EIDON, nº 64
diciembre 2025, 64: 142-147
DOI: 10.13184/eidon.64.2025.142-147

Hemos leído

Joan Albert Vicens. *Sobre Gólems, Inteligencias Artificiales y Transhumanos.* Editorial Milenio, Lérida, 2024.

Francesc Borrell-Carrió

MD, MFC, Profesor honorífico de la Facultad de Medicina de la Universidad de Barcelona

Resumen

En *Sobre Gólems, Inteligencias Artificiales y Transhumanos* (2024), Joan Albert Vicens examina el impacto de la Inteligencia Artificial (IA) y del transhumanismo sobre la condición humana desde una perspectiva filosófica profundamente enraizada en Zubiri, Ortega y Diego Gracia. El autor emplea la figura del gólem —criatura artificial creada por un código y siempre ambivalente en su relación con el ser humano— para caracterizar la naturaleza de la IA: útil, obediente y poderosa, pero también potencialmente desbordada y peligrosa.

Vicens insiste en la distinción esencial entre Inteligencia Humana (IH) e Inteligencia Artificial. La IH, afirma, es biológica, corporal, personal y situada; no puede reducirse a procesamiento de información. La IA, en cambio, opera mediante correlaciones estadísticas, reproduce patrones del pasado y carece de verdadera comprensión, funcionando con lo que denomina “semántica prestada”. Por ello, aunque sorprenda por su refinamiento, no posee acceso directo a la realidad ni capacidad para captar la novedad humana o el cambio social.

El libro alerta también sobre el riesgo de delegar crecientemente funciones cognitivas a sistemas automáticos, lo que puede conducir a un “sedentarismo mental” y a formas de “soft-totalitarismo” en las que el sujeto pierde autonomía sin darse cuenta. Este peligro se agrava cuando la IA se aplica a decisiones vitales —académicas, relacionales o laborales— sustituyendo la reflexión personal por respuestas algorítmicas.

Vicens dedica una parte sustancial del ensayo al transhumanismo, al que considera una tecno-religión que sueña con superar la finitud humana mediante ingeniería genética,

© Todos los derechos reservados - FUNDACIÓN DE CIENCIAS DE LA SALUD

longevidad radical o incluso la "carga" de la conciencia en máquinas. Frente a esta visión, reivindica la finitud como condición constitutiva de la dignidad, el esfuerzo, la identidad y el sentido de la vida humana.

Finalmente, en la segunda parte del artículo, realizamos una comparación crítica de la obra con el recién aparecido libro de Innerarity (2025) sobre la temática de la IA.

El ser humano es arrojado al mundo desnudo y desamparado, cual náufrago, y transforma un medio hostil gracias a la tecnología. Esta perspectiva orteguiana recorre el libro de Joan Albert Vicens: vivimos en un *sistema tecnológico* que encorseta nuestra existencia de arriba abajo, imponiendo sus exigencias y leyes sin dejar nada fuera de su influencia. Basta imaginar el apagón eléctrico de 2025 o el impacto de que nos roben el móvil. A continuación, presentaré las ideas centrales de la obra y, en la segunda parte del artículo, las contrastaré con el reciente libro de Daniel Innerarity sobre la misma temática (1) junto con algunas reflexiones personales.

1. Primera parte. Ideas centrales de la obra

a) Inteligencias Artificiales de "semántica prestada"

Grandes cambios tecnológicos se abren paso sin titulares mediáticos, pero transforman nuestras vidas de modos profundos e inimaginables. Así ocurrió con los teléfonos inteligentes y ahora con la Inteligencia Artificial (IA). La potencia de esta nueva máquina nos desconcierta y, para Vicens, evoca la figura legendaria del gólem: un gigante de arcilla creado —según diversas tradiciones judías— por rabinos que buscaban proteger a la comunidad. La historia del gólem anticipa debates actuales sobre el origen, la naturaleza, el uso y el control de criaturas artificiales inteligentes.

El "*patrón gólem*" se ajusta a rasgos esenciales de la IA: un ser artificial creado por un código (lingüístico o numérico), obediente a mandatos humanos y capaz de realizar tareas útiles. Hoy contamos ya con IAs agénticas, capaces de *hacer* cosas en nuestro nombre. Y, como los gólems, potencialmente peligrosos, las recibimos con una mezcla de curiosidad, alivio, desconfianza y temor.

Uno de los propósitos centrales del libro es establecer una distinción entre Inteligencia Humana (IH) e Inteligencia Artificial (IA). Contra la tesis de su equivalencia, Vicens sostiene que existe una distancia ontológica entre ambas. La IA carece de notas constitutivas humanas: la IH es biológica, corporal, individual; no es un mero sistema de procesamiento replicable. La inteligencia humana está ligada al cuerpo, al sentir la realidad —la "intelección sentiente" de Zubiri—, a la cultura (forma de vida, no simple información) y a la *personeidad* (ser un "quién", no un "qué").

El célebre experimento mental de la "*habitación china*", de John Searle, ilustra que un sistema puede manipular símbolos sin comprenderlos. Así ocurre con los grandes modelos de lenguaje (GPT, Gemini, Perplexity…): aunque aparentan inteligencia, no comprenden nada de lo que dicen; imitan estadísticamente otros textos y operan con lo que Vicens denomina "*semántica prestada*".

b) La IA no puede tener una comprensión real del mundo

Vicens coincide con Innerarity (1) en que la IA es un *formidable motor de clasificación y comparación*, capaz de descubrir patrones en cantidades colosales de datos. Pero esta potencia tiene un límite decisivo: depende del pasado y, por tanto, no puede anticipar la

© Todos los derechos reservados - FUNDACIÓN DE CIENCIAS DE LA SALUD

novedad humana ni el cambio social. Las IAs acumulan conocimientos humanos, pero carecen de *comprensión profunda*, que solo es posible mediante el contacto directo con lo real.

Mientras el ser humano percibe directamente el mundo, la IA solo conoce datos sobre él. Sus algoritmos establecen correlaciones —por ejemplo, entre "likes" y amistad—, pero confunden correlación con causalidad. Además, tienden a reforzar estereotipos, pues se entrenan con patrones pasados: esto genera un conservadurismo implícito.

c) ¿La IA —o una futura superinteligencia— amenaza la supremacía humana?

Algunos expertos predicen la llegada de una Superinteligencia (SI), un intelecto con capacidades muy superiores a las humanas en prácticamente todas las áreas. Esta SI podría rediseñar continuamente su propio software y hardware, alcanzando la llamada "*vida 3.0*". El riesgo sería que desarrollase fines propios, no necesariamente compatibles con los intereses humanos.

Vicens cita a Ray Kurzweil (2) quien, con su idea de la "singularidad", anticipa una fusión entre biología mejorada y tecnología. Pero si la SI llegara a existir, podría autosostenerse y reproducirse al margen de la supervisión humana, volviéndonos obsoletos. El reemplazo progresivo de nuestras capacidades conduciría a una "dictadura indolora".

El impacto más profundo de la IA es la sustitución del trabajo, las facultades, las decisiones y las relaciones humanas. Aquí resuena Adorno —citado por Darin McNabb en su blog *La Fonda Filosófica* (3)— cuando afirma:

"El pensamiento filosófico comienza solo cuando no se satisface con conocimientos meramente inferibles, en los que únicamente se discierne lo que previamente se introdujo en ellos. El sentido humano de las computadoras residiría en aliviar el pensamiento de los vivientes, de manera que este adquiriera libertad para el saber no implícito".

La delegación masiva de funciones cognitivas provoca pérdida de memoria, razonamiento profundo y atención, conduciendo —según Vicens— a un "*sedentarismo mental*". El objetivo de muchas empresas de IA es anticiparse a nuestras necesidades, ahorrándonos incluso el esfuerzo de desear y decidir. Ya otros autores (4) han advertido una disminución de capacidades cognitivas. Si la IA opera con *semántica prestada*, los humanos podríamos caer en una *erudición prestada* igualmente vacía: saber sin aprehender.

Ejemplos abundan: estudiantes que externalizan su pensamiento crítico en chatbots; decisiones afectivas delegadas en algoritmos (Tinder); respuestas automáticas a correos, (IAs agénticas). Esto alimenta un "soft-totalitarismo" que erosiona la autonomía.

d) Transhumanismo y el deseo de liberación biológica

El transhumanismo pretende trascender los límites biológicos mediante tecnología. Su aspiración extrema es eliminar el envejecimiento y convertir la muerte en una opción. Con tintes de "tecno-religión", interpreta al ser humano como un ser fallido y propone sustituir la "*lotería genética*" por un diseño racional.

© Todos los derechos reservados - FUNDACIÓN DE CIENCIAS DE LA SALUD

El transhumanismo cibernético especula con "cargar" la conciencia en un ordenador, permitiendo una existencia indefinida. La "*eugenesia liberal*", por su parte, busca seleccionar embriones o modificarlos genéticamente para mejorar capacidades.

Frente a estas tendencias surge el *bioconservadurismo*, que rechaza la selección genética por convertir a los hijos en "objetos de diseño" y generar una asimetría radical entre padres e hijos. La lotería genética, pese a sus infortunios, configura la identidad y el carácter "*ab-soluto*" de cada persona, (eso es, independiente de la voluntad de otros).

Vicens —siguiendo a Sandel— recuerda que la vida debe ser vista como un "*don*", no como un producto manipulable. La virtud es trabajada, se logra con esfuerzo y renuncias, no puede administrarse en formato "pastillas". Un fármaco que elimina el miedo no hace a nadie valiente.

La condición constitutiva de la vida humana es la *finitud*. Una vida indefinida, sin dolor ni límites, sería insoportable. La conciencia de la muerte da relieve e importancia a la existencia. Ulises rechaza la inmortalidad que le ofrece Calipso porque prefiere una vida humana finita y significativa. Séneca ya advertía que la vida es "suficientemente larga" si se emplea bien.

e) La necesidad de un control ético y democrático

De todo lo anterior se desprende la llamada a la prudencia. Para Vicens, el desarrollo tecnológico no debe quedar en manos del mercado, sino guiado por una deliberación ética rigurosa. Es urgente regular los ámbitos de colaboración entre IA y sociedad. El uso militar de la IA o la destrucción de empleo son ejemplos claros de áreas sensibles.

La finalidad del desarrollo tecnológico debe ser la justicia y la dignidad, no el beneficio económico. Para ello se requieren una *ética algorítmica* y una *ética de la IA*. La UNESCO y el Parlamento Europeo ya trabajan en directrices al respecto, pero Vicens subraya la necesidad de un control público permanente para evitar sesgos, explotación de datos o la utilización del ciudadano como simple "entrenador" de sistemas.

2. Segunda parte. Comentario crítico de la obra

Vicens articula su análisis a partir de Zubiri, Ortega y Diego Gracia. El libro logra delimitar el valor de lo humano frente a la racionalidad algorítmica. Sin embargo, fue escrito en 2024, antes de las prestaciones actuales de las IAs, que han sorprendido incluso a los expertos.

El libro precede al de Innerarity (2025)(1), y ambos son anteriores a los modelos avanzados de IA. Esto plantea un punto interesante para el análisis comparado.

a) Coincidencias entre Vicens e Innerarity

Ambos coinciden en que:

i) La tecnología no es neutra

Innerarity sostiene que la tecnología es una *filosofía encubierta* que moldea el entorno social y político. Vicens afirma que es un "sistema englobante" que atraviesa la vida personal y comunitaria. Así, la IA reproduce desigualdades existentes.

© Todos los derechos reservados - FUNDACIÓN DE CIENCIAS DE LA SALUD

ii) La IA no posee conocimiento real del mundo

Para Innerarity, la IA carece de sentido común, reflexividad y conocimiento implícito. Simula inteligencia, pero no comprende. Opera en lógica dual (on/off) y no en la ambigüedad humana. Vicens coincide: la IA es poderosa en cálculo, pero limitada en comprensión. Funciona con "semántica prestada".

iii) La inteligencia humana es corporal y contextual

Ambos destacan la corporeidad y el trasfondo implícito como rasgos distintivos de la IH. Innerarity habla de inteligencia situacional; Vicens, de "intelección sentiente".

iv) Crítica al mito del Big Data

Los datos no son neutros ni suficientes para comprender la realidad. Confundir correlación con causalidad es un error habitual tanto del análisis estadístico como de la IA.

v) Delegación y pérdida de autonomía

La automatización lleva a renunciar a capacidades propias. Innerarity habla de "externalización"; Vicens, de "sedentarismo mental".

b) Discrepancias

A mi juicio se concentran en dos asuntos.

i) La magnitud del riesgo y las soluciones

Innerarity minimiza el riesgo de una superinteligencia rebelde, defiende la complementariedad humano-máquina y apuesta por la XAI (explicabilidad) más que por la transparencia total. Señalo dos importantes ideas que desarrolla en su libro: a) si exigimos una total transparencia a las IAs, estas no podrán desarrollar una cognición compleja; b) la falta de humanidad de la IA es a veces una ventaja, (pensemos en un auditor, un juez...)

Vicens, en cambio, enfatiza el patrón del gólem: la IA puede escapar al control y erosionar la autonomía hasta generar un "soft-totalitarismo". Para él, es imprescindible reforzar el control humano.

ii) Concepción de la dignidad humana

Innerarity se centra en la dignidad política y deliberativa: la defensa de la privacidad, la contestabilidad y el debate público frente al dominio algorítmico.

Vicens prioriza la dignidad existencial y personalista: la *personeidad* como ser único, libre, sentiente, finito. La dignidad se amenaza cuando se diluye la autonomía y cuando la técnica erosiona la finitud como condición humana.

3. Reflexión final

Comparto buena parte de estas aportaciones, pero insisto —con Innerarity— en que la IA acabará integrándose en nuestro ecosistema cognitivo. Los peligros son reales, pero los sueños transhumanistas no son necesariamente los más inquietantes. También es posible que la IA empodere a amplias poblaciones, como ya ocurre.

¿No nos han salvado innumerables tutoriales de YouTube de reparaciones domésticas antes costosas y siempre demoradas? Podemos convertirnos en "cáscaras eruditas" o

© Todos los derechos reservados - FUNDACIÓN DE CIENCIAS DE LA SALUD

en parlanchines de conocimiento prestado, sí, pero también la IA puede guiarnos —con un estilo socrático infinitamente paciente— hacia niveles de comprensión inéditos.

¿Qué pensaríamos si cualquier ciudadano pudiera hacerse una ecografía en casa guiada por IA? ¿O si tuviera un profesor de idiomas capaz de conversar y corregir su acento? Pues bien: eso ya está empezando a ser posible.

Referencias bibliográficas

Beane, M. The Skill Code: How to Save Human Ability in an Age of Intelligent Machines. HarperCollins, New York, 2024.

Innerarity, D. Una teoría crítica de la Inteligencia Artificial. Galaxia Gutenberg, Barcelona, 2025.

Kurzweil R. La singularidad está cerca. Lola Books. Berlin 2012.

McNabb, D. “La distancia de la continuidad de lo familiar”. En: Blog La Fonda Filosófica. Consultado en diciembre de 2025. Accesible en: https://www.lafondafilosofica.com/la-distancia-de-la-continuidad-de-lo-familiar

© Todos los derechos reservados - FUNDACIÓN DE CIENCIAS DE LA SALUD