

Álvaro Carvajal Villaplana

Claudio Gutiérrez: en los albores de la inteligencia artificial¹

Resumen: *El artículo expone y revisa algunas de las ideas más destacadas en relación con la posibilidad de pensar de las máquinas computadoras u ordenadores, en discusión contra los argumentos presentados por Searle y Penrose. Se analiza el concepto de algoritmo y su importancia para la informática, la robótica y la IA. Luego, se expone su idea sobre la informática como «ciencia»; por último, se plantean los problemas sociales que genera la informática y la IA.*

Palabras clave: *Claudio Gutiérrez Carranza, Inteligencia Artificial, informática, humanismo, algoritmo.*

Abstract: *In this article we present and analyze some of Gutierrez's more relevant ideas related to the possibility of machine thinking, in particular his replies to the objections raised by Searle and Penrose. The notion of algorithm is explained, as well as its relevance for computer science, robotics and artificial intelligence. We consider the scientific status of this field as defended by the author. Lastly, we pose some social problems that informatics and artificial intelligence may give rise to.*

Keywords: *Artificial Intelligence, computer science, humanism, algorithm.*

Introducción

El pensamiento de Claudio Gutiérrez sobre la informática y la inteligencia artificial se encuentra en tres obras: (1) el libro *La sociedad computarizada* de 1990 (Volumen III, *Obras Completas*), la cual recoge y analiza una serie de temas; además, es una guía de lectura para una antología de artículos; (2) de 1993, *Epistemología e informática* (Volumen I, *Obras completas*); (3) en el 2006, *El humanismo replanteado: genes y menes en la sociedad globalizada* ([2006] 2011, *Obras completas*, Volumen V); y, por último, (4) en el volumen III de la *Obras completas* (2011) una serie artículos periodísticos entre 1968-2003. La edición de las *Obras completas* estuvo a cargo del Dr. Luis Camacho Naranjo, fue editada por la Editorial de la Universidad de Costa Rica, en el 2011. Los dos primeros textos muestran la vocación didáctica por parte del autor en estudio y fueron publicados por la Universidad Estatal a Distancia (UNED-Costa Rica), son textos didácticos de apoyo a los cursos de dicha universidad.

En general, llama la atención que en dichos textos Gutiérrez muestra un gran optimismo hacia la informática y la Inteligencia Artificial (IA), tal como si se tratase de una especie de utopía, en tanto que representan la solución de todos los problemas. Sin embargo, al mismo tiempo, Gutiérrez señala los problemas que trae la introducción de la IA a la sociedad y la cultura.



El planteamiento de dichos cuestionamientos, se encuentra en sus escritos de los años noventa. Ahora, si bien evidencia tal optimismo ante los logros y el acelerado avance de la informática y la inteligencia artificial; al mismo tiempo, se muestra prudente, por cuanto considera que dichos resultados solo han revelado el problema que significa crear una inteligencia artificial, por lo que considera que todavía falta mucho por hacer para lograrlo. Lo cierto es que Gutiérrez nunca asume una oposición radical contra la informática y la IA, sino que hace críticas a aplicaciones informáticas específicas. En todos los casos analizados por Gutiérrez, si bien se resaltan las dificultades que presenta la informática y la IA, a tales problemas las percibe como obstáculos por superar, lo anterior desde una óptica de la prudencia razonable. Cabe recalcar que solo en un aspecto Gutiérrez mantiene un principio como absoluto e imperioso: el respeto de la privacidad de las personas.

En este artículo se expone y revisa algunas de las ideas más destacadas en relación con la posibilidad de pensar como los humanos por parte de las máquinas computadoras u ordenadores, en contraste y en discusión de los argumentos escépticos presentados por Searle y Penrose. Igualmente, se analiza el concepto de *algoritmo* y su importancia para la informática, la robótica y la IA. Luego, se expone su idea de la informática como «ciencia». Por último, se plantean los problemas sociales y éticos que genera la informática y la IA.

La posibilidad del pensar artificial

En los textos enunciados en el acápite I, Claudio Gutiérrez se nos presenta como un promotor de la informática, los robots y la inteligencia artificial. Está convencido, a pesar de algunos matices o reservas— que las máquinas pueden pensar, de tal manera que, si no lo hacen como el ser humano, al menos lo pueden emular. La inteligencia artificial para Gutiérrez, siguiendo a John McCarthy, consiste en el

intento de capacitar a las máquinas para realizar actos que, si los hiciera un ser humano,

diríamos que requieren inteligencia. Esta capacitación se trata de impartir dotando a la máquina en cuestión de algoritmos informáticos que sean capaces de solucionar problemas como los que enfrentamos rutinariamente; o en situaciones muy específicas, como al jugar ajedrez u otros juegos de salón, probar teoremas, resolver acertijos, entender o producir lenguaje, reconocer objetos de manera visual, dar diagnóstico médico o armar un itinerario de viaje, evitar obstáculos al caminar, etcétera (...) ([2006] 2011, 18).

Este entusiasmo se evidencia tanto en el libro *El humanismo replanteado: genes y menes en la sociedad globalizada* ([2006] 2011), como en *Epistemología e informática* ([1993] 2011). Sin embargo, él considera que la IA tiene sus escépticos, en particular se refiere a los argumentos de Searle y Penrose, una vez expuestos se dedica a contrargumentar.

El análisis del argumento de Searle es importante para él, en tanto que le sirve para definir las potencialidades y las limitaciones del paradigma informático, ya sea como (a) instrumento para entender la mente humana y sus fenómenos cognitivos o (b) como una herramienta que emula a la mente humana; al respecto revisa el argumento de Searle (1980) del *cuarto chino*, el cual se trata de un experimento mental, que —según Gutiérrez— intenta refutar la tesis de la capacidad de los incipientes programas informáticos para la comprensión del lenguaje natural, tal comprensión del lenguaje es solo un caso particular de la actividad mental humana, realizada por las máquinas (1993, 258). Dicho experimento presume un ataque a la IA Fuerte, no así a la IA Débil, ya que alude a una persona que no sabe nada de chino, la que se encuentra encerrada en una habitación, con unas instrucciones en otro idioma, tales instrucciones son unas listas que indican cómo manejar las formas de los símbolos en chino; de tal manera que cuando dicha persona recibe una lista de símbolos en chino, luego devuelve otra lista de símbolos en chino, la que es una respuesta adecuada a la primera. En tales condiciones para un(a) observador(a) exterior a la habitación, las respuestas de la persona encerrada en la habitación no se distinguiría de la de

un(a) hablante de chino. Sin embargo, la persona encerrada en la habitación produce sus respuestas tan solo manipulando símbolos, no las interpreta, ni las comprende. Para Searle, una computadora se comporta igual que la persona encerrada en la habitación china, de tal manera, que da la impresión de pensar y comprender, pero en realidad se limita a manipular símbolos formales que no comprende. En todo caso, ambos (la persona y la máquina pueden superar la prueba de Turing); empero, eso no significa que puedan pensar y comprender como los seres humanos. Para Searle las únicas máquinas pensantes son los cerebros, es decir, los entes biológicos. Las computadoras solo están dotadas de sintaxis (símbolos y reglas), no conocen la semántica, no saben del significado de los símbolos.

El argumento del *cuarto chino* se basa en que no es suficiente saber manipular símbolos formales no interpretados para atribuir que se conoce un idioma, para el caso: el chino. De tal manera que una computadora -adecuadamente programada- comprende y explica la comprensión el chino ([1993] 2011, 262). Según Gutiérrez, Searle se opone a la idea de que sea posible producir inteligencia por medios estrictamente formales, lo cual no tiene que ver con el soporte material de la mente o la comprensión. La crítica de Searle se dirige a la que Gutiérrez llama «posición fuerte» o IA Fuerte, aquella sostenida por los(as) investigadores(as) que intentan producir una inteligencia igual a la humana. Por otra parte, Searle no critica la «posición débil» o IA Débil, aquella que se conforma con la idea de que los programas lo que muestran son ciertos aspectos de la comprensión humana, con lo que contribuyen a explicar el funcionamiento de la inteligencia.

Un contraargumento que presenta Gutiérrez contra la idea del *cuarto chino* es el que desarrolla Dennett, cuando afirma que si bien en el *cuarto chino* el programa no logra una comprensión plena², es el caso de un chiste; no obstante, esa limitación no invalida a la inteligencia artificial, ya que ésta es concebida como una complejidad, de cuyas partes pueden emerger las características que no están en las partes, sino tales características están en la totalidad, por lo que Dennett no descalifica que

a los billones de acciones de todas estas partes sumamente estructuradas como capaces de producir una auténtica comprensión. Después de todo, eso es lo que tenemos de base para suponer que hace el cerebro; integrar de una manera congruente también billones de acciones de partes sumamente estructuradas (...). (Gutiérrez 1990, 267).

Para Dennett y Gutiérrez, Searle al hacer esto, está rescindiendo al cerebro, es decir, a nosotros mismos, al humano del que se supone que en algún momento tiene una auténtica comprensión (267). La analogía sería en que el cerebro es un sistema complejo, compuesto por las partes y sus relaciones, de tal manera que si bien las partes no comprenden, la totalidad como tal sí comprende. De tal manera que para Gutiérrez el argumento de Searle es una falacia de composición; especialmente porque para él, si las partes no presentan una comprensión, entonces el todo no tiene comprensión; siendo que lo que sucede en cerebro es lo contrario, de tal manera que dicha complejidad y de la interacción de las partes, por sí misma, y no por un milagro, es lo que produce la comprensión, como ya se dijo, esto es así conforme al argumento de Dennett. La falacia se daría por ignorancia, según Gutiérrez, ya que, por parte de Searle, «no habría familiaridad con la programación electrónica o cualquier otro tipo de actividad constructiva o analítica del mismo carácter. Ello porque cualquier construcción suficientemente grande produce de camino «propiedades emergentes» que no se daban en las etapas intermedias» ([2006] 2011, 267).

Ahora, según la visión materialista, «se tiene que la comprensión genuina se produce por un proceso compuesto de interacciones entre una hueste de subsistemas ninguno de los cuales entiende nada por sí mismo» (Gutiérrez 1993, 267). Aplicada esta analogía a la IA, esto implicaría que la mayor complejidad de un programa o algoritmo es lo que produciría la comprensión; aunque, cada una de sus partes no entienda nada por sí misma. Al respecto Gutiérrez asume que la comprensión sería una propiedad emergente, la cual es independiente del soporte físico, por lo que dicho soporte podría ser el cerebro o cualquier otro, y esto, por así decirlo, se produciría

«naturalmente» o de manera espontánea (267). Es claro que Gutiérrez defiende una IA Fuerte³. Pero un argumento opuesto al contraargumento de Gutiérrez señala que la comparación entre el cerebro y el cuarto chino es absurda, ya que supone que la persona en el cuarto chino no entiende chino; empero, la habitación en su conjunto sí que entiende el chino. Para Searle esto sería un absurdo, pues entonces la habitación ¡es un ente que piensa y se comunica en chino!

En todo caso, Gutiérrez interpreta que la objeción central de Searle a la inteligencia artificial, no reside en que el pensamiento solo se produzca por vías biológicas; ya que si fuese así la IA sería imposible, en tanto que la inteligencia solo puede surgir por dicha vía. Él interpreta que no se excluye el desarrollo de la IA por medios puramente formales; por lo cual, continuando con la analogía máquina-cerebro, Gutiérrez entiende que lo que dice Searle consiste en que para producir el pensamiento se deben seguir los mismos principios, y probablemente los mismos materiales, de que están hechas las «máquinas pensantes naturales (los cerebros humanos)»; aunque, Gutiérrez despacha rápido la discusión, ya que, según él este aspecto tiene que ver con un debate empírico que no puede aclararse por el momento ([1993] 2011, 263). En razón de dicha interpretación es por lo que Gutiérrez insiste en la tesis funcionalista que asevera que no es importante la soporte material, sino que los aspectos funcionales y estructurales son suficientes para crear la inteligencia artificial. Pero como ya se indicó, para Searle la intencionalidad, la conciencia y la comprensión son un fenómeno biológico por lo que dependen de la bioquímica. Gutiérrez considera que la posición teórica de Searle lo que hace es introducir nuevo dualismo de tipo cartesiano, ya que distingue entre fenómenos estrictamente biológicos frente a los sistemas mecánicos y formales (264); aunque no es claro en qué sentido dicha distinción se trata de un dualismo cartesiano, tal como Ryle interpreta el dualismo de sustancia de Descartes.

Gutiérrez considera que el argumento de Searle es un cuento de hadas, el cual quiere refutar, para ello de nuevo recurre a un argumento de Dennett expuesto en *Explicación de la ciencia* (1991). Él indica que según Dennett, el

argumento de Searle ha recibido muchas objeciones, las cuales son emitidas por diferentes pensadores(as), y que el argumento del experimento mental del cuarto chino, al tener tal nivel de contraargumentos ha de deberse a alguna razón, y tal razón reside «en el hecho que la gente no imagina verdaderamente la situación descrita con el detalle que esa misma descripción exige» ([1990] 2011, 265). Para Gutiérrez el «punto crucial es que el Cuarto Chino habría pasado la prueba de Turing» (265); claro es posible que esto sea así, ya que la prueba lo que dice es que si algo, un algoritmo parece comportarse como inteligente, entonces, es inteligente; pero se trata de una prueba de imitación. Ahora, si bien una IA puede comportarse como inteligente, esto no quiere decir que sea una inteligencia como la humana y que comprenda. La superación de la prueba de Turing, para el caso en estudio, lo que muestra es la posibilidad de la IA Débil, con lo cual, efectivamente Searle no estaría en desacuerdo. Este argumento es semejante al que plantea Descartes (Véase Carvajal 2020). En todo caso, como ya se indicó, Gutiérrez lo que argumenta es que si bien el algoritmo no tiene plena comprensión (el chiste), esto no descalifica que la máquina tenga posibilidades de adquirir una auténtica comprensión; pero, ¿cómo se sabe que la máquina adquiere una plena comprensión?

En relación con la consecuencia que se desprende de la tesis de la emergencia de la comprensión o la inteligencia desde sistema complejo, y recurriendo a Newell y Simon, Gutiérrez asume la idea de dichos autores de que un sistema de símbolos tiene las condiciones necesarias y suficientes para la acción inteligente natural ([1993] 2011, 152), por lo que, se trata de una ley de estructura cualitativa; además, de una hipótesis empírica. En este punto, insiste en que el desarrollo de la inteligencia no tiene que ver con el soporte material, sino que lo importante es el sistema de símbolos, por ende, no la materia; es decir, la función de las partes en dicho sistema, eso es lo que explica la inteligencia (153). Empero, al analizar Gutiérrez este asunto a partir de la *teoría de los niveles epistemológicos* de Dennett, tal parece que de dicha explicación se desprenden algunos límites a la defensa que hace de la IA Fuerte. Tal teoría

distingue tres niveles o perspectivas: (1) físico: la construcción de una computadora, (b) Diseño: la perspectiva que adopta el programador (a) y (c) intensional⁴: el que corresponde al usuario de un programa en cuanto a la actitud de un sujeto hacia el objeto (crear, desear, tener, querer, etc.) ([1993] 2011, 157). Él concibe la intencionalidad como «todo aquello que tiene la naturaleza de un deseo o creencia, incluso el conocimiento como creencia justificada» (157). Siguiendo a Dennett, también considera que la intencionalidad «se relaciona con las estrategias de alguien para tratar de explicar o predecir un comportamiento. Por ejemplo, podemos considerar como sistema intensional a algo tan simple como un termostato» (157). Con lo cual usa una noción muy amplia de *intencionalidad*.

Tal uso de la noción de *intencionalidad* —según Dennett— sirve para explicar el comportamiento de un programa inteligente, en tanto *diseño óptimo* del programa. Este supuesto es semejante al de la *racionalidad*, lo que hacemos cuando interactuamos con otras personas y le atribuimos a la otra persona racionalidad ([1993] 2011, 157). Gutiérrez agrega que tal presunción de racionalidad no implica que la computadora sea una persona, sino que se trata solo de considerarla como un «sistema intensional», lo cual —según Dennett— tiene que ver con un «sistema capaz de actuar como si tuviera creencias, propósito, planes, etc. No tenemos completamente claro en qué consista para un humano tener creencias, propósitos, etc.» (158). De tal manera que cuando una computadora hace un movimiento en un juego de ajedrez, simplemente hay que afirmar que tiene una «mente», algo semejante a lo que tendría un ser humano en las mismas circunstancias. La diferencia entre la mente y la computadora reside en que en el caso de la última se sabe cómo funciona esa «mente», ya que ha sido diseñada por un ser humano. Esto supone que no se sabe cómo funciona la mente humana. La siguiente pregunta es saber por qué la IA hizo lo que lo hizo, según Gutiérrez eso es más difícil de responder, a lo que recurre es la perspectiva del diseñador, quien no podría dar una respuesta definitiva de por qué lo hizo; aunque, para él, lo más fácil es saber por qué no lo hizo. En todo caso, la computadora o el algoritmo, por el

mismo, tampoco puede dar una explicación de por qué hace algo.⁵

Igualmente, Gutiérrez considera que llanamente se le ha de atribuir comprensión a la IA, en el caso del juego de ajedrez, como si esta fuese un jugador humano ([1993] 2011, 160); de la misma manera, tal atribución no exige que se le trate como una persona. Empero, esta forma de argumentar ya no parece referirse a la IA Fuerte, sino que remite a la IA Débil; en el sentido de que se trata de un juego de imitación, como ya lo estableció Descartes (véase Carvajal, 2020) y luego Turing. En todo caso, la consideración de *hacer como si* _____, no quiere decir que exista una inteligencia igual a la humana. De tal manera que Gutiérrez no se compromete con la existencia ontológica de una IA Fuerte, sino que se limita a afirmar que lo que hay es un «sistema intensional», lo cual remite de nuevo al planteamiento de la IA Débil. Por ello, el «sistema intensional» —según Gutiérrez— no es metafísica, no es un alma inmortal, ni una constitución biológica o una conciencia moral. Ya que, es más fácil decidir si una máquina es un sistema intensional, a solventar que una máquina piensa o tiene conciencia o es moralmente responsable. Pero, si esto es así, entonces, ¿por qué rechazar el argumento de Searle?

Ahora, según Gutiérrez puede adoptarse —analogicamente— la perspectiva de la persona ([1993] 2011, 161), solo como un artilugio para explicar el comportamiento de los seres humanos; pero, eso mismo funciona para explicar el comportamiento de la máquina. De tal manera que las potencialidades y los límites del paradigma informático tiene dos ámbitos: (a) como instrumento para entender la mente humana y (b) como herramienta para simular a la mente humana. Aquí de nuevo, aparece la idea de que la IA es tan solo una simulación de la mente humana.

En relación con los argumentos de Roger Penrose, resalta la idea de dicho autor sobre la oposición entre «el carácter mecánico o ciego de la ejecución de algoritmos y el carácter creativo, indeterminado y sujeto a errores, propio de la inteligencia humana (1989). Tal contraste lo lleva a negar que la inteligencia artificial sea realmente inteligente y que la inteligencia humana pueda estar basada en algoritmos» (Gutiérrez

[2006] 2011, 18). La crítica de Gutiérrez, a dicha distinción, reside sencillamente en explicar qué significa o qué es la creatividad de la inteligencia humana. Para él, Penrose ignora los logros y los avances de la IA, los cuales Gutiérrez considera que son creativos por parte de la IA. Sin embargo, Gutiérrez tampoco define la noción de *creatividad*. Por otra parte, conduce el argumento de la diferencia establecida por Penrose a su punto de vista, ya que indica que

desde luego que los programas de inteligencia artificial son ciegos y mecánicos en un sentido, puesto que los ejecuta una máquina; pero igualmente lo son el disparo de nuestras neuronas y su contacto a través de sinapsis, y sin embargo eso no nos hace menos inteligentes. (18)

Si bien reconoce que existe una diferencia entre la creatividad humana y la de la IA, en tanto que la IA todavía no llega a ser como la humana. Para él, «todavía no hemos terminado el arduo trabajo de diseñar algoritmos artificiales tan buenos como los que han forjado millones de años de selección natural» (19); así las deficiencias actuales de la IA son momentáneas, esto resalta el optimismo de Gutiérrez en torno a lograr una inteligencia artificial semejante a la humana.⁶

El algoritmo

En el libro *El humanismo replanteado. Genes y menes en la sociedad globalizada* ([2006] 2011), Gutiérrez desarrolla otros conceptos importantes como la *informática*, la *robótica* y la *inteligencia artificial*. Un concepto de interés es el de *algoritmo*. Tal como lo define Gutiérrez, y en razón de las distintas áreas y disciplinas a la que se aplica, tal parece que se trata de una constante de la realidad física y social. Él define *algoritmo* como «un procedimiento seguro y mecánico que logra un cierto resultado, especialmente en matemática y en informática» ([2006] 2011, 11). El concepto se transfiere de la matemática y la informática, y se aplica a otros campos como la economía, la ingeniería, la selección natural y

la lógica (14). Por otra parte, distingue entre los *viejos algoritmos* de la matemática, la lógica y la economía, los cuales se usaron antes de los años setenta, caracterizados porque solo explican la lógica de la acción; en contraste, con los *nuevos algoritmos*, los que surgen a partir de dicha década. Estos últimos son creados con la aparición de la computadora, lo novedoso, para él, es que no solo explican la lógica de una acción, sino que a la vez la ejecutan ellos mismos (14), estos son los que llama *algoritmos informáticos*; aunque si bien con calcos de la matemática, surgen con la característica de la autonomía que les daría la capacidad de cambiar, según él, lenta, pero de manera segura la estructura de la sociedad, la concepción de ser humano y del mundo. Para él, la proeza de estas máquinas también reside en:

(...) poner a las máquinas a realizar procesos racionales, antes reservados solo a los humanos, subraya más que nunca la extraordinaria capacidad de concebir y manejar símbolos, exclusiva de nuestra especie dentro del reino animal, como bien lo había categorizado Aristóteles. Pero por otro lado, el gran logro que significa una máquina capaz de realizar actividades simbólicas, nos apremiaba a una introspección inédita sobre la naturaleza y el alcance de nuestra «diferencias específica». En nuestra larga lucha por superar los atavismos ancestrales de nuestra especie, que nos asimilan a los animales más salvajes, percibíamos ahora que la racionalidad (...) nos asimila en algún sentido a las nuevas máquinas. ([2006] 2011, 14).

La última idea expresada en la cita es intrigante, ya que puede interpretarse como que la racionalidad humana nos asemeja a la máquina, como si la máquina fuese más perfecta que el ser humano. Mientras que el ser humano se encuentra atado a su carga atávica ancestral, la máquina es pura lógica, razonamiento y algoritmo capaz de ser eficiente en la solución de problemas. Esta idea parece más un sueño que una certeza. En todo caso, considera que el cuerpo humano es una máquina, la más delicada de todas, incluyendo el cerebro y todos sus procesos ([2006] 2011, 15). Esto para ejemplificar el impacto de

la redefinición de *ser humano*. Así los algoritmos «se habían soltado por el mundo y andaban transformando todas y cada una de las actividades humanas, sin excepción alguna, desde las más materiales hasta las más espirituales» (15). Se trata de una revolución recursiva que caracterizan la era de la computación.⁷

En *Epistemología e informática* ([1993] 2011), en el capítulo 2, titulado «Lógica e informática», asevera que los algoritmos son estructuras de razonamiento lógicos, y más precisamente, lo que él llama «líneas de prueba», es decir, «un grupo de premisas nos sirve para demostrar algo, las que, a su vez, junto con nuevas premisas, puede llevarnos a probar todavía otra conclusión» (119), tal línea de prueba es el algoritmo. Ésta manera de entender el algoritmo cumple con la definición aportada por Gutiérrez, ya que, se trata de un *procedimiento eficaz*, para llegar a los enunciados, los que suponemos creíbles (sin aportar su demostración) a la conclusión buscada. Lo último, porque para Gutiérrez el razonamiento también sirve para convencer, partiendo de lo que la persona sabe, para llevarla por etapas a algo que interese que llegue (119).

Los programas de computación son los mediadores entre las construcciones intelectuales (los algoritmos y las estructuras de datos) y las máquinas que realizan los respectivos y poderosos efectos computacionales. En otras palabras, los programas permiten el paso de la lógica y el razonamiento al programa que hace realidad esa lógica en las operaciones físicas de la computadora ([1993] 2011, 119-120). Para él, la *computadora* es una encarnación posible de la máquina abstracta universal propuesta por Turing. Esta máquina es modelo general de un algoritmo que es capaz de imitar a cualquier otro algoritmo, por que responde a instrucciones elementales (120)⁸. Así, los algoritmos (los razonamientos) son útiles para resolver problemas, esta es la capacidad de razonar, lo que significa —según Gutiérrez— producir y aplicar algoritmos; tal capacidad es la que ha dado ventaja al ser humano (121). El algoritmo es como una especie de nicho evolutivo del ser humano, al igual que la información, pues el ser humano procesa, produce e interpreta; además, la información es algo que se encuentra en su entorno. Con la informática, según Gutiérrez,

ahora deseamos poner esos algoritmos, podríamos decir «naturales», en forma de programas y en máquinas, lo artificial; en último término obedecible por una máquina, la tarea consistirá en desmenuzar el problema para expresarlo en términos del conjunto de estados del universo, según el discurso del que se trata (122).

La informática como ciencia

En el acápite II se plantea el asunto de la exploración de los límites y las posibilidades de la informática y la IA. Al respecto, Gutiérrez asevera que tales preocupaciones conllevan a definir la informática como una ciencia. Por una parte, (a) como una ciencia teórica, es decir, se trata de una teoría de la computabilidad, que determina qué cosas son imposibles, y explora los límites, así como las posibilidades de la disciplina a partir del análisis de la máquina de Turing, y la idea de la inteligencia. Por otro lado, (b) ve a la informática como una ciencia empírica, igual que cualquier otra ciencia, ya que según él, se trata de inventariar fenómenos computacionales, y formula una suerte de hipótesis sobre su naturaleza y características ([1993] 2011, 147); aunque, tal vez propiamente puede considerarse como una disciplina que desarrolla una teoría tecnológica sobre la computabilidad⁹; que señala cómo se puede construir tecnologías informáticas. Por otro lado, Gutiérrez asevera que para considerar a la informática como ciencia, él se vale de la filosofía de la ciencia de Popper, la que le proporciona el marco conceptual para la articulación de las ciencias del conocimiento. Para él, la informática es una ciencia como las otras, con hipótesis que pueden ser refutadas por la realidad, o falseados por el experimento; no obstante, la máquina de Turing no es falsa, pues es una noción abstracta, el algoritmo es una definición matemática, no hay experiencia posible para refutarla, es una disciplina analítica y a priori (147). Aunque no indica como se detectaría el error.

Si bien podría pensarse que la informática es más una ciencia empírica que teórica, el sostiene que no es así, que existe un continuo entre ambas. La teórica es valiosa porque dice qué

cosas son posibles o las que no se deben intentar, pero dicen muy poco de cómo son posibles ([1993] 2011, 147); aunque, en todo caso, si se considera la definición de la informática como ciencia, tal parece que se refiere a una noción de ciencia formal. De tal manera, que el método analítico de la teoría computacional determina los límites de la disciplina de manera a priori, una vez aclarados los límites, sigue la investigación empírica (147). Para Newell y Simon, esta ciencia tiene dos hipótesis: (a) hipótesis de los sistemas simbólicos y (b) hipótesis de búsqueda heurísticas (148-149). En todo caso, el método heurístico no es exclusivo la informática, sino que está presente en todas las ingenierías, entre otros. Además, es un método más creativo que estrictamente científico, para la solución creativa de problemas tecnológicos.

Los problemas sociales de la informática y la IA

En *El humanismo replanteado: genes y menes en la sociedad globalizada* ([2006] 2011), en el acápite «El algoritmo informático», Gutiérrez asevera que el algoritmo es objeto de estudio de la filosofía, en sus palabras:

el algoritmo puede dar materia para reflexión filosófica importante. A parte de los trabajos en matemática, que tienen gran interés filosófico en cuanto análisis de los límites del conocimiento -por ejemplo, el teorema de Gödel demuestra que las matemáticas son o incompletas o incongruentes, lo que nos infunde un sentido de humildad y de búsqueda permanente-, han surgido en nuestros días varios tópicos muy generales relacionados con los algoritmos de considerable contenido filosófico» (17).

Como se analizó, efectivamente, el estudio filosófico de la informática, la robótica y la IA tiene que ver con la pregunta de si es posible crear una inteligencia artificial; así como hacer un análisis de los argumentos a favor y contra de dicha posibilidad. Ahora, lo curioso es que Gutiérrez cita el argumento de la incompletitud

de Gödel para resaltar la humildad en el estudio de estos temas; sin embargo, dicho argumento se usa para mostrar las dificultades existentes para el desarrollo de la IA Fuerte. Por otra parte, la importancia de tal reflexión filosófica no se queda en los aspectos más teóricos y abstractos sobre la informática, la robótica y la IA, sino que también se dedica a los más prácticos. Se ha de recordar que en *El humanismo replanteado: genes y menes en la sociedad globalizada* ([2006] 2011), él señala que los algoritmos cambiarán la estructura de la sociedad, al ser humano y al mundo (15). A esta transformación la llama revolución, apogeo y proeza.

El libro *La sociedad computarizada* ([1990] 2011) es llamativo porque es la primera aproximación de Gutiérrez al tema de la informática y la IA en lo relativo a las implicaciones de dicha tecnología para la cultura, la sociedad y la economía. El problema que intenta responder en la *Prefacio de la versión original* del texto es que, en este «mundo de conocimientos y destrezas cada vez más especializados, ¿cómo podemos salvar la integridad de una visión humanista de la cultura y la unidad de acción del ciudadano responsable?» ([1990] 2011, 9). Con lo que refuerza su visión humanista presente en su texto del 2006, *El humanismo replanteado: genes y menes en la sociedad globalizada*, en su comparación con el libro de 1990, *La sociedad computarizada* en su edición de 2011 en las *Obras completas*.

Para el estudio filosófico de las implicaciones prácticas, Gutiérrez procede de manera analítica, descomponiendo la totalidad en sus partes para obtener un conocimiento más preciso, y trabaja cada una de ellas como si fuesen la totalidad, y suponemos que se sucesivamente hasta llegar al límite del análisis. Tal fragmentación no está en la realidad, sino que es un asunto de método, en donde cada parte se trabaja como si fuera una totalidad ([1990] 2022, 9). Por eso en *La sociedad computarizada* aparecen una serie de temas de la informática, la robótica y de la IA que tienen un impacto en diferentes ámbitos de lo social; esto sin que la especialización produzca una huida o falsificación de la realidad; ya que estamos en un mundo complejo, en donde hay interdependencia de disciplinas y factores que condicionan nuestro ambiente (9).

En su vocación docente, él considera que a pesar de esa fragmentación, el estudiantado tiene que aprender a trabajar de manera interdisciplinaria. Por lo cual, se ha de dotar a los estudiantes de una cultura científica (10), además, que ha de tomar decisiones con base en la ética. La informática es una disciplina que tiene sus raíces en la cultura, como en la lógica, la física y la filosofía, con repercusiones en otras disciplinas y con sus implicaciones humanistas (11).

El libro *La sociedad computarizada* ([1990] 2011), como se indicó en la introducción, es un libro de texto para un curso de la Universidad Estatal a Distancia (UNED-Costa Rica), el cual va acompañado de una antología, los temas en los que se descompone el impacto de la tecnología informática, la robótica y la IA en la sociedad, el cual se estructura en capítulos (ejes temáticos), y estos en subcapítulos (áreas concretas). Cada tema tiene sus objetivos de aprendizaje, una introducción teórica y un comentario a los textos de la antología. La edición del libro para las *Obras completas* se ha independizado el texto de la antología, por eso el autor incluye citas de los textos de la antología. En el libro se estudian 11 casos de impacto: la industria, el comercio, la oficina, el trabajo en el hogar (el trabajo masculino y femenino), la privacidad (derecho a la privacidad como un imperativo), el poder, el capitalismo, el Tercer Mundo (desarrollo y subdesarrollo), la educación, el pensamiento del ser humano y el pensamiento de la máquina, la ética desde la perspectiva deontológica: la responsabilidad de las computadoras, propiedad intelectual y libertad de pensamiento (programas libres), y las teorías éticas. Sobre estos temas se expondrán algunas reflexiones.

En las reflexiones sobre los temas específicos, Gutiérrez analiza los argumentos en contra y favor de las aplicaciones de la informática, la robótica y la IA a diferentes esferas de la sociedad, la economía y la cultura. Así, por ejemplo, analiza cómo estas tecnologías afectan a la industria, cómo la automatización aumentaría la productividad, pero reduciría el trabajo; por lo cual, no admite cualquier tipo de introducción de dichas tecnologías, sino que propone que habría que contar con planes para permitan el cambio. Si bien las tecnologías van a influir en el

comercio, la oficina, y el hogar; para él es claro que la innovación tecnológica provocará cambios en el comportamiento de la sociedad, cuya asimilación toma más tiempo, ya que exige a su vez cambios en la manera de percibir las cosas y las relaciones con las personas ([1990] 2011, 72). Él parece imaginar cómo la tecnología informática configura el futuro, por ejemplo, el trabajo virtual o el teletrabajo, lo cual se experimentó con la pandemia del COVID-19, en este sentido a pesar de las dificultades que Gutiérrez observa en la introducción a la informática y la inteligencia artificial en la sociedad, tal configuración de futuro es percibida con gran optimismo. En todo caso los problemas generados por la tecnología requieren de una reflexión.

Un aspecto en el que muestra una gran convicción, y una enfoque ético absoluto es en relación con la privacidad, de tal manera, que Gutiérrez enfatiza o asume o se propone como objetivo formular las razones para la defensa de la privacidad en el contexto de la democracia moderna ([1990] 2011, 84). Al mismo tiempo que resalta los peligros de la introducción de las computadoras en la vida social para la privacidad. Él tiene la convicción filosófica de que la vida privada debe ser respetada, aunque no es algo evidente en sí mismo es un imperativo (84). Para él, la protección de la privacidad es una base para la democracia, ello implica la protección de los bancos de información (85).

En relación con el poder, analiza los efectos de la informática y la IA, los cuales tienen que ver con las posibilidad de que este poder sea centralizado o descentralizado (92); los efectos en las campañas electorales, cita como ejemplo, varias campañas electorales en Costa Rica. Otro aspecto del poder que requiere una reflexión gira en torno al manejo de datos para la toma de decisiones, en especial aquellas decisiones que tienen que ver con las personas subordinadas, a los que están debajo de las jerarquías (93). Ahora, al respecto de este manejo de cómo la tecnología informática puede llevar a la concentración o no del poder, él cree en la neutralidad de la tecnología, ya que «no parece que las computadoras de por sí tiendan a favorecer la centralización o la concentración del poder. En realidad, estás

máquinas pueden ser usadas de muchas maneras y para diversos propósitos». (97)

Estos son algunos ejemplos de sus reflexiones filosóficas sobre algunos temas; pero en general, Gutiérrez considera la investigación en informática y la inteligencia artificial ha tendido muchos avances, pero que todavía falta mucho por hacer, esto en el contexto temporal de los años noventa, en sus palabras:

No obstante sus logros, dicha investigación ha servido hasta el momento más para revelar la magnitud del problema de crear una inteligencia artificial que para reproducir a ésta en un nivel satisfactorio. Muchísimo trabajo resta por hacer, y gran parte de ese proyecto depende de la disponibilidad de computadoras con mucho más memoria que las que actualmente se construyen». ([1990] 2011, 131)

La cita refiere a la investigación en inteligencia artificial con base en la máquina de von Neuman, la cual según Gutiérrez, mostró sus limitantes, por lo cual otros investigadores optaron por modelos conexionistas, los cuales sugieren que la inteligencia se ha de dar en forma espontánea a partir de configuraciones computacionales de gran conectividad, semejantes a las neuronas del sistema nervioso. En todo caso, en los años noventa, él considera que independiente de cualquier resultado, que cualquier tipo de investigación es una gran oportunidad a las ciencias del conocimiento para estudiar los fenómenos mentales ([1990] 2011, 133).

Conclusiones

Las reflexiones de Claudio Gutiérrez se encuentran en varios niveles, desde el análisis más teórico y desde los aspectos históricos del desarrollo de la informática y la IA, pasando por los aspectos técnicos y de programación, hasta la reflexión filosófica sobre a la posibilidad de crear una inteligencia artificial semejante a la humana. Por otra parte, se trata de una reflexión en el ámbito humanístico y ético, en donde propone una formación humanista y ética para las

personas especialista en informática, robótica e inteligencia artificial. Pero a pesar los problemas que genera la posibilidad misma de generar una inteligencias artificial igual que la humana, así como los problemas sociales que genera la introducción de las tecnologías informática en la sociedad, siempre mantuvo un fuerte optimismo acerca del desarrollo de dicha tecnología. Por otra parte, su convicción de que realmente se puede crear una inteligencia como la humana, resulta inconsistente, según como se ha mostrado, o al menos falta una mayor precisión a la hora de discutir los argumentos a favor y contra. De tal manera que a pesar de que argumenta a favor de una IA Fuerte, realmente, sus argumentos apuntan a una IA Débil.

Notas

1. El presente artículo es el resultado de una mesa redonda en homenaje póstumo al Dr. Claudio Gutiérrez Carranza, profesor de la Universidad de Costa Rica, destacado humanista, filósofo y divulgador de la informática y la Inteligencia Artificial (IA) en Costa Rica. La actividad se denominó *Mesa Redonda. Homenaje al Dr. Claudio Gutiérrez Carranza (1930-2023)*, y se realizó el día 04 de mayo de 2023, a las 15 horas de Costa Rica. Además, en la mesa redonda participaron el Dr. Luis Camacho Naranjo, con su ponencia *De la lógica a la epistemología*; el Lic. Guillermo Coronado Céspedes, con *Don Claudio*; la Dra. Marcela Ehandi con *Impacto de su pensamiento en mi formación. Reflexiones sobre un nuevo humanismo en el pensamiento de Claudio Gutiérrez*. La moderación correspondió a la Dra. Dina Espinosa Brilla.
2. Gutiérrez afirma que su crítica supone una teoría de Roger Schank sobre la comprensión del lenguaje por computadora —la cual se basa en la teoría de Marvin Minsky—, tal teoría propone la idea de marcos para programar a las computadoras en la comprensión del lenguaje ordinario. A la vez la teoría se basaba en la hipótesis de sistemas de símbolos físicos de Newell y Simon (véase págs. 260-261).
3. Al respecto -como lo veremos más adelante- Gutiérrez se posiciona en una defensa de la IA débil; aunque, su ataque al argumento de Searle parece más bien defender la IA fuerte. No

- obstante, como veremos, tal parece que Gutiérrez, asume, junto con Dennett, un enfoque que se parece a la IA débil.
4. Gutiérrez escribe la palabra intencional con «s», así «intencional» se refiere a la significación escolástica —introducida en la filosofía contemporánea por Franz Brentano—. De referencia a un objeto «contenida en todos los conceptos que expresan actitud de un sujeto a un objeto (tales como «creer», «desear», «temer», «querer», «conocer», «creer», etcétera). Para lo que sigue, vamos a llamar intencional a todo aquello que tiene la naturaleza de un deseo (positivo o negativo) o de una creencia» ([1993] 2011, 157). Ese uso de la palabra puede deberse a un error tipográfico.
 5. Una discusión interesante sobre cómo mostrar que alguien tuvo la intención de hacer algo, o a su capacidad de responder a la pregunta ¿por qué?, se encuentra en Elizabeth Anscombe ([1957]1991), en *Intención*.
 6. Para Gutiérrez la paradoja de la inteligencia humana y artificial se resuelve al concebir que «la estrategia de sus programas combina dos subalgoritmos, cada uno de los cuales realiza su función de manera mecánica. El primero de ellos se encarga de *generar*, de manera más o menos arbitraria, una serie de hipótesis que podrían ser solución al problema que enfrenta; por su parte, el otros se ocupan enseguida de *escoger*, de conformidad con una función heurística, aquella que resulte más promisoría para los objetivos del programa. Esa función es falible, aunque su aplicación sea totalmente determinista» ([2006] 2011, 18).
 7. Una lista de esos cambios es la siguiente: el procesador de texto transformó la práctica de la escritura; la hoja electrónica alivió el trabajo de los(as) contadores(as). Las bases de datos un enjambre de interconexiones lógicas; los juegos electrónicos y la realidad virtual, la transformación de las operaciones bancarias, y del comercio y la industria, la robotización de la industria automovilística, se creó la manufactura flexible, las oportunidades del trabajo desde la casa, la instrucción de la imagenología digital en su relación con el cuerpo humano. La secuenciación del genoma humana, la comunicación digital, entre otros cambios ([2006] 2011, 15-16).
 8. Según Gutiérrez, Turing dotó a esta máquina universal de un lenguaje sencillo, ya que tenía que demostrar que todo algoritmo podía expresarse en ese lenguaje, y para intentar mostrar que no podía existir un algoritmo que resuelva todos los problemas matemáticos ([1990] 2011, 120), que se concretan en lenguajes de programación, a los que llama *máquinas virtuales*.
 9. Para el autor del artículo, la informática más que una ciencia es una tecnología, la cual elabora una teoría tecnológica para explicar como dicha ingeniería procesa la información, construye algoritmo, e indica cuál es su naturaleza y características. El término *informática* es el equivalente a *teoría de la información*. El término proviene del francés *informatique*, implementado por el ingeniero Philippe Dreyfus a comienzos de la década de los años sesenta. La palabra es, a su vez, un acrónimo de *information y automatique*. En Castellano el término *informática* proviene de otros dos conceptos: *información y automática*; aunque en los textos en general se dice que es un «ciencia» que estudia el Tratamiento Automático de la Información. El concepto se usa para designar un conjunto de ciencias y técnicas que tienen por objeto la información y la comunicación. Este tratamiento automático es el que ha propiciado y facilitado la manipulación de grandes volúmenes de datos y la ejecución rápida de cálculos complejos, por medio de los sistemas informáticos o computacionales. Tales sistemas deben contar con la capacidad de cumplir tres tareas básicas: entrada (captación de la información), procesamiento y salida (transmisión de los resultados). El conjunto de estas tres tareas se conoce como algoritmo. Por eso, un tema principal es la estructura o posibilidad de los llamados servomecanismos, ordenadores y computadoras. Un aspecto importante es el proceso de diseñar y crear una aplicación para poder realizar tratamiento de textos, datos, sistemas, es una tarea informática. Además, estudia lo que los programas son capaces de hacer (teoría de la computabilidad), de la eficiencia de los algoritmos que se emplean (complejidad y algorítmica), de la organización y almacenamiento de datos (estructuras de datos, bases de datos) y de la comunicación entre programas, humanos y máquinas (interfaces de usuario, lenguajes de programación, procesadores de lenguajes), entre otras cosas.

Referencias

- Anscombe, Elizabeth. [1957] 1991. *Intención*. Barcelona: Paidós.
- Camacho, Luis. 2001. Introducción a las *Obras Completas*. En *Obras completas*. Vol. I. San José: Editorial UCR.
- Carvajal Villaplana, Álvaro. 2020. Máquinas y autómatas contribuciones al diseño de lo artificial en Descartes. En *Azur. Revista Centroamericana de Filosofía* 1, No. 1: 87-101.
- Gutiérrez, Claudio. [2006] 2011. *El humanismo replanteado: genes y menes en la sociedad globalizada*. En *Obras completas*. Vol. V. San José: Editorial UCR.
- Gutiérrez, Claudio. [1993] 2011. *Epistemología e informática*. En *Obras completas*. Vol. I. San José: Editorial UCR.
- Gutiérrez, Claudio y Castro, Marlene. [1990], 2011. *La sociedad computarizada*. Vol. III. San José: Editorial UCR.

Álvaro Carvajal Villaplana (alvaro.carvajal@ucr.ac.cr) Docente e investigador de la Escuela de Filosofía y la de Estudios Generales de la Universidad de Costa Rica. Docente e investigador de la Escuela de Ciencias Sociales del Instituto Tecnológico de Costa Rica.

Recibido: 30 de octubre, 2023.
Aprobado: 6 de noviembre, 2023.