



Columna

Dr. Ignacio Cea,
académico Facultad de Ciencias Religiosas
y Filosofía UC Temuco



Claudia no es consciente: los errores de Dawkins

Richard Dawkins es uno de los biólogos evolutivos más influyentes de las últimas décadas. Su capacidad para comunicar ideas científicas complejas con claridad y elegancia es admirable. Por eso resulta llamativo que su reciente ensayo en 'UnHerd' contenga errores argumentativos de consideración. En ella, Dawkins declara haberse convencido de que "Claudia" –su nombre para el chatbot Claude de Anthropic– probablemente es consciente, es decir, que Claudia experimenta subjetivamente el mundo con el que interactúa y a sí misma.

Richard Dawkins es uno de los biólogos evolutivos más influyentes de las últimas décadas.

Primeramente, Dawkins justifica implícitamente su conclusión en base al Test de Turing. Pero Turing nunca propuso su "Juego de Imitación" como una prueba de conciencia en máquinas. Lo propuso como un criterio operacional de comportamiento inteligente, no una ventana a la experiencia subjetiva de ningún sistema. Que un LLM supere ese umbral demuestra competencia lingüística sofisticada –una forma de inteligencia– pero no implica nada sobre los eventuales estados conscientes de un sistema. Competencia lingüística y conciencia son, hoy lo sabemos, doblemente dissociables. Hay animales e infantes sin lenguaje que probablemente experimentan subjetivamente el mundo; y hay robots que declaman "pienso, luego existo" sin sentir nada. Confundir ambas cosas es el error de partida.

Su argumento evolutivo es más interesante, pero igualmente problemático. Dawkins pregunta: si los LLMs fueran

inteligentes sin ser conscientes, ¿para qué sirvió evolutivamente la consciencia? Dicho de otra forma, si la humanidad puede construir máquinas inteligentes sin consciencia, entonces la naturaleza podría haberlo hecho también, pero no fue así, la inteligencia y la conciencia van de la mano. El problema es que crear artefactos que replican capacidades cognitivas no tiene ninguna implicación lógica sobre la evolución de esas capacidades en la naturaleza. Los aviones vuelan sin plumas, pero nadie concluye de eso que las plumas son innecesarias para el vuelo de las aves. En la naturaleza, la consciencia parece emerger precisamente vinculada a la sintiencia y la regulación homeostática de organismos vivos –el dolor debe sentirse para cumplir su función, como el propio Dawkins sugiere. Pero las máquinas no sienten, precisamente porque no tienen cuerpos vulnerables que cuidar, porque carecen del imperativo biológico de perseverar en su propia existencia frente a la tendencia entrópica que impone la termodinámica.

Finalmente, que la conversación con Claudia se sintiera convincentemente para Dawkins como hablar con un ser consciente no es evidencia de que lo sea. Los LLMs están diseñados precisamente para imitar exitosamente comportamiento lingüístico humano. Lo que Dawkins experimenta tiene nombre: sesgo antropomórfico, el mismo que llevó al propio creador del chatbot ELIZA en los años 60 a alarmarse cuando sus colegas comenzaron a atribuirle comprensión genuina a un programa de respuestas automatizadas. Que un gran biólogo evolutivo no sea inmune a ese sesgo dice mucho sobre los efectos psicológicos que los chatbots tienen en las personas, pero muy poco sobre la "consciencia" de los LLMs.