tanto que proceso finito y limitado sometido a múltiples variaciones y connotaciones– determina su correspondiente capacidad predictiva y, por tanto, su capacidad de casación será una cuestión de probabilidad –que mejora con la acumulación de experiencias–. De manera más habitual, consideramos que las relaciones de identidad son las que mejor habilitan al usuario para establecer comparaciones. Vamos a concretar esta afirmación. Cada objeto –datos, imágenes, videos, informes...– implementado en la ciber-realidad² será capaz de generar una serie de respuestas lógico-matemáticas que, a través de los programas correspondientes, serán traducidas en las soluciones demandadas por los usuarios.

Los ciber-objetos³ de los que se nutren los programas de IA, al contrario de los objetos de la realidad física, no tienen que manifestar una forma y cualidades constantes e invariables. La identidad es una relación que se ve desfigurada precisamente por la discontinuidad a la que están sometidos dichos ciber-objetos.

Las relaciones de identidad, que los usuarios externos asimilarán debido a las respuestas que reciben desde la IA, supondrán la desconexión de dichos usuarios con sus referencias experienciales previas, precisamente, por la falta de identidad de los ciber-objetos que son considerados. Una irremediable variación del concepto tradicional de identidad deberá ser revisada pormenorizadamente, pero esta cuestión queda fuera del rango del presente estudio. Como consecuencia, la causalidad se anula al no tener que ser, necesariamente, coincidentes las respuestas emitidas por el sistema de IA y las respuestas conocidas por las experiencias previas de los usuarios.

Debemos percatarnos de que, por más que la arquitectura lógica que articula las respuestas del sistema siga un determinado patrón, dichas respuestas del sistema estarán condicionadas por una infinidad de datos variables que forman parte de cada proposición interna que es tratada. Esta circunstancia motiva que las respuestas del sistema puedan adquirir una uniformidad o variabilidad impredecibles.

Llegados a este punto, advertimos que el problema de la inducción no será tratar de probar que determinadas generalizaciones empíricas, adquiridas previamente por cada usuario, puedan servir con mayor o menor fiabilidad en predicciones futuras. La cuestión que deberá ser resuelta consistirá en despejar y caracterizar el grupo de algoritmos que permiten al sistema de IA producir una proposición concreta.

El uso descontrolado de esta tecnología conduciría al fin de la supremacía de lo experiencial

Nuestro análisis lógico-filosófico nos desvela que cuando nos enfrentamos con el análisis de la verdad y la causalidad en sistemas de IA –constituidos a partir de la aplicación y la convergencia de diferentes lenguajes formales muy sofisticados– nuestra labor crítica debe transcurrir en sentido contrario. Nuestro primer cometido consistirá en ser capaces de entender cómo se ha producido la construcción lógica de ese determinado entorno digital. Ese será el único modo de advertir las bondades y los sesgos que adquirimos al emplear una tecnología determinada.

Conclusiones

Primera conclusión: es fácil advertir que el uso descontrolado de esta tecnología conduciría, de manera inexorable, al fin de la supremacía de lo experiencial.

A la hora de sumergirnos en técnicas de IA es pertinente señalar que el grado de complejidad y la enormidad de los datos consultados impide a los usuarios disponer de criterios para validar o desdeñar –de manera inmediata– las respuestas recibidas. La totalidad de los ciber-objetos o relaciones implementados en la ciber-realidad podrán ser caracterizados, plenamente, en base a su estructura lógico-formal. El estudio –minería, análisis, topología...– de grandes volúmenes de ciber-datos nos permitirá revelar dichas estructuras constitutivas y primitivas al analizar las repuestas emitidas –el análisis pormenorizado de muchas de las respuestas o propuestas que nos llegan desde la IA exigirá un largo y complicado estudio de detalle–.

Segunda conclusión: su implementación deberá realizarse de modo que evitemos el desmantelamiento de las más profundas y peculiares habilidades cognitivas del ser humano. Es difícil imaginar qué ocurriría si la experiencia natural acumulada por cada usuario externo dejase de suponer una referencia fiable o, dicho de otro modo, que dejase de implicar una ventaja o mejora adaptativa, incluso, llegado el caso, que pudiera convertirse en una desventaja porque la experiencia previa resulte inadecuada según la respuesta que nos ofrece la IA

Tercera conclusión: nuestro análisis lógico-filosófico señala que toda posible normativa encaminada a la regulación de la IA, que no tenga en cuenta la caracterización de la estructura lógica que sustenta a cada tipo de sistema particular, será tan vulnerable como imprecisa.

Boole, G. (1847): *The mathematical analysis of logic*. London, Macmillan.

Carnap, R. (1928): *Der Logische Aufbau der Welt*. Berlin-Schlachtensee, Weltkreis-Verlag.

Frege, G. (1879): *Conceptografía, un lenguaje de fórmulas, semejante al de la aritmética, para el pensamiento puro*. México: UNAM, Instituto de Investigaciones Filosóficas, 1972.

García-Baró, M. (1993): *Categorías, intencionalidades y números*. Madrid, Tecnos.

Russell, B. (1903): *Principles of mathematics*. Cambridge, Cambridge University Press.