El mito de lo mental: el proyecto de investigación de la inteligencia artificial y la transformación hermenéutica de la fenomenología (Primera parte)

Jethro Masís
Julius-Maximilians-Universität Würzburg

Resumen

En sus dos partes, este estudio pretende reconstruir con cierto detalle el fiasco del proyecto de investigación de la Inteligencia Artificial y las consecuencias de la crítica devastadora de ese proyecto llevada a cabo por Hubert Dreyfus en su magnum opus *What Computers Can't Do* (1972, 1979, 1992). Parte de estas consecuencias, es la emergencia de un grupo de investigadores en este campo especializado que se han llamado a sí mismos 'heideggerianos'. Esta autodefinición será criticada en la segunda parte de este estudio. En esta primera parte, se consideran (i) los sueños estrafalarios y las falsas expectativas de los investigadores en IA, y (ii) la propia crítica dreyfusiana a partir de supuestos de naturaleza fenomenológica.

Abstract

In its two parts, this study intends to reconstruct with some detail the fiasco of the Artificial Intelligence research project and the devastating critique carried out against it by Hubert Dreyfus in his magnum opus *What Computers Still Can't Do* (1972, 1979, 1992). Part of these consequences is the emergence within this specialized field of a group of scholars who have called themselves 'Heideggerian'. This definition shall be dealt with and criticized in the second part of this study. In this first part, we go on to analize (i) the bizarre dreams and false expectations of AI researchers, and (ii) Dreyfus' own criticism and its definitive thrust of phenomenological nature.

Palabras Clave: Inteligencia artificial, teoría de la subjetividad, filosofía de la mente, transformación hermenéutica de la fenomenología, Martin Heidegger.

Keywords: Artificial Intelligence, theory of subjectivity, philosophy of mind, hermeneutical transformation of phenomenology, Martin Heidegger.

El mito de lo mental: el proyecto de investigación de la inteligencia artificial y la transformación hermenéutica de la fenomenología (Primera parte)

Jethro Masís
Julius-Maximilians-Universität Würzburg

> "Los matemáticos quieren tratar los asuntos de la percepción matemáticamente y, con ello, se ridiculizan a sí mismos. […] La mente lo hace todo tácita, naturalmente, y sin reglas técnicas".
>
> **—PASCAL, *Pensées***

1. Antecedentes: Sueños Estrafalarios y Falsas Expectativas

Comencemos con unas lapidarias sentencias de Hubert Dreyfus que se encuentran a la altura del cierre de la introducción de su obra seminal, *What Computers Still Can't Do. A Critique of Artificial Reason* (1992):[1]

> Si estamos en el umbral de crear Inteligencia Artificial, estamos a punto de ser testigos del triunfo de una concepción muy especial de la razón. En efecto, si la razón puede ser programada en una computadora, esto confirmará una concepción del ser humano como objeto, detrás de la cual han andado a tientas los pensadores occidentales en los últimos dos mil años, pero que sólo ahora tienen las herramientas para expresar e implementar. La encarnación de esta intuición, cambiará drásticamente nuestra comprensión de nosotros mismos. Si, por otra parte, resultase que la Inteligencia Artificial es imposible, entonces nos veremos en la obligación de distinguir entre la razón humana y la artificial, y esto también cambiará radicalmente nuestra visión de nosotros mismos. Así que ha llegado el momento de, o bien enfrentar la verdad de la intuición más profunda de la tradición, o bien abandonar la explicación mecánica de la naturaleza humana que se ha ido desarrollando gradualmente en los últimos dos mil años (1992: 78-79).[2]

Se trata, a no dudarlo, de sentencias tajantes y —todo parecería indicarlo— de una encrucijada decisoria, puesto que el fracaso o éxito del proyecto de investigación de la Inteligencia Artificial (IA en lo sucesivo) pareciera dejarnos sopesar concluyente y definitivamente la concepción de la razón forjada en Occidente que se nutre en sus rasgos fundamentales de la suposición (y absolutización) de la objetividad. Además, se trata aquí, ni más ni menos, que de nuestra propia concepción respecto de nosotros mismos y de nuestro encuentro con las cosas. En suma, se trata de la interpretación de qué sea lo humano. Todo

[1] La edición revisada de The MIT Press, *What Computers Still Can't Do* (1992), es precedida por las ediciones de 1972 y 1979, intituladas de forma algo diferente: *What Computers Can't Do. The Limits of Artificial Reason*. El cambio del título mediante la agregación del adverbio *still*, quiere dar a entender que los 'resultados' centrales del libro *todavía* siguen manteniendo vigencia, puesto que el problema de la Inteligencia Artificial es de naturaleza ontológica.

[2] La traducción de las citas que provengan de textos en lengua extranjera, son exclusiva responsabilidad del autor. Se traducirán todos los textos citados en el cuerpo principal del texto. Las citas que se coloquen en el aparato de notas aclaratorias al pie de la página, sin embargo, se dejarán intactas en la lengua original.

apunta a que en este reto que lanza Dreyfus, es decir, el de enfrentar la verdad de la intuición más profunda de la tradición occidental, que no es más que el descubrimiento de la objetividad, se juega nuestra propia valoración de este triunfo de la tradición filosófica y científica; o bien por el contrario, en caso de que nos hallemos ante resultados anómalos respecto de nuestras grandes expectativas, lo que nos espera es llamar a cuentas a una concepción de las cosas y de nosotros mismos que, en el fondo, resulta insuficiente para explicar, precisamente, qué somos. Parece que, si Dreyfus está en lo cierto, nos encontramos en el umbral de comprobar los alcances genuinos y los límites del pensamiento objetivante, es decir, se trata de dimensionar, de colocar en un terrreno acotado, a la objetividad. Y por objetividad hemos de entender la concepción teorética que se propone estudiar, precisamente, objetos y que no concibe diferencias —a menos que sólo lo sean de grado— entre el ser humano y el ser de los objetos, y que, por ende, estudia todo lo que compete a lo humano (el espíritu, la conciencia, la subjetividad, la mente, el cuerpo, etc.) como otro objeto entre muchos. La tematización de la objetividad, la sola posibilidad de esta tematización, es el asunto que acá puede ser confirmado, o que puede ulteriormente disolverse en la nada.

Como se verá en lo que sigue, si la concepción de la objetividad resulta errática en el respecto comprensivo que se irá desarrollando en esta meditación (y que denominamos fenomenología hermenéutica), no sólo quedará demostrado que no puede llevarse a cabo un estudio objetivante del ser humano, sino que la misma noción de objeto aparecerá como un trasunto abstracto e inoriginario incluso cuando se aplica a las mismas cosas. La coseidad (*Dinglichkeit*), de tal forma, quedará cuestionada en un doble sentido y, con ello, desactivada en su eficacia explicativa. En este sentido, es la *Geganständlichkeit* de los *Objekte*, es decir, el carácter de cosa de los objetos, lo que en segundo lugar viene a cuestión. Se trata, entonces, de explicar también esta cosificación y sus poderes para trasuntarse y aparecer como independiente.

Solamente por lo antedicho, debería adivinarse de antemano que *What Computers Can't Do*, en efecto, levantó inmediatamente una serie de reacciones que, cuando menos, pueden ser caracterizadas de virulentas. Pero todo esto es comprensible, porque Dreyfus lanzó una bomba en el mismo campo de juego de los filósofos analíticos y atacó su más preciosa idea: la de comprobar, por medios tecnológicos, la adecuación del análisis filosófico y, junto con ello, la primacía de la abstracción logicista. En fin, no sólo se trataba de filosofar hablando y discutiendo, sino sobre todo de *realizar* la abstracción consecuente en un aparato tecnológico.

En este respecto, Seymour Papert, uno de los fundadores del laboratorio de IA del Instituto Tecnológico de Massachusetts (MIT, por las siglas en inglés), fundador del lenguaje de programación LOGO y un reputado científico computacional y matemático, puede ser considerado como el primero de todos los críticos de Dreyfus. En 1968, se abocó afanosamente a la redacción de un extenso memorando, de más de setenta páginas de extensión, acerca del primer escrito que dedicó Dreyfus a los avances de la IA intitulado, *The Artificial Intelligence of Hubert L. Dreyfus. A Budget of Fallacies* (cf. 1968). Según recuerda Dreyfus, su propio reporte para la Corporación RAND acerca del estado de la investigación, *Alchemy and Artificial Intelligence*, fue calificado de "siniestro, deshonesto, risiblemente gracioso y una increíble tergiversación de la historia" (1992: 87). En su reporte, intitulado quizá de forma insidiosa por su referencia a la alquimia, Dreyfus constató lo que le parecía ser un patrón en el desarrollo de la IA: éxito temprano acompañado, seguida e ineludiblemente, de signos inequívocos de estancamiento y degeneración (cf. 1965: 9-17). *Alchemy and Artificial Intelligence* esboza los temas que serán desarrollados en la obra más conocida de Dreyfus, pero lo que dio pie a una más bien sañuda recepción quizá fue la

sugerencia de que la forma en la que una máquina procesa información (*machine information processing*) difiere esencialmente de aquella que atañe a los agentes humanos (cf. 1965: 18-46), y de que ciertas concepciones falsas de lo mental por parte de los investigadores de la IA (como la *petitio principii* según la cual los seres humanos, de hecho, 'procesan información') enmascaraban —en vez de hacerle frente— la seriedad de las dificultades teoréticas que iban saliendo al paso de la investigación (cf. 1965: 46-64).

La inmediata reacción de Papert respecto del texto de Dreyfus se ensañó en primer lugar éticamente, señalando la supuesta falta de responsabilidad del reporte. Según Papert, los 'hechos' de Dreyfus "casi siempre son incorrectos y su noción de lo que sea la programación es tan pobre, que clasifica como imposibles programas que cualquier principiante podría desarrollar" (1968: 0-2).[3] Además, cada página del texto está llena de "sinsentido técnico" (*idem*), puesto que "Dreyfus comprende mal sistemáticamente incluso artículos elementales sobre el tema que pretende denunciar" (1968: 0-7), y su procedimiento denunciativo está viciado por la falta de integridad: "Gran parte del 'análisis penetrante' de Dreyfus (como lo ha llamado Oettinger) se genera a partir de seleccionar dificultades específicas señaladas por Simon y Newell como problemas *técnicos* para programas *particulares* y simplemente declararlos obstáculos *absolutos* para *todos los programas posibles*" (1968: 0-7). Papert se siente incluso facultado para reflexionar acerca del lugar de las humanidades en la academia:

> Es una cobardía el responder [a las demandas de un pensamiento humanista] llenando los departamentos de 'humanidades' con 'fenomenólogos' quienes nos aseguran que la computadora ha sido desnudada por sus estados numéricos finitos, y esto sin asediar ulteriormente las áreas de actividades que los fenomenólogos califican de 'estrictamente humanas' (1968: 0-2).[4]

En suma, de lo que carecería Dreyfus, según Papert, no es sólo de las competencias técnicas que lo capacitarían para ser un crítico serio y fructífero de la IA, sino sobre todo de integridad académica; juicio con pretensiones de rotundidad que le permite a Papert canturrear edificantemente sobre los fines de la cultura: "Nuestra cultura se halla, en efecto, en una situación desesperadamente crítica si sus valores han de ser defendidos de la deposición de la integridad académica por parte del pensamiento confuso (*muddled thinking*)" (*idem*). En concordancia con Papert, Dreyfus peca de una suerte de cripto-romanticismo teologal según el cual ni es deseable ni es factible simular tecnológicamente lo específicamente humano;[5] esto, puesto que "la combinación de una imagen exageradamente romántica de sí mismo con una imagen exageradamente simplificada de la compudora, hace que el neófito [en este caso, Dreyfus] se aterre ante la sugerencia de que un robot podría tomar un dictado tan bien como su secretaria" (1968: 0-6).

Ahora bien, son reacciones de este talante, por lo demás, las que tuvieron el efecto inmediato de incidir en que la honestidad intelectual de Dreyfus fuese cuestionada y que su trabajo profesional en la universidad se viera amenazado. Todo lo cual es comprensible, puesto que las opiniones de Dreyfus fueron, por lo demás, el principal detonante del llamado 'invierno de la IA' (*Artificial Intelligence winter*):

[3] Citamos el memorando de Papert según la paginación original, que se estructura según el guarismo de cada una de las secciones, seguido del número de página.
[4] Según Papert, "his [de Dreyfus] arguments must be read as literary conceits with deep 'humanist' content" (1968: 0-2).
[5] En este contexto, Papert denuncia los apegos de Dreyfus a la 'falacia super humana' (*superhuman fallacy*): "Inability to imagine the kind of formalism that could describe certain aspects of human behavior leads us to say that this behavior *cannot* in principle be encompassed by formal theories", 1968: 0-6.

el período de tiempo en que se redujo notablemente el presupuesto para investigación en IA después de la subsecuente frustración ante la manifiesta falta de resultados del proyecto;[6] período de tiempo, cabe añadir, que abarca desde los años setenta y que llega hasta nuestros días. La IA se ha convertido en una serie de sub-áreas muy especializadas pertenecientes principalmente a la robótica, pero el proyecto basado en la hipótesis fuerte de la IA ha sido abandonado.[7]

¿Cuáles fueron las expectativas que no llegaron a buen término? Nos haremos una buena idea de ellas si nos remitimos a las estrafalarias aserciones hechas por Herbert Simon y Allen Newell, ambos científicos pioneros del proyecto, en 1958:

> No tengo intenciones de sorprenderlos o conmocionarlos… Pero la forma más simple en la que puedo resumirlo es diciendo que ya hay en el mundo máquinas que piensan, que aprenden y que crean. No sólo eso sino que su habilidad para hacer estas cosas va a aumentar rápidamente en un futuro visible, hasta que el rango de problemas que pueden manejar las máquinas sea coextensivo con el rango de problemas de los que se encarga la mente humana (Simon & Newell, 1958: 8).

Que sepamos, aún hoy, es decir, cinco décadas después de las declaraciones de Simon y Newell, no contamos con máquinas que realmente piensen, ni que sean capaces de aprender o de crear (a menos que entretengamos ideas muy básicas de 'pensar', 'aprender' y 'crear', o que concibamos, con Hobbes, que "pensar no es más que calcular").[8] Las declaraciones públicas de Simon y Newell parecen, desde nuestro punto de vista actual, simplemente falsas y exageradas. Ahora bien, si este era supuestamente el estado de las máquinas inteligentes en 1958, es de suponer que las predicciones futuristas no serían menos extravagantes. En este famoso artículo que estamos citando y en que introducen la heurística (*heuristics*) como la disciplina teorética que permitiría simular los procesos de la resolución de problemas del nivel humano en computadoras digitales,[9] Simon y Newell hacen cuatro predicciones puntuales respecto de lo que sería esperable de las máquinas inteligentes en el futuro:

1. Que en diez años, una computadora digital será el campeón mundial de ajedrez, a

[6] El 'invierno de la IA' ha tenido varias etapas, entre las cuales pueden destacarse las siguientes: 1966: la aceptación definitiva del fracaso del proyecto de la traducción mecánica. 1970: el abandono del conexionismo (el modelamiento de los fenómenos mentales a partir de procesos emergentes de redes interconectadas de unidades más simples). 1971-1975: la frustración de DARPA (*Defense Advanced Research Projects Agency*) con el proyecto del reconocimiento del habla por parte de las computadoras (*computer speech recognition*). 1973: los efectos del reporte del profesor James Lighthill (cf. 1973), cuya prognosis negativa puso en suspenso la importancia de invertir en proyectos de IA en Inglaterra. 1973-1974: la suspensión del presupuesto para la investigación en IA por parte de DARPA. 1988: la cancelación de más presupuesto por parte de la *Strategic Computer Iniciative*. 1993: el decaímiento de las altas esperanzas depositadas en los *expert systems*. 1993 en adelante: la creciente mala reputación de la IA como 'ciencia'. Sobre el estado actual del invierno de la IA, cf. Hendler, 2008.

[7] 'Strong AI hypothesis' es la expresión acuñada con el fin de referir la expectativa fundante del proyecto: la de que, eventualmente, la acción inteligente general (*general intelligent action*) podría ser simulada en una máquina digital. Cf. al respecto, Goertzel & Pennachin (eds.) 2007.

[8] De acuerdo con con la afamada aserción de *Leviathan* (1651), razonar (o pensar) es calcular: "Out of which we may define (that is to say determine) that which is meant by this word *reason* when we reckon it amongst the faculties of the mind. For REASON, in this sense, is nothing but *reckoning* (that is, adding and substracting) of the consequences of general names agreed upon for the *marking* and *signifying* of our thoughts; I say *marking* them, when we reckon by ourselves; and *signifying*, when we demonstrate or approve our reckoning to other men" (2005: 34). Es, precisamente, Leibniz a quien se le atribuye la exhortación *calculemus!*, dirigida contra quienes dudasen de las bondades de su máquina de cálculo de cuatro funciones (al respecto, cf. Stein *et al*.: 2006).

[9] "[W]e have now the elements of a theory of heuristic (as contrasted with algorithmic) problem solving; and we can use this theory both to understand human heuristic processes and to simulate such processes with digital computers" (Simon & Newell, 1958: 6).

menos que las reglas la excluyesen de la competición.

2. Que en diez años, una computadora digital descubrirá y probabá un nuevo e importante teorema matemático.

3. Que en diez años, una computadora digital compondrá música que será aceptada por los críticos como de valor estético considerable.

4. Que en diez años, la mayoría de las teorías de la psicología tomarán la forma de programas computacionales, o de proposiciones cualitativas acerca de las características de los programas de las computadoras (Simon & Newell, 1958: 7-8).

Solamente la cuarta profecía llegó a cumplirse, y se llegó a instalar culturalmente en el 'sentido común'. Pero, ¿cuáles son las razones que yacen detrás de estas expectativas y profecías tan ilusorias? ¿Por qué se esperaba tanto del proyecto de investigación de la IA? No sólo se trataba de un optimismo epistemológico sin fundamento, sino sobre todo de la asunción sin cuestionamientos de una serie de suposiciones de raigambre filosófica. Concretamente, se trata de cuatro asunciones que subyacían al persistente optimismo de los investigadores en IA: respectivamente, las asunciones biológica, psicológica, epistemológica y ontológica.[10] Pero tendremos la oportunidad de ir desgranando cada una de estas asunciones indiscutidas a partir de una lectura de dos artículos clásicos sobre el tema de Simon & Newell: "Heuristic Problem Solving: The Next Advance in Operations Research" (1958) y "Computer Simulation of Human Thinking and Problem Solving" (1962).

La propuesta programática de la heurísica (*heuristics*) fue escrita conjuntamente con Newell, pero leída y presentada ante la *Operations Research Society of America* por Herbert Simon en 1958. Simon, laureado con el Premio Nobel en 1978 y entrenado en los campos de la economía, que le mereció el mentado galardón, y las ciencias políticas (disciplinas que él mismo llama 'suaves' —*soft fields*— si se las compara con los 'logros superiores' de las ciencias naturales, 1958: 2), ha sido un personaje muy influyente en el estudio sociológico de las organizaciones empresariales.[11] Y es que el criterio con que parecen anunciarse los avances de la investigación operacional (*operations research*), dentro de la cual vendría a jugar un papel progresivo y decisivo la heurística, no deja de ser empresarial y ulteriormente industrial. Simon inserta históricamente su propio proyecto en conjunción con los intentos mecanicistas decimonónicos de Charles Babbage (1791-1871), a quien se atribuye el haber "sin lugar a dudas… comprendido e inventado la computadora digital" (1958: 3). Simon también señala la idea de Gaspard de Prony (1755-1839) de una producción masiva de las tablas matemáticas, la cual sugirió a Babbage "que la maquinaria podría reemplazar el trabajo humano en las fases de oficinista de la faena, y que lo impulsaron a la empresa de diseñar y construir una máquina automática de calcular" (*idem*). Estamos ante la invención de un mecanismo matemático que podría resolver problemas humanos, porque estos criterios industriales y administrativos conducen a Simon a concluir que un problema bien estructurado (*well-structured*) a contrapelo de uno mal estructurado (*ill-structured*), debería satisfacer los siguientes criterios:

1. Puede ser descriptible en términos de variables numéricas, de cantidades vectoriales y a nivel de escala.

[10] Estas asunciones son presentadas y discutidas detalladamente en Dreyfus 1992, 153-227.
[11] Algunas reflexiones sobre racionalidad económica y planeamiento social, se encuentran en Simon, 1996.

> 2. Las metas que se persiguen pueden ser especificadas en términos de una función objetiva bien definida (por ejemplo, la máximización de la ganancia o la minimización de los costes).
>
> 3. Existen rutinas computacionales (algorítmicas) que permiten que la solución sea hallada y atestada en términos numéricos actuales. Ejemplos comunes de tales algoritmos, que ya han jugado un papel importante en la investigación operacional, son los procesos de maximización en el cálculo y el cálculo de las variaciones, la programación lineal de algoritmos como los métodos *simplex* y *stepping-stone*, las técnicas de Monte Carlo, etc. (1958: 4-5).[12]

Todo esto quiere decir que "los problemas bien estructurados son aquellos que pueden ser formulados explícita y cuantitativamente, y que pueden ser resueltos por técnicas computacionales conocidas y factibles" (1958: 5). Simon incluso no deja de mostrar su vergüenza ante las innumerables situaciones en que las variables utilizadas para solucionar un problema no son numéricas, "sino simbólicas o verbales", y esto quiere decir finalmente "vagas y no cuantitativas" (*idem*). Por ello, la conclusión es esperable: "[H]ay muchos problemas prácticos —sería más adecuado decir que 'la mayoría de los problemas prácticos'— para los cuales simplemente no existen algoritmos computacionales" (*idem*). Lo peor del caso, según la argumentación de Simon y Newell, es que estos problemas mal planteados que aún se escapan al dominio del cálculo y de la cuantificación, terminan por ser exclusivamente parte de la provincia del mero juicio y de la intuición, es decir, resultan ser más un asunto de una corazonada que del cálculo (*a matter of hunch than of calculation*, cf. *idem*).

Es en este contexto (en los sueños de la administración empresarial y de los procesos industriales de la total calculabilidad), en que la heurística hace su aparición como el método que, en principio, permitiría deshacer el embrujo de todas estas inexactitudes que no permiten reducir el comportamiento humano bajo la cuantificación y la calculabilidad total. La heurística vendría a ser la 'mecánica del juicio práctico' (una suerte de mecánica para lo social) y, de tal forma, quedarían resueltos los problemas metodológicos que se cifraron desde el siglo XIX en las vicisitudes de las ciencias humanas por alcanzar un estatuto cabal de cientificidad. Estaríamos, así, a las puertas de encontrar, en esta mismísima conferencia de Simon y Newell, la solución a los problemas que turbaron en antaño a Wilhelm Dilthey (pero esto, desde luego, sin reconocer la distinción diltheyana entre las ciencias naturales y las del espíritu).[13]

Pero, ¿en qué consiste exactamente el método de la heurística anunciado tan pomposamente por Simon y Newell? Básicamente, en una teoría de la resolución de problemas que puede utilizarse para dar

[12] El algoritmo *simplex*, desarrollado por el matemático estadounidense George Dantzig (1914-2005), consiste en una serie de soluciones numéricas para la programación lineal. El *simplex* es un politoto de $N + 1$ vértices en N dimensiones: un segmento de línea sobre una línea, un triángulo sobre un plano, un tetraedro en un espacio de tres dimensiones y así sucesivamente. Por otra parte, el *stepping-stone* (StSt), es un tipo de medida de seguridad computacional que consiste en colocar sistemas de seguridad lógicos, utilizados como servidores de autentificación, en una disposición serial que emula un estrecho canal físico, análogo al camino físico formado por una sucesión de piedras en un río que servirían para cruzarlo. Finalmente, los métodos de Monte Carlo, se utilizan en matemática financiera para evaluar y analizar las inversiones mediante la simulación de la incertidumbre que hipotéticamente afectaría dichas finanzas y que terminaría por determinar el valor promedio por encima del rango de los resultados alcanzados. Al respecto, cf. Anderson *et al*., 2008; principalmente el capítulo 17 y ss.

[13] He tratado extensamente las aporías decimonónicas del historicismo y las contradicciones de Dilthey, el fundador programático de las ciencias del espíritu (*Geisteswisseschaften*), en otro trabajo. Cf. Masís, 2009.

cuenta de "los procesos heurísticos de la comprensión humana y para simular dichos procesos por medio de computadoras digitales. La intuición, el tacto comprensivo [*insight*] y el aprendizaje no son más posesiones exclusivas de los seres humanos: cualquier computadora de gran velocidad puede ser programada también para exhibir estas capacidades" (Simon & Newell, 1958: 6). El programa heurístico, que es el resultado de las investigaciones llevadas a cabo durante los años cincuenta y sesenta por Simon, Newell y J. C. Shaw para la Corporación RAND, otorgaría a una computadora "la habilidad de descubrir pruebas para los teoremas matemáticos; no de verificar pruebas, debe notarse, puesto que un simple algoritmo puede diseñarse para eso, sino de llevar a cabo las actividades 'creativas' e 'intuitivas' propias de un científico que busca la prueba de un teorema" (Simon & Newell, 1958: 7). Ulteriormente, es decir, en las etapas más avanzadas que se barruntaban del método heurístico, se cumplirían los sueños del pensamiento occidental, porque "[l]a investigación en la resolución heurística de los problemas será aplicada a la comprensión de la mente humana. Con la ayuda de los programas heurísticos, ayudaremos al ser humano a obedecer el mandamiento antiguo: Conócete a ti mismo. Y conociéndose, podrá aprender a utilizar los avances del conocimiento para beneficiar, en vez de destruir, a la especie humana" (Simon & Newell, 1958: 8).

Como se desprende de lo antedicho, los investigadores pioneros de la IA no trabajaban simplemente para conseguir resultados parciales o meramente ingenieriles. Se trataba, ni más ni menos, que de simular la inteligencia humana y todos sus atributos, incluso de reproducir artificialmente la mente y, por ello, no es que haya acá un entrometimiento de la metafísica en la ingeniería (como había sugerido Papert, cf. 1968), sino que, como dice bien Germán Vargas, el proyecto es de suyo ontológico: "La cosa misma de la que se ocupa un interés fenomenológico por la IA no es el conjunto de los 'mecanismos', sino la *esencia de la subjetividad protooperante*" (2004: 106), es decir, el interés fenomenológico que motiva una mirada filosófica a estos problemas es la pretensión de un grupo de investigadores de convertir el racionalismo moderno en un proyecto de investigación tecnológica. La conclusión claramente cientificista (y comercial) de Simon y Newell habla a favor de esta interpretación: "Cuando las máquinas tengan mentes, podremos crear copias de estas mentes de la forma tan barata como hoy en día se imprimen libros" (1958: 9). Según nuestra forma de interpretar la historia de la IA, 'strong AI' es una caracterización tardía del proyecto que asume que el designio original por programar la subjetividad protooperante tuvo que confesarse fallido. Pero es que no se trataba de cualquier gazapo que fuera superable mediante el mejoramiento de las teorías respecto de los mecanismos, sino que con sólo echar un vistazo a la concepción que se entretenía de la mente, es dable pensar que se trataba —si se me permite decirlo con García Márquez— de una 'crónica de una muerte anunciada'.

En "Computer Simulation of Human Thinking and Problem Solving" (1962), Simon y Newell presentaron un sistema metódico que llamaron GPS: *General Problem Solver*. En este escrito programático, comenzaron anunciando lo que les parecía una obviedad y que solamente un ludita se atrevería a disputar:

> Ya no es necesario argumentar que las computadoras pueden utilizarse para simular el pensamiento humano o para explicar en términos generales cómo dicha simulación puede llevarse a cabo. Más de una docena de programas computacionales, que han sido escritos y probados, llevan a cabo algo de las tareas interesantes de la resolución de problemas y de la manipulación simbólica que los seres humanos pueden realizar, y lo hacen en una forma que simula, al menos en los aspectos más generales, el modo en que los seres humanos cumplen con estas tareas. Los programas de las computadoras juegan

> ahora ajedrez y tablero, encuentran pruebas para teoremas en geometría y lógica, memorizan sílabas sin sentido, forman conceptos y aprenden a leer (1962: 137).

En concordancia con sus creadores, el GPS es "un sistema de métodos… que resulta bastante útil en muchas situaciones en que una persona enfrenta problemas para los cuales no posee métodos especiales de resolución" (1962: 138). De tal forma, cuando una persona tiene que enfrentar un problema, sigue estrictamente una sucesión reglamentaria inconsciente, un mecanismo que está funcionando cada vez que dilucida alternativas para la resolución de una dificultad. La heurística, al menos a nivel hipotético, permitiría simular la reglamentación algorítmica de ese proceso resolutivo que se suele denominar 'pensamiento'.

La asunción básica que supone la viabilidad del GPS es de raigambre filosófica: habría, en principio, una estructura subagencial de cómo piensan los seres humanos constituida principalmente por un sistema lógico de reglas. Pero esta es —puede decírselo sin ambages— la concepción analítica de la filosofía. El filósofo oxoniense Peter Strawson, uno de los pensadores analíticos más eminentes del siglo XX, lo afirma explícitamente: "[A]sí como el gramático… trabaja para elaborar una explicación sistemática del sistema de reglas que observamos sin ningún esfuerzo cuando hablamos gramaticalmente, el filósofo [analítico] lo hace para conseguir una explicación sistemática de la estructura conceptual general de la que nuestra práctica diaria muestra que tenemos un dominio tácito e inconsciente" (1997: 50). Los investigadores de la IA estaban trabajando como filósofos analíticos, pero ahora no sólo discutiendo y dilucidando teorías, sino poniendo en marcha un proyecto de investigación que concretaba tecnológicamente mediante métodos ingenieriles la suposición filosófica de que, en efecto, se poseía algo así como una comprensión de esa estructura sistemática y conceptual del nivel subagencial del pensamiento y del obrar comprensivo humano. Ahora bien, lo tácito y lo inconsciente es aquello que, mediante la asignación de ciertos algoritmos a una máquina, podría simular los procesos heurísticos de que estaría supuestamente constituida la mente humana. Esto también está supuesto en el GPS y, si bien Simon y Newell lo afirman en relación con la ingeniería computacional, repiten la misma concepción strawsoniana de la existencia de una estructura reglamentada del pensamiento:

> En tanto teoría de la resolución humana de los problemas, el GPS sostiene que los estudiantes universitarios resuelven problemas… llevando a cabo esa suerte de análisis organizado en términos de medios y de fines. No sostiene que este proceso se lleve a cabo de manera consciente, pues es fácil mostrar que muchos pasos en el proceso de la resolución de problemas no alcanzan a llegar a lo consciente [*do not reach conscious awareness*]. Tampoco sostiene la teoría que el proceso parecerá particularmente ordenado para un observador que no conozca el programa en detalle o el mismo solucionador de problemas. Sí sostiene que, si comparamos esa parte que podemos observar en el comportamiento del sujeto humano cuando soluciona problemas (los pasos que sigue, sus verbalizaciones) con el proceso llevado a cabo por la computadora, serán substancialmente los mismos (1962: 141).

En esto, tanto Simon y Newell como Strawson, parecen estar de acuerdo: hay un proceso inconsciente, una perspectiva de tercera persona y, por ende, subagencial, que sostiene todo nuestro afrontamiento[14] consciente. El punto es demostrar que la IA, de hecho, puede coadyuvar a poner en

[14] En este trabajo, el uso técnico del término 'afrontamiento' obedece a un intento por traducir el término inglés 'coping'. La expresión verbal 'to cope with' significa básicamente arreglárselas, habérselas conductual, corporal y comprensivamente (esto es,

práctica esta asunción, puesto que si la suposición se probase cierta aunque fuese a un nivel de escala reducido, es decir, si realmente fuese cierto que los seres humanos utilizan procesos heurísticos para la resolución inteligente de las tareas problemáticas y que estos pueden ser programados tecnológicamente, entonces la investigación sucesiva podrá felicitarse eventualmente de la esperanza, ahora asegurada para la posteridad, de alcanzar resultados cada vez más alentadores y exitosos.

Esta era, al menos, la misma esperanza de aquellos estudiantes pertenecientes al Laboratorio de IA del MIT, que se acercaron al curso sobre Heidegger que Dreyfus impartía en el primer lustro de los años sesenta, para señalarle las falencias de la filosofía: "Ustedes los filósofos han estado reflexionando desde sus sillones en los últimos dos mil años y aún no comprenden la inteligencia. Nosotros hemos asumido el control desde al laboratorio de IA y estamos teniendo éxito en todo aquello en que los filósofos han fallado" (2007: 247).

No obstante, la respuesta de Dreyfus es categórica:

> Pero en 1963, cuando fui invitado a evaluar el trabajo de Allen Newell y de Herbert Simon sobre los sistemas físico-simbólicos, me encontré con la sorpresa de que, lejos de reemplazar a la filosofía, estos investigadores pioneros habían aprendido bastante, directa e indirectamente, de la filosofía: a saber, de la convicción de Hobbes de que razonar es calcular, de las representaciones mentales de Descartes, de la idea de Leibniz de una 'característica universal' (un entramado de rasgos primitivos en que todo el conocimiento podía ser expresado), de la concepción de Kant de que los conceptos son reglas, de la formalización de Frege de tales reglas, y del postulado de Wittgenstein de átomos lógicos en el *Tractatus*. Por decirlo de forma resumida: sin darse cuenta de ello, los investigadores de la IA estaban trabajando afanosamente en convertir la filosofía racionalista en un programa de investigación.
>
> Pero comencé a sospechar que la ideas formuladas desde instancias existencialistas, especialmente de Heidegger y de Merleau-Ponty, eran malas noticias para los investigadores de la IA: que mediante la combinación de representacionalismo, conceptualismo, formalismo y atomismo lógico en un programa de investigación, los investigadores de la IA habían condenado su empresa a la recreación de un fracaso rotundo (2007: 247-248).

Es cayendo en la cuenta del tiempo en que Dreyfus hace sus primeras y acertadas denuncias, hace ya casi cuarenta años, que su apuesta se nos muestra como un total atrevimiento. ¿No debería haber esperado unas cuantas décadas con el fin de vislumbrar el desarrollo de máquinas tanto más potentes y desarrolladas? ¿No es, precisamente, en la rapidez vertiginosa con que se desarrolla la tecnología que, las más de las veces, nuestras propias predicciones se nos muestran de poquísimo alcance? La estocada certera, incluso intempestiva, de Dreyfus radica en haberse dado cuenta de que lo que no marchaba con el proyecto, y la razón por la cual los orondos investigadores habían condenado su empresa al fracaso (*para siempre*, podría decirse incluso sin exagerar), era una determinada concepción filosófica abstracta,

de forma *encarnada* y prácticamente *situada*) en medio de un contexto determinado en el que se está envuelto de manera ejecutiva y práxica. En cuanto *terminus technicus*, afrontamiento significa, por tanto, el despliegue práctico del agente humano en su vivencia en el mundo, al lado de la comprensión que de suyo acompaña este afrontamiento.

plagada de supuestos logicistas y metodizantes; era también una concepción analítica, que partía de la hipótesis implausible, pero ahora por fin simulable en un mecanismo tecnológico, según la cual mediante la asignación de un listado de predicados sobre una serie de hechos atómicos, podría reproducirse la totalidad del sentido, puesto que el mundo no era sino una colección de objetos con propiedades y los procesos resultantes de estos objetos fijos y presentes. Supuestos, por cierto, que ya décadas atrás, habían sido desmantelados hasta los añicos por la fenomenología: una forma de practicar la filosofía que, muy tardíamente y no sin dificultades comprensivas básicas, pudo paulatinamente hacer su lugar en los países anglosajones.[15] Dreyfus ha sido tan importante en la introducción de la fenomenología en Estados Unidos, que no sólo ha incidido en que ciertos científicos cognitivos hayan cambiado su estimación hacia la filosofía europea (de cariz fenomenológico-hermenéutico), sino que seguramente a él también se debe el que el Laboratorio de IA del MIT se haya convertido reconocidamente 'heideggeriano' en orientación.

2. El Fiasco Señalado

Pero antes de ese viraje fenomenológico, autorreconocido incluso como 'heideggeriano', tuvo que haber acontecido un fiasco en esa investigación por simular la mente que con tan ostentosas predicciones se exornaba.

En efecto, Marvin Minsky, el codirector del laboratorio de IA del MIT de 1959 a 1974, aprendió bastante del GPS y del método heurístico de Simon y Newell, y "estaba convencido de que representando unos cuantos millones de hechos sobre objetos, incluyendo sus funciones, resolvería lo que llegó a llamarse el problema del conocimiento del sentido común" (Dreyfus, 2007: 248). John McCarthy, quien acuño la expresión *Artificial Intelligence* en 1956, escribió al respecto un reporte para el MIT intitulado curiosamente: *Programs with Common Sense*, y vislumbraba la posibilidad de construir un programa (llamado *Advice Taker*) que manipulara proposiciones instrumentales comunes en un lenguaje formal adecuado (cf. 1958). Empero, el primer obstáculo para simular de forma tecnológica la mente, no fue sino, precisamente, el sentido común, aunque se tratase del más común de los sentidos. Aunque, tal como advirtiera Voltaire, *el sentido común no es nada común*, los investigadores de la IA estaban convencidos de que cualquier dificultad debería resolverse descubriendo las artimañas funcionales de un complejísimo mecanismo natural. Pero con esto hacían caso omiso de aquel decir de Goethe, de que *la inteligencia y el sentido común se abren paso con pocos artificios*.

Cuando en una entrevista en la revista *Wired*, se le preguntó a Minsky por qué un nombre tan representativo de la IA como el suyo la había declarado cerebralmente muerta desde los años setenta, Minsky respondió: "No hay una computadora que tenga sentido común".[16] Este reconocimiento se retrotrae al fracaso resolutivo del problema del marco (*frame problem*).

El problema del marco atañe a la relevancia contextual en una situación dada y, en cuanto tal, tiene que ver con la atención fenomenológica. A una máquina se le programan proposiciones de objetos y

[15] El lugar de la fenomenología en los Departamentos de Filosofía de las principales universidades estadounidenses e inglesas, sin embargo, no puede decirse que esté definitivamente asegurado. Para la filosofía analítica que domina esos departamentos académicos, la fenomenología con su apelación a la intuición o a un ámbito donde no cabe la cuantificación, no es más que el 'intuicionismo filosófico' que, según Mario Bunge, "terminó, pues, convirtiéndose en una filosofía de los perversos para los irracionales" (2005: 56). Este texto que citamos, *Intuición y Razón* de Bunge es, por cierto, una excelente muestra de *lo que no es* la intuición en sentido fenomenológico.

[16] "Why AI is Brain Dead". Entrevista con Marvin Minsky. *Wired*. Issue 11, 08 Agosto de 2003.

de funciones, y esto supone que el mundo es algo así como un entramado fijo de objetos funcionales, respecto del cual nuestra conciencia estaría 'al tanto'. Al 'saber' o al intentarlo —y este 'saber' (entrecomillado) en este contexto significa: haciendo retrorreferencia a su almacenamiento recursivo de conceptos y representando simbólicamente el 'mundo exterior'—, una máquina se encuentra violentamente, en primera instancia, con el problema de la relevancia del contexto. El 'mundo exterior', representado desde los recursos simbólicos asignados al programa computacional, no parece hallar parangón con el mundo humano, donde los cambios plásticos del entorno son constantes y variables, es decir, donde al parecer no hay objetos fijos con funciones permanentes, y donde el sentido de las acciones no recae quizá sobre las funciones prefijadas de una colección de objetos constituidos. El problema del marco, ha sido definido por Michael Wheeler en *Reconstructing the Cognitive World* (2005) de la siguiente forma:

> Dando por supuesto que el mundo es dinámico y cambiante, ¿cómo puede un sistema que no es mágico… dar cuenta de aquellos estados cambiantes del mundo… que importan, y de aquellos estados no cambiantes del mundo que importan, siempre que ignore aquellos que no importan? ¿Y cómo puede un sistema semejante reparar y (si es necesario) revisar, a partir de todas las creencias que posee, sólo aquellas creencias que son relevantes en un contexto particular de acción? (2005: 179).

El mundo, según Wheeler, es dinámico y cambiante, y si no se trata de un mundo mágico —como parece sugerirlo nuestro conocimiento actual de la biología, la física y la química— debe haber una cierta lógica y una cierta reglamentación nómica que dé cuenta satisfactoriamente del sentido del mundo humano. Ante este problema temprano, pero que terminó dando al traste con todo el proyecto, la solución de Minsky fue sugerir que se utilizaran descripciones de situaciones vitales típicas, entre las que podrían contarse las fiestas de cumpleaños, las clases de filosofía, el ordenar un platillo en un restaurante, el departir en una reunión con amigos, etc. Es claro que la programación del predicado 'blanco' en el programa, podría ser cabalmente incomprensible si no se justifican los diversos contextos vitales, o marcos, en que el predicado no significa lo mismo: por ejemplo, cuando decimos 'blanco' de alguna persona de raza caucásica no lo hacemos en el mismo sentido que cuando lo predicamos de la pizarra acrílica que está en el aula, o cuando decimos, verbigracia, que la pared es blanca. Pero incluso con esta salvedad, al solo reconocimiento de un marco de relevancia contextual le sería menester el agregarle otra serie de marcos menores para la sola desambiguación de la terminología y de sus matices, de forma que parecía obvio, tal como señaló Dreyfus, "que cualquier programa de IA que utilizara marcos quedaría atrapado en un regreso a marcos para reconocer marcos relevantes que reconocieran hechos relevantes y, por ende, el problema del almacenamiento del conocimiento del sentido común y de la actualización de la información no era solamente un *problema*; era un signo inequívoco de que algo andaba seriamente mal con todo el enfoque" (2007: 248).

Esta fase del proyecto de investigación, en que por primera vez asoma el nefando semblante el problema del conocimiento del sentido común y en que se propone el esquema de los marcos contextuales, ha sido denominado por John Haugeland *Good old fashioned AI* (o GOFAI, por las siglas en inglés, cf. 1996). La 'IA pasada de moda' —como quizá podríamos traducir la expresión inglesa de Haugeland— consistió en una serie de fracasos sucesivos que, sin embargo, eran excusados bajo la consigna de falta de mayor investigación y de la necesidad de contar con máquinas con un mayor potencial de almacenamiento y de memoria.

Empero, en su *magnum opus*, Dreyfus puso en evidencia que las suposiciones de las cuales partían los investigadores pioneros en IA eran simplemente falsas e insostenibles; si no es que todo el modelo filosófico subyacente. Acto seguido, es menester desnudar las asunciones (2.1) biológica, (2.2) psicológica (2.3), epistemológica y (2.4) ontológica, que privaban de su realización el vaticinio exagerado con que Minsky abría su libro *Computation: Finite and Inifite Machines*: "en una generación… pocas áreas del intelecto permanecerán fuera del rango de la máquina. El problema de crear IA estará substancialmente resuelto" (1967: 2).

(2.1) La Asunción Biológica

La asunción biológica adquiere la forma de una teoría del procesamiento de la información (cf. Minsky, 1969) y supone que el cerebro humano está estructurado organizativamente como una computadora. Se trata del "supuesto ingenuo de que el ser humano es un ejemplo clarísimo de un exitoso programa de computación digital" (Dreyfus, 1992: 159). Desafortunadamente, esta imagen funcionalista del cerebro no ha pasado a formar parte del museo de las asunciones filosóficas fallidas y ampliamente descartadas, sino que incluso goza del estatuto de ser una de las ideas claves del proyecto de la ciencia cognitiva.[17]

Es bajo este supuesto, desde luego, que algo así como un programa heurístico de la resolución de problemas cobra pleno sentido. El cerebro humano sería, así, un artefacto simbólico-manipulativo de propósito general (*general-purpose symbol-manipulating device*) que 'operaría' tal cual lo hace una computadora digital, es decir, acatando los dictámenes metódicos de una reglamentación discreta. Si se lograse, correlativamente, simular procesos de inteligencia simples en computadoras digitales, cabría esperar el desarrollo de modelos más complejos, cada vez más semejantes a los supuestos procesos cognitivos de los seres humanos. Lo curioso, naturalmente, es que los investigadores no propusieran una pesquisa teorética *ex hypothesi* respecto de este modelo digital aplicado a la inteligencia humana con el fin de comprobar su efectiva plausibilidad, sino que lo asumieran desde el principio sin ningún cuestionamiento, como si fuera simplemente obvio e indubitable. Pero bastaba con que el cerebro 'procesara información' de otra forma (o, dicho más radicalmente, que no 'procesara' nada de nada), para que esta asunción de raigambre informática (nótese que no biológica) se hiciera pedazos.

Y eso fue lo que, de hecho, sucedió. Ya en 1966, en un artículo sobre cibernética y el cerebro humano, Walter Rosenblith del MIT confesaba que "[l]as comparaciones detalladas de la organización de los sistemas computacionales y la de los cerebros resulta igualmente frustrante e inconclusa" ("On Cybernetics and the Human Brain". *The American Scholar*, 274. Citado por Dreyfus, 1992: 162). Dreyfus concluye contundentemente:

> No pueden sacarse argumentos acerca de la posibilidad de la IA a partir de la evidencia empírica actual acerca del cerebro. De hecho, la diferencia entre la naturaleza 'altamente interactiva' del cerebro y el carácter no interactivo de la organización de la máquina sugiere que, en tanto sean relevantes los argumentos tomados de la biología, la evidencia está en contra de la posibilidad de utilizar computadoras digitales para producir inteligencia (*idem*).

[17] Las asunciones de que la mente es una suerte de sistema que procesa información, lo mismo que un artificio representacional y, en cierta medida, una computadora, son aún hoy en día ampliamente aceptadas. Cf. el prefacio de los editores al *Blackwell Companion to Cognitive Science*. Bechtel & Graham (eds.), 1999.

(2.2) La Asunción Psicológica

La pregunta que habría que traer a colación es si efectivamente el modelo cibernético puede justificar el uso de la hipótesis de la computadora digital en psicología. La psicología reinvindica para sí la prerrogativa explicativa de un nivel particular de funcionamiento de la inteligencia humana, que, sin negar las explicaciones físico-químicas del cerebro, puede dar cuenta de ciertos procesos mentales. La mente, según este supuesto, lleva a cabo tareas computacionales como comparar, clasificar, revisar listas de datos, etc., y sería el resultado de estos procesos lo que llamamos 'inteligencia'.

La asunción de que las teorías psicológicas adquirirían ineludiblemente el aspecto de programas computacionales, ya había sido vaticinada por Simon. Y Simon no se quedó sentado esperando que esto sucediera, sino que él mismo acometió la tarea de diseñar una serie de programas de computación que consistían en simular los pasos conscientes e inconscientes pretendidamente seguidos por una persona cuando realiza competencias cognitivas. Los entusiastas de la hipótesis psicológica asumían que cuando los seres humanos se comportan inteligentemente, se someten al seguimiento de una serie de reglas heurísticas similares a aquellas que le permiten a una computadora digital funcionar. Para Ulrich Neisser, el autor de una *Cognitive Psychology*, "la tarea del psicólogo que desea comprender la cognición humana es análoga con la de quien intenta descubrir cómo una computadora ha sido programada" (1967: 6). Por ello, la tarea del psicólogo cognitivo en la era de la cibernética sería el descubrir 'el programa' de la mente humana.

La asunción psicológica se ha valido de una metáfora, pero que ha sido asumida como un hecho constatado, que casi se ha convertido en una verdad de perogrullo para quienes vivimos en la era cibernética: la mente humana sería un mecanismo que 'procesa información'. Empero, tanto la forma de 'procesamiento' como el significado de 'información' son aquí términos bastante ambiguos. En primer lugar, aunque fuera cierto que hay algo así como 'procesos cerebrales', de ello no se sigue el que éstos consistan en estar programados, es decir, que sigan una serie de operaciones discretas determinadas de antemano. Si no hay operaciones discretas, todo el proyecto de la simulación cognitiva se viene abajo. En este caso, Dreyfus señala una falacia: aquella que consiste en "saltar del hecho de que el cerebro transforma de alguna forma sus entradas [*inputs*] a la conclusión de que la mente realiza alguna secuencia de operaciones discretas" (1992: 166). Pero ni siquiera es claro qué significa 'información' en este contexto cognitivo. Claude Shannon, el ingeniero y matemático que es recordado como el padre de la teoría de la información, ya había advertido que todos los aspectos semánticos de la información no formaban parte del problema estrictamente ingenieril de la comunicación (cf. 1948). Esto quiere decir que desde el punto de vista técnico de la ingeniería, 'información' no significa 'información con sentido' o 'información significativa'. En concordancia con Warren Weaver, incluso "dos mensajes, uno de los cuales esté cargado con mucho significado, y otro que sea un puro sinsentido, pueden ser exactamente equivalentes desde el presente punto de vista en lo tocante a la información. Esto es, sin duda, lo que quiere decir Shannon cuando afirma que los aspectos semánticos de la comunicación son irrelevantes para los aspectos ingenieriles" (citado por Dreyfus, 1992: 165).

Lo que ha acontecido acá, según Dreyfus, es una transformación ilegítima de la teoría matemática de la comunicación en una teoría del significado, a la que se añade la suposición de que la experiencia puede ser analizada en alternativas atómicas y aisladas (cf. *idem*). El hecho de que, en efecto, cualquier actividad física puede ser en principio formalizada lógica y algorítmicamente y, en

consecuencia, manipulada en una serie de operaciones discretas en una computadora digital, no justifica el que se asuma que la mente sigue esas operaciones. Para Dreyfus, tal cosa es tan absurda como creer que "los planetas están resolviendo necesariamente ecuaciones diferenciales cuando permanecen en sus órbitas alrededor del sol, o que la regla de cálculo (una computadora analógica) sigue los mismos pasos cuando calcula una raíz cuadrada que los que sigue la computadora digital cuando utiliza el sistema binario para calcular el mismo guarismo" (1992: 167). Todo esto significa que, si bien todos los procesos psicoquímicos pueden ser formalizados y calculados discretamente, ello en absoluto implica que haya procesos discretos subyacentes a todas las actividades involucradas en las competencias cognitivas.

(2.3) La Asunción Epistemológica

Aunque pareciera que ya debería ser claro que el modelo de la computadora digital (con la suposición del seguimiento heurístico de operaciones discretas) no provee ninguna clarividencia especial para dar cuenta del funcionamiento de la inteligencia humana, sino que se trata de un enfoque más bien torpe y discreto (en el doble sentido de este término), aún quedaba el subterfugio de la formalización. De éste, se sustenta básicamente la asunción epistemológica, según la cual el comportamiento humano podría ser, con todo, formalizable y simulado en un programa de computación. En este lugar, también podemos traer a colación, con Dreyfus, el ejemplo de los planetas: "No están resolviendo ecuaciones diferenciales mientras le dan la vuelta al sol. No están siguiendo ninguna serie de reglas; pero su comportamiento es, no obstante, legal, y para comprender su comportamiento, echamos mano de un formalismo —en este caso, de ecuaciones diferenciales— que expresa su comportamiento como movimiento *en concordancia con* una regla" (1992: 189). La asunción epistemológica sostiene, así, básicamente dos tesis: (i) Que todo comportamiento no arbitrario puede ser formalizado, y (ii) que el formalismo puede ser utilizado para reproducir el comportamiento en cuestión (cf. Dreyfus, 1992: 190). Dreyfus se desembaraza de estas dos tesis sosteniendo, a contrapelo, que la pretensión de la formalización total involucra una generalización injustificada del éxito de la física. Por lo demás, una teoría de la competencia no debe confundirse con una teoría de la acción.

En su clásico artículo de hace más de medio siglo, "Computer Machinery and Intelligence", en que pretendía dar respuesta a la pregunta *Can machines think?*, Alan Turing definió la computadora como un mecanismo "que debe seguir reglas fijas" (1950: 436), con lo cual quedaba limitada a un apego ajustado a ciertos datos carentes de ambigüedad y a reglas estrictas que se aplicarían inequívocamente sobre esos datos. Se suponía que una máquina de Turing, en cuyo funcionamiento se expresa la esencia de la computadora digital, podría llevar a cabo cualquier tarea realizable por un ser humano. Así lo confesaba Minsky: "No hay razón para suponer que las máquinas tengan limitaciones de las que carecen los seres humanos" (1967: vii). Minsky consideraba el artículo de Turing como una gran contribución a sus propias ambiciones por haber despejado dudas neurálgicas y haber refutado objeciones peregrinas. Una de estas objeciones tenía que ver, de hecho, con la imposibilidad de formalizar todo el comportamiento humano, y Turing la enfrentó haciendo una distinción entre las 'leyes de la conducta' y las 'leyes del comportamiento'.

El argumento de la informalidad del comportamiento, tal como lo denominó Turing, duda de que la *eventualidad* del comportamiento humano sea el resultado del acatamiento de reglas estrictas. Pero Turing repara que las reglas de la conducta, como los preceptos, no son reglas del comportamiento: todas aquellas leyes de la naturaleza aplicables universalmente al comportamiento humano, como, por ejemplo, "si lo pellizcas, gritará". En concordancia con Turing, "no sólo creemos que es verdad que el ser

regulado por leyes del comportamiento implica ser algún tipo de máquina… sino contrariamente, que ser tal máquina implica el ser regulado por tales leyes" (1950: 452). Dreyfus cree que, en este punto, Turing presumiblemente generaliza "el argumento de Wittgenstein de que es imposible proveer reglas normativas que prescriban por adelantado el uso correcto de un término en todas las situaciones" (1992: 192). El argumento de Turing, en suma, funcionaría más o menos de esta forma: si bien no podemos formular las reglas normativas para la correcta aplicación de un predicado particular, esto no demuestra definitivamente que no podamos formular las reglas que describan cómo, de hecho, un individuo particular aplica un predicado semejante (cf. Dreyfus, *idem*). O, puesto de otra forma: "Aunque Turing está dispuesto a admitir que podría resultar imposible proveer una serie de reglas que describan lo que una persona *debería hacer* en todas las circunstancias… no hay razón para dudar de que se podría en principio descubrir una serie de reglas que describan lo que esa persona *haría*" (1992: 193). Pero no parece haber tampoco ninguna razón que nos obligue a creer que, incluso si existiesen las leyes en cuestión, estas podrían formalizarse en una computadora digital.

El argumento ciertamente quiere sustentarse en la ambigüedad de la misma expresión 'leyes del comportamiento' que podría referirse, por un lado, a las acciones humanas significativas o a los movimientos físicos del organismo humano. Dado que los cuerpos humanos pertenecen al mundo descrito por la física, debería suponerse que, en tanto objetos físicos, obedecen cierto a comportamiento regular que, en cuanto tal, podría ser formalizable tal como la trayectoria de cualquier proyectil o como la caída de los objetos. Sin embargo, la idea de Turing y Minsky de la inteligencia de las máquinas no afirma que éstas resuelven ecuaciones físicas, sino que procesan datos que representan hechos sobre el mundo mediante el recurso a operaciones lógicas. La IA, en esta temprana fase como GOFAI, postula la existencia de una mente que manipula datos que representan el mundo y no pretende resolver las ecuaciones físicas que describen los objetos físicos. De hecho, utilizar las leyes de la física para calcular detalladamente el movimiento de los cuerpos podría ser físicamente imposible. Según el 'principio del límite' de Bremermann, "ningún sistema procesador de datos, artificial o viviente, puede procesar más de 2×10^{47} bits por secundo por cada gramo de su masa" (1962: 93). Esto quiere decir que, dado que hay $\pi \times 10^{7}$ segundos en un año, que la edad de la tierra es más o menos 10^{9} años, y que su masa está constituida por menos de 6×10^{27} gramos, incluso una computadora del tamaño del planeta tierra no podría procesar más de 10^{93} bits durante un tiempo igual a la edad de la tierra. Si Bremermann está en lo cierto, "las dificultades en los campos del reconocimiento de patrones y en la prueba de teoremas no serán resueltas por el mero aumento de la velocidad del procesamiento de datos por parte de las futuras supercomputadoras" (1962: 94). Pero aún así, "los investigadores en el campo de la IA, de Turing a Minsky, parecen refugiarse en esta confusión entre las leyes físicas y las reglas del procesamiento de la información para convencerse a sí mismos de que hay razón en suponer que el comportamiento humano puede ser formalizado, y que el peso de la prueba recae sobre aquellos que creen que hay procesos que no pueden ser descritos en un lenguaje formal" (Dreyfus, 1992: 197).

(2.4) La Asunción Ontológica

En concordancia con la asunción ontológica, se supone que "todo lo esencial para el comportamiento inteligente puede ser comprendido en términos de una serie de elementos independientes y determinados" (Dreyfus, 1992: 206). Habría, así, una actividad sintética de la mente que, a partir del reconocimiento de esos rasgos independientes, formaría la totalidad del significado. El sentido sería el resultado de la suma de estos elementos factuales objetivos y de sus funciones. Programar una máquina con sentido principia con la asignación a su órgano de almacenaje de todos los

objetos del mundo y de todas las funciones y relaciones de esos objetos.[18] Porque, ¿qué más puede haber qué objetos, funciones, procesos y relaciones entre esos objetos y esas funciones? Con todo, cabría preguntar ¿qué pasaría si hubiese otras determinaciones, no presentes ni objetivas, que fuesen *lo determinante* en los aspectos más importantes del sentido? ¿Qué pasaría si más bien lo que le da sentido al mundo es lo ausente, lo que no está nunca presente como tal en las cosas, sino sólo lo latente y lo sugerido, es decir, lo que no se anuncia? ¿Y qué pasaría si el sentido fuese más posibilidad que actualidad, si fuese algo más que lo objetivo y actual?

En la refutación de esta asunción ontológica —cuya indubitabilidad deriva de dos mil años de metafísica occidental junto con una mala interpretación del éxito de la física— Dreyfus comienza a mostrar su semblante más filosófico, porque no es cierto que a nivel fenomenológico, es decir, que al nivel de nuestra experiencia vital, las cosas se presenten como objetos o como entes determinados:

> Hasta un silla no es comprensible en términos de cualquier serie de hechos o de 'elementos del conocimiento'. Reconocer una cosa *como* una silla, por ejemplo, significa comprender su relación con otras cosas y con los seres humanos, lo cual implica un contexto humano total de actividad, en donde la forma del cuerpo, la institución de la mueblería y la inevitabilidad de la fatiga, constituyen solamente una pequeña parte de ese entramado (Dreyfus, 1992: 210).

Lo que hay en el pensamiento objetivante y calculador, y en la suposición heurística del seguimiento de reglas discretas en que confiaban los investigadores pioneros en IA, no es más que lo que el joven Heidegger había llamado desmundanización (*Ent-weltlichung*), privación de vida (*Ent-lebung*), deshistorización (*Ent-geschichtlichung*) y designificación (*Ent-deutung*) (cf. GA 56/57). El mundo humano tiene sentido, y el sentido es precisamente lo que no ha sido posible programar a causa de las asunciones filosóficas expuestas anteriormente. Pero el mundo 'mundea' (*es weltet*); frase acuñada en un ejercicio de imaginación que Heidegger, en su primera lección universitaria en Friburgo en 1919, lanzó a sus estudiantes para que comprobasen por sí mismos el proceso de privación de vida (*Ent-lebung*) a que se somete toda experiencia originaria en la objetivación teorética. Se trata de la experiencia cotidiana de Heidegger como profesor de llegar al aula y ver la cátedra desde la que impartirá sus lecciones:

> [E]ntro al aula y veo la cátedra… ¿Qué 'veo'? ¿Superficies marrones que se cortan en ángulo recto? No, veo otra cosa. ¿Veo acaso una caja, más exactamente, una caja pequeña colocada encima de otra más grande? De ningún modo. Yo veo la cátedra desde la que debo hablar, ustedes ven la cátedra desde la cual se les habla, en la que yo he hablado ya. En la vivencia pura no se da ningún nexo de fundamentación, como suele decirse. Esto es, no es que yo vea primero superficies marrones que se entrecortan, y que luego se me presentan como caja, después como pupitre, y más tarde como pupitre académico, como cátedra, de tal manera que yo pegara en la caja las propiedades de la cátedra como si se tratara de una etiqueta. Todo esto es una interpretación mala y tergiversada, un cambio de dirección en la pura mirada al interior de la vivencia. Yo veo la

[18] En concordancia con Turing, una computadora digital consiste en tres partes: "(i) Store. (ii) Executive unit. (iii) Control. " (1950: 437). La unidad de almacenamiento (*store*), guarda información, mientras que la unidad ejecutiva (*executive unit*) realiza las operaciones individuales involucradas en el cálculo. Por último, la unidad de control (*control*), se ocupa de que las instrucciones sean cumplidas al pie de la letra y en el orden adecuado. Cf. Turing, 1950: 437-439.

> cátedra de golpe, por así decirlo; no la veo aislada, yo veo el pupitre como si fuera demasiado alto para mí. Yo veo un libro sobre el pupitre, como algo que inmediatamente me molesta (un libro, y no un número de hojas estratificadas y salpicadas de manchas negras) (GA 56/57: 71).
>
> En la vivencia de ver la cátedra se *me* da algo desde un entorno inmediato. Este mundo que nos circunda… no consta de cosas con un determinado contenido de significación, de objetos a los que además se añada el que hayan de significar esto y lo otro, sino que lo significativo es lo primario, se me da inmediatamente, sin ningún rodeo intelectual que pase por la captación de una cosa. Al vivir en un mundo circundante, me encuentro siempre rodeado de significados por doquier, todo es mundano, *mundea* [*es weltet*] (GA 56/57: 72-73).

No se trata, desde luego, de volvernos místicos y de buscar en consecuencia una "exhalación nebulosa de turbios 'sentimientos del mundo' [*Weltgefühlen*] que, además, se presentan tan pomposamente y actúan detrás de la luz" (GA 61: 101), tal como ya lo advirtió Heidegger en sus lecciones fenomenológicas sobre Aristóteles. No es el caso de una singularidad inmediata, de lo vivido puro tal que, si lo afirmamos, terminamos por disolverlo.[19] No es el no-saber, inconcebible e inefable, del que habla Jean-Hippolite cuando explica a Hegel (cf. 1996: 13-33). Contrariamente, no es de otro mundo del que hablamos. El mundo de los objetos determinados, cuyas funciones son siempre pasibles de formalización lógica… ese sí que es otro mundo. Pero no es el mundo humano, tal y como se nos da primariamente en la experiencia del sentido.

En la segunda parte de este estudio, veremos la influencia de estas ideas fenomenológicas en algunos investigadores de la IA que han querido desarrollar programas 'heideggerianos'.

Bibliografía

Anderson, David *et al*. (2008) *An Introduction to Management Science. Quantitative Approaches to Decision Making*. Ohio: Thomson South-Western.

Bechtel, William & George George (eds.): *A Blackwell Companion to Cognitive Science*. New York: Blackwell, 1999.

Bremermann, Hans-Joachim (1962) "Optimization Through Evolution and Recombination". En: *Self-Organizing Systems*. Ed. M.C. Yovitts *et al*.Washington, D.C.: Spartan Books, 93–106.

Bunge, Mario (2005) *Intuición y Razón*. Buenos Aires: Editorial Sudamericana.

Dreyfus, Hubert (1965) *Alchemy and Artificial Intelligence*. RAND Papers. P-3244

_____. (1992) *What Computers Still Can't Do*. Cambridge, MA: The MIT Press.

_____. (2007) "Why Heideggerian AI Failed and how Fixing it would Require making it more Heideggerian". *Philosophical Psychology*. Vol. 20 No. 2, 247-268.

Goertzel, Ben & Cassio Pennachin (eds.) (2007) *Artificial General Intelligence*. Berlin – Heidelberg – New York: Springer.

Haugeland, John (1996) *Artificial Intelligence. The Very Idea*. Cambridge, MA/London: The MIT Press.

[19] El *esto* de la certeza sensible e inmediata es, para Hegel, la mayor de las trivialidades: "Como un universal *enunciamos* también lo sensible" (1994: 65). En ese acto de la enunciación, se disuelve su pretendida consistencia singular. Es la inevitabilidad de la mediación que comienza con el lenguage, que "es lo más verdadero" (*idem*), la que nos prohibe tener acceso al puro ser inmediato de la certeza sensible (cf. Hegel, 1994: 63 ss).

Hegel, G. W. F. (1994) *Fenomenología del Espíritu*. Trad. W. Roces. México: Fondo de Cultura Económica.

Heidegger, Martin (GA 56/57) *Zur Bestimmung der Philosophie*. [KNS 1919]. Gesamtausgabe Bde. 56/57. Ed. B. Heimbüchel. Frankfurt am Main: Vittorio Klostermann. 1987. [Versión castellana de una de las tres lecciones: *La Idea de la Filosofía y el Problema de la Concepción del Mundo*. Trad. J. A. Escudero. Barcelona: Herder. 2005].

_____. (GA 61) *Phänomenologische Interpretationen zu Aristoteles. Einführung in die phänomenologische Forschung*. [WS 1921-1922]. Gesamtausgabe Bd. 61. Ed. W. Bröcker & K. Bröcker-Oltmanns. Frankfurt am Main: Vittorio Klostermann. [1985] ²1994.

Hendler, James (2008) "Avoiding Another AI Winter". *IEEE Intelligent Systems*. Vol. 23, No. 2, pp. 1-4.

Hobbes, Thomas (2005) *Leviathan (Parts I and II)*. Ed. A. P. Martinich. Ontario: Broadview Press.

Hyppolite, Jean (1996) *Lógica y Existencia*. Trad. L. Medrano. Barcelona: Herder.

Lighthill, James (1973) "Artificial Intelligence: A General Survey". *Artificial Intelligence: A Paper Symposium*. Science Research Council.

Masís, Jethro (2009) "De la Vida Histórica: Auge y Aporías del Historicismo Decimonónico". *Konvergencias. Filosofía y Culturas en Diálogo*. Vol. VII, No. 21, 208-250. URL: <http://www.konvergencias.net/jmasis252.pdf>.

McCarthy, John (1958) *Artificial Intelligence Project: Programs with Common Sense*. AI Memo. No. 17. Cambridge, MA: MIT Artificial Intelligence Laboratory.

Minsky, Marvin (1967) *Computation: Finite and Infinite Machines*. New Jersey: Prentice-Hall.

_____. (ed.) (1969) *Semantic Information Processing*. Cambridge, MA: The MIT Press.

Minsky, Marvin & Seymour Papert (1970) *Proposal to ARPA for Research on Artificial Intelligence at MIT (1970-1971)*. AI Memo. No. 185. Cambridge, MA: MIT Artificial Intelligence Laboratory.

_____. (1971) *Proposal to ARPA for Research on Artificial Intelligence at MIT (1971-1972)*. AI Memo. No. 245. Cambridge, MA: MIT Artificial Intelligence Laboratory.

_____. (1972) *Artificial Intelligence. Progress Report*. AI Memo. No. 252. Cambridge, MA: MIT Artificial Intelligence Laboratory.

Neisser, Ulrich (1967) *Cognitive Psychology*. New York: Appleton-Century-Crofts.

Papert, Seymour (1968) *The Artificial Intelligence of Hubert L. Dreyfus. A Budget of Fallacies*. AI Memo. No. 154. Cambridge, MA: MIT Artificial Intelligence Laboratory.

Shannon, Claude (1948) "A Mathematical Theory of Communication". *Bell Sytem Technical Journal*. Vol. 27, 379-423.

Simon, Herbert & Allen Newell (1958) "Heuristic Problem Solving: The Next Advance in Operations Research". *Operations Research*. Vol. 6, No. 1, 1-10.

_____. (1962) "Computer Simulation of Human Thinking and Problem Solving". *Monographs of the Society for Research in Children Development*. Vol. 27, No. 2, 137-150.

Stein, Erwin *et. al*. (2006) "Calculemus! New Research Results and Functional Models Regarding Leibniz' Four Functional Calculating Machine and the Binary Calculating Machine". *Foundations of Civil and Environmental Engineering*. No. 7, 319-332.

Strawson, Peter (1997) *Análisis y Metafísica.Una Introducción a la Filosofía*. Trad. N. Guasch. Barcelona – Buenos Aires – México: Paidós.

Turing, Alan (1950) "Computing Machinery and Intelligence". *Mind*. Vol. LIX, No. 236, 433-460.

Vargas Guillén, Germán (2004) "Psicología y Fenomenología Trascendentales en el Proyecto de la Inteligencia Artificial". *Revista de Filosofía de la Universidad de Costa Rica*. Vol. XLII, No. 106-107, 105-118.

Wheeler, Michael (2005) *Reconstructing the Cognitive World: The Next Step*. Cambridge: The MIT Press.