

## ¿La inteligencia artificial como mejora cognitiva?: de los Sistemas de apoyo a la decisión (DSSs) a las *Reflection machines*

ZAIDA ESPINOSA ZÁRATE\*<sup>1</sup>

Universidad Loyola Andalucía (España)

zespিনosa@uloyola.es

### Resumen

El presente trabajo analiza si los Sistemas de apoyo a la decisión (DSSs) y otros asistentes para su uso, como las *Reflection machines* o los *Personal Assistants that Learn* (PAL), contribuyen de hecho a una mejora cognitiva, como habitualmente se tiende a asumir. Es decir, se examina si su potencial para expandir e impulsar la acción de las facultades cognoscitivas se ve efectivamente actualizado y, en consecuencia, si sirven para reafirmar el sentido capacitante de la IA y la extensión de la agencia del sujeto a través de ella. Para ello, se pone de manifiesto que el diseño y aplicación de estos sistemas están mediados por una comprensión del conocimiento que actúa a modo de marco que limita sus posibilidades de mejorar las capacidades cognoscitivas humanas. Se analizan las condiciones en que la mejora cognitiva es posible a través del uso de estos sistemas, que remiten a una aproximación integradora que no tiene en cuenta solo el factor tecnológico, sino que repara en cómo, en su uso, intervienen factores gnoseológicos, así como psicológicos y sociológicos, que median en la capacidad efectiva de estos sistemas de ampliar las facultades de conocimiento, y de los que las tecnologías se sirven para ejercer su influencia.

**Palabras claves:** Sistemas de apoyo a la decisión (DSSs), inteligencia artificial, mejora cognitiva, conocimiento, máquinas de reflexión.

---

\* Doctora en Filosofía. Profesora Contratada Doctora en el Departamento de Humanidades y Filosofía de la Universidad Loyola Andalucía (España). Sus artículos recientes incluyen: ¿Es el lenguaje tecnología? Análisis de un argumento transhumanista (2022, *Acta Philosophica*), Problem-Based Service Learning (PB-SL): Constructing a pedagogy of poverty based on Ignacio Ellacuría (2022, *Educational Philosophy and Theory*), El holismo cosmoteándrico de Raimon Panikkar y el holismo sustancial de Fernando Inciarte. ¿Dos metafísicas de sistema? (2020, *Areté. Revista de Filosofía*), Ciencia y fe, hacia un pensamiento integral: la propuesta de Raimon Panikkar (2019, *Veritas. Revista de Filosofía y Teología*), Voluntad de saber en el tiempo de la posverdad (2019, *Educación XXI*). Es autora del libro: *Percepción, pensamiento y lenguaje. El realismo radical de Reinhardt Grossmann y el realismo moderado de Fernando Inciarte* (2015, EUNSA).

## *Artificial intelligence as cognitive enhancement? From Decision Support Systems (DSSs) to Reflection machines*

### **Abstract**

*The present article questions whether Decision Support Systems (DSSs) and other assistants for their use, such as Reflection machines or Personal Assistants that Learn (PAL), do in fact contribute to cognitive enhancement, as it is usually taken for granted. In other words, it is analysed whether their potential to effectively expand or augment cognitive capacities is actualised and, therefore, if the capabilities and the agency of the person appear in fact extended through artificial intelligence. For that purpose, it is highlighted that the design and implementation of AI systems are dependent on a certain understanding of human knowledge that acts as a framework that limits their actual capacity to enhance the cognitive faculties of the mind. The conditions under which this cognitive enhancement is possible through these systems are analysed and require an integrative approach that not only takes technology into account but realises that, in its use, gnoseological, as well as psychological and sociological factors are involved and mediate in the actual capacity of these systems to extend cognitive functions. These factors are also crucial for these technologies to exert their influence.*

**Keywords:** *Decision Support Systems, artificial intelligence, cognitive enhancement, knowledge, reflection machines.*

### INTRODUCCIÓN

Muchos han dedicado su tiempo a lograr estrategias de pensamiento más efectivas y eficaces, ante la constatación de que las decisiones que tomamos son muchas veces precipitadas o simplemente erróneas, y nuestras habilidades de pensamiento son deficientes, especialmente las de orden superior (Swartz, Costa, Beyer, Reagan & Kallick, 2008). Las consecuencias de esta debilidad de pensamiento son de gran alcance, como se ha podido comprobar en los resultados contrastantes -en términos humanos, económicos y sociales- a los que la gestión de la pandemia Covid-19 ha conducido en los diferentes países (Kristjánsson, Fowers, Darnell & Pollard, 2021).

La promesa de la era digital de aumentar radicalmente nuestra capacidad cognoscitiva se ha visto frustrada, al reconocer que “su potencial de mejora cognitiva no es tan claro como originalmente se asumía”, sino que, frente a lo inicialmente esperado, “el mero acceso a una vasta cantidad de información no es suficiente para dar lugar a una mejora cognitiva robusta” (Voinea, Vica, Mihailov & Savulescu, 2020: 2358-2359) e incluso puede ser contraproducente: “Puede conducir a una reducción seria del rendimiento cognitivo del usuario” (Voinea et al., 2020: 2358-2359).

También las neurotecnologías se han implementado con el propósito de mejorar las capacidades humanas de conocimiento a través del registro y la estimulación de la actividad cerebral, tanto a través de procedimientos invasivos como no invasivos. En concreto, se han empleado para extender funciones cognitivas como la memoria, la atención, la percepción y comprensión de una situación (*situation awareness*), la comunicación, la toma de decisiones individual y colectiva o la resolución de problemas, como Cinel, Valeriani y Poli (2019) recogen en su revisión del estado del arte de estas tecnologías de la neurociencia. Sin embargo, se reconoce que su uso no terapéutico o curativo -es decir, referido a individuos sanos para la mejora cognitiva-, está todavía infrarrepresentado y hasta ahora no ha sido suficientemente relevante.

El intento por mejorar las capacidades cognoscitivas y asistir al ser humano en el proceso de toma de decisión se ha procurado también a través de las ciencias de la computación, y ha dado lugar a la implementación de sistemas de Inteligencia Artificial (IA) como los Sistemas de Apoyo a la Decisión (*Decision Support Systems* o DSSs, por sus siglas en inglés), con el propósito de llevar a un nivel superior la sabiduría práctica (*phronesis*) de la que es capaz el ser humano por sus propios medios. Como parte crucial para ella, la mejora cognitiva, entendida como la “ampliación o extensión de capacidades centrales de la mente mediante la mejora o el aumento de sistemas internos o externos de procesamiento de información” (Bostrom & Sandberg, 2009: 311), se ha perseguido a través del uso de potentes tecnologías de procesamiento de datos. Sin embargo, la implementación de estos sistemas ha traído consigo otros riesgos, dando lugar a la situación paradójica de que, para obtener mejores resultados en términos de decisión, se ha acabado empobreciendo la capacidad cognoscitiva humana, al generar aquellos sistemas una excesiva dependencia de las máquinas.

El mismo patrón se ha identificado respecto a Internet: si bien su característica distintiva de conectar mentes e información, al proporcionar acceso a la información y abrir posibilidades de cooperación o colaboración social, le hacía parecer originalmente una tecnología poderosa de mejoramiento cognitivo -en palabras de Persson y Savulescu, “the most realistic means of substantial cognitive enhancement” (2008: 167)-, sus efectos positivos sobre las facultades de conocimiento no son tan claros bajo cualquier uso de él, y esto no solo por su reconocido efecto debilitante de la memoria o porque anima a un procesamiento de la información superficial y que reduce la capacidad de contemplación (Carr, 2011), sino porque, cuando se experimenta un exceso o sobrecarga de información como la que se da en la *web*, esta puede resultar paralizante, bloquear el razonamiento y conducir al sujeto a buscar atajos para reducir

la complejidad y la incertidumbre, resultando con ello perjudicada también la toma de decisión. En los entornos en los que el sujeto se vive como desbordado o abrumado por un exceso de información, numerosos estudios han mostrado que esta tiende a filtrarse eligiendo aquella que concuerda con la propia concepción del mundo, la que resulta más llamativa o la que primero se recibió (Voinea et al., 2020). En el primer caso, se provoca un reforzamiento de la propia opinión que magnifica su validez y que tiende a mermar la capacidad del sujeto de lidiar con ideas contrarias, lo cual repercute en su capacidad democrática (Gozálvez-Pérez, 2011). Si a esto se suma el hecho de que en el entorno digital la información no se ordena ni estructura desde un criterio de calidad, sino de mercado, diseñado para captar la atención del usuario (“*distraction by design*”) aprovechándose de varias “vulnerabilidades humanas: la orientación de la atención a la novedad”, la necesidad de pertenencia al grupo y el hecho de que las personas “tienden a encontrar más placer en recursos informativos que confirman sus creencias ya existentes” (Voinea et al., 2020: 2352), la acción cognoscitiva puede verse seriamente afectada desde su punto de partida, cuando no deformada por ser estratégicamente dirigida a informaciones falsas.

Desde este marco, el presente trabajo pretende poner de manifiesto cómo la aplicación de sistemas de IA como los Sistemas de apoyo a la decisión (DSSs) está mediada por cierta comprensión del conocimiento y de la actividad de conocer que actúa a modo de marco que permite -y limita- las posibilidades de estos sistemas de efectivamente mejorar las capacidades cognitivas humanas. Esto es, se mostrará cómo la gnoseología que se asuma tiene implicaciones en la aplicación de la IA en general, y esto afecta de manera específica a algunos campos que lidian explícitamente con la creación de conocimiento y el aprendizaje, como el educativo.

La estructura que se seguirá es la siguiente: tras analizar los Sistemas de apoyo a la decisión (DSSs), las consecuencias de su uso y las propuestas de solución para contrarrestar sus efectos secundarios indeseados, se abordará críticamente la noción de conocimiento que está operando implícitamente en la generación e implementación de estos sistemas y que algunos críticos han puesto de manifiesto, para contraponer una concepción más ajustada a la condición humana que permita impulsar y aumentar, y no reducir, el sentido capacitante de la IA, es decir, su potencial colaborador en la expansión de las facultades operativas humanas y, por tanto, en la agencia del sujeto a través de ella. Esta concepción que se presenta comprende el conocimiento y, en consecuencia, su mejora a través de la tecnología, en un contexto más complejo o desde una aproximación más integradora que no tiene en

cuenta solo el factor tecnológico, sino que repara en cómo, en su uso, intervienen factores gnoseológicos, así como psicológicos y sociológicos que median en la capacidad efectiva de mejorar las capacidades humanas a través de la IA, y de los que las tecnologías se sirven para ejercer su influencia.

## 2. LOS DSSS O SISTEMAS DE APOYO A LA DECISIÓN: ¿HACIA UNA MEJORA COGNITIVA?

Los sistemas de apoyo a la decisión (DSSs) son sistemas que “no toman decisiones independientemente, sino que sugieren o recomiendan hipótesis o decisiones a unos supervisores humanos que las aprueban o rechazan” (Cornelissen, van Eerdt, Schraffenberger & Haselager, 2022: 18-19). En palabras del National Forum on Education Statistics, un DSS es un “sistema integrado de hardware y software específicamente diseñado para manipular datos y permitir a los usuarios sintetizar y compilar información útil procedente de fuentes diversas de datos en bruto o sin procesar para apoyar la resolución de problemas y la toma de decisión” (2006: 2). Los elementos nucleares de los DSSs radican en la integración que aportan de diversas fuentes de datos, de modo que no se reducen a una base de datos o red, y en su interactividad, a través de la cual se establece una relación entre máquina y persona. Su implementación es especialmente útil en relación a procesos no estructurados de toma de decisiones, donde el problema que se quiere abordar no ha aparecido antes, es complejo o impreciso (Bonczek, Holsapple & Whinston, 1981; Bennet & Bennet, 2008), y cuando las decisiones deben tomarse en un breve lapso de tiempo y en situaciones de incertidumbre. Los DSSs incluyen varios componentes: las bases de datos, el sistema de software y la interfaz de usuario. Su base o fundamento viene dado por la calidad de los datos en los que se apoyan, definida en términos de su utilidad o relevancia para la toma de decisión; validez, en tanto que son resultado de mediciones precisas, fiables y, por tanto, replicables, y oportunidad, esto es, disponibilidad a tiempo para la toma de decisión. Esta calidad suele aumentar cuando el individuo se siente personalmente responsable de ellos (National Forum on Education Statistics, 2006). Los DSSs se han convertido en una herramienta habitual en campos como la medicina, el derecho, las finanzas, los seguros, la búsqueda de pareja, etc., aunque no todos los campos los han integrado con la misma intensidad. Por ejemplo, Celik, Dindar, Muukkonen y Järvelä (2022) reportan cómo notoriamente el educativo está a la cola en la aplicación de IA.

Si bien hay estudios que concluyen que los DSSs mejoran la efectividad de la decisión, es decir, conducen a una decisión de mayor

calidad o a un rendimiento o resultado mejor, cabe preguntarse por los efectos que tienen en las facultades cognoscitivas, especialmente en el proceso de razonamiento que lleva a la apropiación y comprensión de lo que está en juego en cada ocasión. Uno de los problemas identificados en su uso es que, como reportan Cornelissen et al. (2022), su adopción ha conducido a una mayor pasividad del sujeto, que tiende a asumir *sistemáticamente* la decisión sugerida por los sistemas de IA. Es decir, se plantea que muchos usuarios “*have moved from active participants in the task to more passive monitors of automation performance*” (Merritt, Ako-Brew, Bryant, Staley, McKenna, Leone & Shirase, 2019: 2). Si consideramos que los sistemas de IA se califican de ‘inteligentes’ o como sistemas ‘cognitivos’, resultado de una ‘computación cognitiva’ (Noor, 2015) precisamente porque parecen aprender y, por ello, cometen errores, los riesgos que esto entraña son graves. En efecto, Cornelissen et al. temen que en períodos de tiempo prolongados o “con el paso del tiempo ellos [los humanos] puede volverse incapaces de cumplir su papel de monitoreo o supervisión de manera significativa, y confiar en el DSS, aportando meros sellos humanos de aprobación” (2022: 19). Mackay también recoge de Todd y Benbasat una conclusión semejante: que “los sujetos que usaban un DSS seleccionaban la estrategia menos costosa” o que menos esfuerzo requería (Mackay, 1992: 650).

Merritt et al. (2019) se refieren a este fenómeno derivado de la automatización como autocomplacencia o un monitoreo sub-óptimo o insuficiente de la actuación de los sistemas (“*complacency, or sub-optimal monitoring of automation performance*”), que está detrás de accidentes. Por ello, proponen y validan una escala para medir la susceptibilidad a ella en el uso de sistemas automáticos. El riesgo de autocomplacencia es mayor cuando la alfabetización que se propone de ellos es meramente técnica y no crítica, como suele ser habitual en las primeras fases de implementación de las tecnologías. Precisamente por esto se reclama la participación en el proceso y su supervisión por parte de expertos implicados en la propia configuración del sistema, aplicando una metodología de uso responsable de estas tecnologías desde el diseño. Sin embargo, cuando se trata del usuario final, esta participación puede verse comprometida.

Para evitar este efecto, Cornelissen et al. (2022) sugieren el uso de “*Reflection machines*”, que son sistemas de IA que se añaden a los Sistemas de apoyo a la decisión y guían la acción cognoscitiva tecnológicamente mediada por ellos. En concreto, las máquinas de reflexión invitan a una involucración activa en el proceso de razonamiento, a través de una actitud crítica con los resultados de los DSSs, de modo que el sujeto se vea empoderado en lugar de reducido a mero apéndice validador del proceso técnico. Se podría decir que se le incita a hacer una “lectura de resistencia”

(Walton, 2012: 227) de las conclusiones arrojadas por los DSSs, lo cual le permite mantener el control del proceso y, en definitiva, seguir siendo su sujeto o, como se dice en el ámbito pedagógico, su protagonista y no mero receptor y aplicador de su contenido o resultado. Cornelissen et al. definen las *Reflection machines* como “sistemas de apoyo adicionales... que proveen de *feedback* ante las decisiones humanas, con el propósito de aumentar el control y la implicación de los humanos en el proceso de toma de decisiones” (2022: 19). Lo hacen a través de imágenes o textos breves que hacen preguntas, apuntan a otros factores que podrían ser relevantes, a datos inconsistentes con la decisión adoptada o mediante la sugerencia de alternativas. Wang y Wang recogen que una interfaz gráfica ha demostrado ser más efectiva que una narrativa, favoreciendo la interactividad, y permite que “el que tiene que tomar la decisión pueda elegir entre una lista más acotada de alternativas, y pueda llevar a cabo análisis de sensibilidad del tipo ‘qué pasaría si...’” (2022: 129).

También Voinea et al. recomiendan asistentes de IA en el uso de Internet “que nos guíen a través del laberinto informacional... supervisen y nos hagan conscientes de nuestras acciones online... nos adviertan sobre los peligros y amenazas” (2020: 2357) para entrenar a los usuarios en un uso activo y crítico que ayude a discriminar y categorizar la información, e incluso la filtren desde criterios aportados por el usuario. Avanzando un paso más en esta dirección, Noor se refiere a los “ecosistemas cognitivos” del futuro desde una perspectiva holista (en la forma de *smart cities*, empresas o sistemas de producción cognitivos, etc.), en los que los asistentes cognitivos o “*personalized assistant that learns*” (PAL) (2015: 85) - que son “agentes de software inteligentes especializados y proactivos diseñados para seguir e interactuar con humanos y otros asistentes cognitivos a través de los entornos” (2015: 75)- y los humanos trabajarán juntos para ampliar las capacidades humanas, logrando “un aprendizaje oportuno, interesante, personalizado y colaborativo y una toma de decisiones efectiva”, “estimulando la creatividad y la innovación” (2015: 75). Esta concepción está en línea con la tesis de Andy Clark de que lo distintivo de la plástica inteligencia humana radica “en la capacidad de entrar en relaciones profundas y complejas con constructos no biológicos, soportes y ayudas”, colaborando y *fusionándose* con estas tecnologías “para resolver mejor los problemas de la supervivencia y la reproducción” (2003: 5). De este modo, según Clark el ser humano es un *cyborg* por naturaleza (*natural-born cyborg*), y esto significa que todas las tecnologías que ha desarrollado -desde las tradicionales, como el papel, el bolígrafo, el reloj, etc., hasta las más novedosas- no son meras *herramientas* de la mente humana, como si fueran cosas distintas de ella, sino que la constituyen por completo, esto es, se identifican con ella: “*Tools-r-Us and always have been*”

(2003: 7). Esto es, las tecnologías afectan la mente de un modo que la constituyen y, en el escenario de hoy en día en el que “nuestro mundo se vuelve más inteligente... resulta cada vez más difícil decir dónde termina el mundo y empieza la persona” (2003: 7), esto es, dónde trazar la línea entre “el yo biológico y el mundo tecnológico” (2003: 8). Sloterdijk también se refiere a un pensamiento ecológico en el que las relaciones del sujeto son múltiples y multidimensionales, de forma que los seres humanos no solo se relacionan entre sí, sino también “con las máquinas, los animales, las plantas... en una relación polivalente con el entorno” (Vásquez Rocca, 2009: 20) y en este sistema la información puede atribuirse de forma *horizontal* y *unívoca* a todo tipo de agentes: “Animales, máquinas y humanos por igual, como si compartieran el mismo estatus ontológico” (Chavarría Alfaro, 2015: 99).

La conveniencia de recurrir a herramientas como los asistentes cognitivos o las *Reflection machines* que se añadan y complementen el uso de los Sistemas de asistencia a la decisión para prevenir los riesgos que estos llevan asociados responde a motivos tanto *legislativos* como *antropológicos* o relacionados con el florecimiento humano. Por una parte, se reconoce “el derecho de los sujetos a que no se tome una decisión sobre ellos sobre la sola base de un proceso automático” (Reglamento General de Protección de Datos de la Unión Europea, artículo 22), sino con intervención humana. Por otra, se observa la necesidad de asegurar la aportación original y novedad de la que el sujeto humano es capaz con su iniciativa, que constituye un atributo radical de la persona que hay que desplegar a través de la acción característica de sus facultades superiores para que sea capaz de “formular y llevar a cabo... una forma de existencia personal, singular” (Bernal Guerrero, 1994: 254), que se percibe amenazada desde determinados usos de los DSSs. Se reclama, entonces, la necesidad de “impulsar el proceso de personalización, mediante el cual el hombre pone en acto sus potencialidades personales” (Medina Rubio, 1989: 17).

Si bien el uso de *Reflection machines* puede mejorar la calidad del proceso de toma de decisión, empujando al sujeto a que lo transite por sí mismo y ejerza así un dominio real sobre él, reafirmandose como su sujeto o protagonista, cabe cuestionar que el recurso a más tecnología sea suficiente para paliar los problemas suscitados por los DSSs, así como si realmente los “ecosistemas cognitivos” del futuro generarán los efectos positivos sobre el conocimiento con los que Noor (2015) los vislumbra.

Los propios Cornelissen et al. (2022) alertan del peligro de que estas máquinas de reflexión se limiten a la forma más básica de preguntar si se está seguro de la decisión tomada, a lo que, por muchas veces que se repita -y precisamente a causa de un uso desmedido de estos recursos- el usuario rápidamente se acostumbra, de forma que no se consigue el objetivo



deseado de estimular la reflexión y el pensamiento crítico. Por ello, los autores recomiendan que el *feedback* de estas máquinas debe ser específico y estar referido a decisiones concretas.

Sin embargo, con frecuencia, a los problemas de la IA se responde con más IA, cuando lo que se necesita es algo de otro tipo. Como Voinea et al. indican: “*While technological solutions are necessary, they are not sufficient... [and] must be complemented by traditional forms of user empowerment*” (2020: 2355). Para un uso efectivo de estos sistemas que realmente conduzca a una mejora, y no reducción, de las capacidades cognoscitivas, es necesaria la formación: una alfabetización que no puede ser solo técnica, sino crítica, y que pasa por un cultivo humanista -y decididamente no transhumano- de las facultades del sujeto, especialmente de las habilidades performativas y de las virtudes intelectuales, que son las que en último término posibilitan los logros cognoscitivos, mediadas por las capacidades morales del sujeto. Esto es, para un uso de los sistemas de IA que efectivamente tenga potencial ampliador de las capacidades de conocimiento, es imprescindible cultivar los hábitos intelectuales que permiten apropiarse de lo conocido a través de la extensión tecnológica y experimentar el resultado como logro cognoscitivo *personal*, esto es, como propio, derivado de la singularidad e intimidad creativa de la persona.

Si, en relación con las tecnologías de la información y la comunicación (TIC), se habla de transitar desde el paradigma TIC “al paradigma TAC (tecnologías del aprendizaje y el conocimiento) y al paradigma TEP (tecnologías del empoderamiento y la participación) (Reig, 2013)” (Macchiarola, Martini, Montebelli & Guazzone, 2015: 103), lo mismo cabe aplicar a los sistemas de IA. En definitiva, se trata de la *apropiación* que el sujeto logra de ellos, a la vista de que el crecimiento exponencial de la penetración de estos sistemas en los diferentes campos de la práctica no se ha traducido en iguales niveles de apropiación por parte de los usuarios, esto es, en un aprovechamiento de sus potencialidades para mejorar tanto las capacidades cognoscitivas del sujeto en un sentido *objetivo* (teniendo en cuenta no solo el resultado logrado, sino la calidad del proceso para llegar a él), como la experiencia *subjetiva* de la acción cognoscitiva en términos de la percepción que el sujeto tiene de la efectiva ampliación o no de su capacidad a través de estos dispositivos. Respecto a esto último, indican Cornelissen et al. (2022) que “sería interesante medir el sentido de agencia del usuario y su dependencia de sistemas externos” (2022: 19) a través de cuestionarios que pregunten si los diagnósticos o decisiones se han tomado de manera completamente independiente, si esos sistemas han influido la decisión, o si algunas preguntas concretas lo han hecho y cómo. Esto apunta a la contribución de estos sistemas al metaaprendizaje del sujeto, esto es, al desarrollo de su capacidad de aprender a aprender, si se

posibilita la transparencia que permite reconstruir el proceso de razonamiento seguido.

Respecto a la IA, debe evitarse cometer el mismo error que ha contribuido a profundizar la brecha digital en relación con otras tecnologías: la mera dotación de infraestructura -en este caso, de sistemas de IA- no basta para asegurar unos mejores resultados en términos cognoscitivos y, en último término, humanos. Por ello, los esfuerzos deben centrarse en comprender el proceso humano de apropiación de ellos para cada una de las prácticas y situaciones distintivas en que se desempeñan los sujetos. Esto es, la aproximación no puede ser generalista, sino enfocada a determinados actores sociales y a las situaciones específicas que en cada contexto se plantean, pues son las necesidades particulares las que marcan los procesos de apropiación (Macchiarola et al., 2015).

En IA, como sucede en cualquier otro asunto, cuando se piensa desde un marco, concepción o imagen del mundo (*mythos*), se ofrecen las soluciones a los problemas disponibles dentro de ella, concebidas dentro de sus posibilidades. Detectar los presupuestos naturalmente asumidos en cada *mythos* desde su contexto histórico permite poner de manifiesto sus contornos y, por consiguiente, la naturaleza de las respuestas que tiene disponibles, los límites (*peras*) en que estas se ofrecen y, en consecuencia, también sus limitaciones, para abrir el horizonte hacia posibilidades de otro tipo. Con este propósito, a continuación, se presenta y revisa la noción de conocimiento que tiende a estar operando detrás del diseño e implementación de los Sistemas de apoyo a la decisión.

### 3. EL CONCEPTO DE CONOCIMIENTO EN EL DISEÑO DE DSSS

“Toda tecnología es siempre un intento por asistir o modificar actividades y prácticas sociales” (Roco, 2007: 369). De hecho, desde una visión relacional y no meramente incrementalista de la tecnología, se entiende que esta puede “crear propósitos o metas nuevas no consideradas por otras herramientas”, de modo que “las herramientas modifican al usuario: sus percepciones, relaciones, prácticas e interpretaciones” (Macchiarola et al., 2015: 95). Aplicando esto a las tecnologías de la computación, “*automation does not simply supplant human activity but rather changes it, often in ways unintended and unanticipated by the designers of automation*”, como indican Parasuraman y Manzey (2010: 381).

Esto es lo que ha sucedido respecto a la acción cognoscitiva: sistemas diseñados para llevar a cotas más altas el conocimiento humano, acaban provocando, bajo ciertas condiciones, el efecto contrario: si, respecto a Internet, se ha constatado que distrae la atención, empobrece la memoria,

vuelve la comprensión endeble, haciéndola superficial, debilita el razonamiento y, con ello, la capacidad de decisión, lo mismo debe revisarse respecto a los DSSs.

Estos efectos indeseados se originan, al menos en parte, a causa de un diseño que parte de presupuestos erróneos relativos a la actividad cognoscitiva humana, es decir, a la manera de comprender el conocimiento, así como de ideas ingenuas o idealistas en torno a la aplicación efectiva de estos sistemas en la realidad, que no toman en cuenta factores psicológicos y sociológicos que median en la manera de poner en práctica las facultades cognoscitivas a través de ellos.

Como Mackay ya indicaba, “historically... DSS research employed an input-output philosophy” (1992: 649), and “virtually all aids to human thinking were developed without significant attention to the thought process itself” (1992: 650). Como el mismo autor afirma, esta aproximación:

is only a simplistic representation of the actual problem-solving process. Problem-solving processes, the link between inputs and outcomes, largely have been treated as “black boxes” that have been recognized but not incorporated into the research design of empirical investigations (1992: 650).

Desde esta perspectiva, se pierde información muy relevante del proceso que impide comprender cuáles son los mecanismos concretos que ayudan a empoderar cognitivamente al sujeto, en lugar de degradarlo a mero apéndice de la tecnología y, en consecuencia, resulta imposible saber qué dinanismos habría que activar con el objetivo de mejorar los logros cognoscitivos a través del uso de estos sistemas. Por ejemplo, esta perspectiva no permite conocer cómo las diferencias individuales influyen en su uso y, en consecuencia, cómo intervienen en la efectiva ampliación cognoscitiva lograda a través de ellos o en su reducción.

En otras palabras, se constata que estos sistemas no son igualmente aprovechados por distintos sujetos, sino que hay que tener en cuenta las diferencias individuales en la extracción de su potencial. Esto es, hay características psicológicas, sociológicas, afectivas, laborales (como la cantidad de experiencia profesional), etc. que influyen en el rendimiento que se les pueda sacar. Por ejemplo, Mackay (1992) recoge de otros autores algunas diferencias en el uso que hacen de ellos expertos y novatos o principiantes, respectivamente: los primeros

are knowledgeable about their domain, [know] how to apply and use... knowledge more effectively... [and] considered several ways of solving a problem, whereas novices tended to use the first method that came to mind... Experts prefer to take a system view of the problem, whereas novices focus on obtaining a solution rather than on understanding the problem (1992: 651).

Una consideración de estos sistemas que obvia el proceso denota un funcionalismo que “valora más [...] las capacidades que el camino” (García-Gutiérrez, Gil Cantero & Rejero García, 2017: 27), es decir, el resultado deseable frente a la *experiencia* vivida, pese a lo subjetivamente valiosa que es esta para la mejora o, más precisamente, para vivir esta como tal, como auténtica mejora. Como indica García-Gutiérrez et al., “es en el camino de la formación donde se encuentran específicamente las claves de la realización humana” (2017: 27). En otras palabras, para una mejora cognitiva *formal*, que sea vivida precisamente en tanto que mejora o crecimiento que se integra con el ser de la persona, es clave la transparencia o la comprensión del proceso, la “conexión experiencial e inteligible” entre los medios usados y el fin logrado (Bostrom & Roache, 2007: 13). Esto implica al menos la conciencia *inobjetiva* o *atemática* de estos procesos, a la que se puede sumar una conciencia *reflexiva* (Millán-Puelles, 1967) del significado que han supuesto para la situación. Sólo esta experiencia subjetiva de distinto orden permite acoger, reconstruir y comprometerse con lo conseguido como genuino logro personal. Sin esta experiencia subjetiva, la persona podría conocer algo temáticamente -el resultado arrojado por el sistema-, pero no se podría conocer, inobjetivamente, como el ser que conoce aquello, si permanece opaco el proceso. Es decir, sin una implicación en este, no se comprende la formación de los hábitos, en este caso, de virtudes intelectuales, ni, en consecuencia, el significado humano de la mejora. Sucedería aquí lo mismo que ocurre respecto a la mejora biotecnológica de corte transhumanista: podríamos “como mucho sentir sus efectos sin entender su significado en términos humanos”, como recogen Bostrom y Roache (2007: 13) de Leon Kass. Frente a esto, “en la mayoría de nuestros esfuerzos cotidianos de automejora... notamos la relación entre nuestro hacer y la mejora resultante, entre los medios usados y el fin buscado. Hay una conexión experiencial e inteligible entre medios y fin” (Bostrom & Roache, 2007: 13).

Esto está en línea con la tesis de la epistemología de la virtud (Sosa, 2007) o lo que se ha denominado la “visión meritoria del conocimiento” (*Credit view of knowledge*) (Greco, 2007; Sosa, 2007), según la que el conocimiento incluye “una dimensión normativa o evaluativa” (Greco, 2007: 57), es decir, ha de entenderse como un logro y, por tanto, como algo meritorio, digno de ser valorado. De hecho, como Greco (2007) indica, en algunas visiones morales como la aristotélica el conocimiento es un constitutivo fundamental del florecimiento humano y, por tanto, posee valor intrínseco (moral) para la persona. Esto es, constituye una instancia o ejemplo de actividad meritoria, a saber, en este caso constituye un logro de tipo cognitivo, y lo es en la medida en que proviene o es resultado de unas facultades cognoscitivas confiables, esto es, de la excelencia de estas

capacidades y no de otro motivo. En términos clásicos, se puede decir que una creencia verdadera constituye auténtico conocimiento si y solo si es la virtud de las facultades cognoscitivas -es decir, su disposición o inclinación a actuar de cierto modo (de modo que conduce a buenos resultados)- la causa por la que se tiene esa creencia, de modo que esta creencia no es resultado meramente accidental ni puede atribuirse a la suerte.

Si se aplica esta comprensión del conocimiento como logro a la cuestión que aquí se analiza, los DSSs aumentarían el conocimiento humano no simplemente cuando aportan una creencia verdadera al sujeto que los utiliza, sino cuando esta creencia es asumida *a causa de* la excelencia lograda en las facultades del sujeto provocada gracias al uso y apropiación de estos sistemas. Es decir, cuando esa creencia *resulta de* la potenciación de las facultades cognoscitivas del sujeto conseguida gracias a estos sistemas, que, en consecuencia, aumentarían su disposición a lograr buenos resultados. En otras palabras, las volverían más virtuosas.

Bostrom y Roache (2007) plantean una objeción a esta necesidad de enfocarse en el proceso (entendido como apropiación o logro) en el caso de la mejora biotecnológica, al referirse a que hay ocasiones en las que sólo importa el resultado, independientemente de que se conozca que hay medios superiores que llevan a él. Por ejemplo, “puede que cojamos el autobús para llegar a algún sitio aunque reconocemos que hay más valor en ir corriendo, o contratamos a un jardinero para cultivar nuestro jardín aunque reconocemos que tiene más valor el hacerlo nosotros mismos” (Bostrom & Roache, 2007: 14). Si aplicamos esto a los sistemas de apoyo a la decisión, se podría decir que en muchas ocasiones lo único que se quiere es un buen diagnóstico o, en general, un resultado certero, independientemente de los medios que se hayan aplicado para llegar a él, siempre y cuando estos sean legítimos. De manera semejante, también el fiabilismo (*reliabilism*) (Bonjour, 2002) considera que el proceso no es siempre relevante para el conocimiento. En concreto, sostiene que el *conocimiento del proceso* confiable no es necesario para que haya conocimiento. A saber, según esta postura, un sujeto tiene conocimiento de que “p” es verdadero si y solo si se cumplen tres condiciones: (a) p es verdadero, (b) el sujeto cree que p es verdadero y (c) el sujeto ha llegado a esa creencia a través de un proceso confiable. Sin embargo, la última condición, (c), no implica que el sujeto conozca en detalle los intrínsecos del proceso que hacen que la creencia sea fiable. Por ejemplo, se puede decir que un sujeto conoce que hay un libro encima de la mesa simplemente porque lo ha percibido visualmente, pero no necesariamente tiene que conocer todos los procesos cognitivos que están implicados en la percepción visual humana. Basta con que la percepción funcionó de manera adecuada para que pueda afirmarse del sujeto que conoce tal cosa.

Por otra parte, Lackey (2009) critica que el conocimiento no se vincula necesariamente o no siempre tiene que ver con a algún tipo de mérito o logro personal, como ocurre en los casos de conocimiento testimonial, esto es, de conocimiento que un sujeto tiene a causa del testimonio ofrecido por otra persona (que puede ser o no experta). En estos casos, el sujeto en cuestión tiene una creencia verdadera, según Lackey; tiene, en efecto, conocimiento, pero carece de mérito propio, puesto que aquella creencia verdadera no puede atribuirse a la excelencia de sus propias facultades cognoscitivas, sino más bien a las del testigo.

A estas tres objeciones se puede contestar lo siguiente. Respecto a la primera de Bostrom y Roache, que resalta la (a veces exclusiva) importancia que a efectos prácticos tiene el resultado de las acciones, quitando valor al medio empleado para llegar a él, se puede notar que esta objeción encierra una visión utilitarista de los hábitos o virtudes del sujeto, a los que se entiende como simple medios para lograr un resultado, sin reparar en su carácter finalístico y, por tanto, en la significatividad humana que tienen en sí mismos. Por tanto, frente a esta primera objeción se puede responder que, aunque ciertamente hay muchos caminos que conducen a un fin y estos pueden equipararse en su calidad de *medios*, esto no significa que no se distingan en su diversa valía para “promocionar lo humano” (García-Gutiérrez et al., 2017: 24), puesto que unos medios sirven mejor que otros para este fin. Esto es semejante a lo que ocurre con la electroestimulación muscular frente a la actividad deportiva habitual: mientras que el resultado puede –aparentemente– ser el mismo (si bien Zohny (2015) lo pone en duda<sup>2</sup>), la primera práctica elimina muchos de los beneficios del deporte para el cultivo de la humanidad. En otras

---

<sup>2</sup> Zohny (2015) se muestra crítico respecto a los resultados logrados mediante el uso de fármacos para la mejora cognitiva: descalifica como un mito la afirmación, ligeramente aceptada, de que ya existen fármacos que mejoran nuestras capacidades cognitivas. Si bien reconoce la evidencia empírica de que, entre los que tienen algún tipo de patología (por ejemplo, los que muestran falta de sueño) estos son efectivos, no ocurre así respecto a individuos sanos, pues las evidencias de eficacia en este caso no son suficientes o no son significativas, además de que no se suelen registrar la cantidad de estudios con resultados nulos. El autor nota que, de hecho, en algunos casos se han encontrado evidencias de que, aunque mejoraban ciertas funciones ya adquiridas, han ralentizado el tiempo de respuesta, dificultando el pensamiento creativo y “tareas que requieren adaptación, flexibilidad y planeamiento” (p. 3). Además, precisa que, si bien la experiencia o la disposición subjetiva inducida por estos fármacos (es decir, el estado de ánimo) es mejorada con su ingesta y, por tanto, resulta más favorable al rendimiento, según reportan diversos sujetos, no se puede decir que “de hecho mejoren el rendimiento objetivo” (p. 4). Así, concluye que “las expectativas... [que hemos puesto en ellos] exceden sus efectos actuales” (p. 3).

palabras, en ella, el resultado no puede atribuirse a la excelencia de las facultades del sujeto.

Respecto a la segunda objeción del fiabilismo (*reliabilism*), según la que, para que haya conocimiento, no es necesario conocer el proceso que conduce a tener una creencia fiable, se puede responder que hay que distinguir dos tipos de conciencia de los actos y procesos mentales: inobjetiva y objetiva. La conciencia inobjetiva o concomitante de ellos es aquella que acompaña los actos mentales, sin que estos resulten tematizados en sí mismos; mientras que la conciencia objetiva o reflexiva de ellos los vuelve el tema u objeto explícito de un nuevo acto de conciencia reflexivo. Desde esta distinción, podría contestarse al fiabilismo que, para que una creencia esté justificada y, por tanto, sea fiable y constituya auténtico conocimiento, basta con tener una conciencia inobjetiva o concomitante de los actos mentales que conducen a ella y, por tanto, no necesariamente reflexiva o metacognitiva. En el ejemplo mencionado arriba sobre la creencia de un sujeto de que hay un libro sobre la mesa -creencia lograda a través de un proceso confiable de percepción- basta con tener conciencia atemática o inobjetiva del acto de percepción para que efectivamente pueda describirse esa creencia como auténtico conocimiento. Un enfoque metacognitivo o reflexivo que analice los actos mentales implicados en el proceso perceptivo no ayudará a que haya más conocimiento sobre el libro, sino que aportará nuevos conocimientos no ya sobre el libro, sino sobre el propio sujeto o sus actos mentales. En otras palabras, el conocimiento no es siempre reflexivo, pero es, a pesar de ello, auténtico conocimiento.

Si esto se aplica a los DSS, se puede considerar que estos sirven para aumentar el conocimiento humano en la medida en que ayudan al sujeto a realizar nuevos actos mentales, explorar nuevas vías para la resolución de problemas, y basta con la conciencia inobjetiva de estas para que pueda considerarse que estos dispositivos potencian el conocimiento humano. Por tanto, la transparencia en el proceso demandada al diseño de estas tecnologías se refiere al menos a la capacidad de los sujetos que emplean estos sistemas para seguir por sí mismos el camino que proponen o señalan estos sistemas, viviendo el proceso por sí mismos al menos de forma inobjetiva, pero no necesariamente metacognitiva.

El riesgo de asumir los sistemas de apoyo a la decisión acríticamente radica en considerar solo la satisfacción que proporciona el logro de un resultado, en base a los sentimientos positivos que genera (un alivio, la reducción de una penosidad o de cierta incertidumbre). Pero la búsqueda de resultados correctos sin más consideraciones es indicativa de la vieja tendencia a identificar el bien con el placer o con la satisfacción obtenida, y el mal con su contrario (Platón, 1992), que “es de una ingenuidad

bastante necia” (Marina, 2003: 13), y da lugar a una autonomía ilusoria: “Quienes entienden la libertad sin lucha, fácilmente se hacen esclavos de sí mismos o de los demás” (Gervilla Castillo, 2003: 108) y, en este caso, de la tecnología. Los propios Bostrom y Roache notan que hay casos en los que podemos querer sentir una emoción negativa o desagradable -en el caso que nos ocupa, llegar a un resultado erróneo-, incluso aunque tengamos los medios para evitarlo a través, por ejemplo, de cierto fármaco o de sistemas de IA: “Puede que queramos ser el tipo de persona que sentiría profunda tristeza ante la muerte de un ser querido” (Bostrom & Roache, 2007: 14). Si esto puede considerarse preferible en algunas circunstancias, debe ser porque también las experiencias negativas o desagradables, al igual que el error, son experiencias humanamente valiosas, y pueden “mejorar nuestra comprensión de nosotros mismos y de otros, y dar a nuestra personalidad una riqueza y profundidad que pueden faltar si solo experimentamos emociones ‘positivas’” (Bostrom & Roache, 2007: 12) o respuestas acertadas. Por esto precisamente hay una tendencia a promover una cultura del error como valor positivo en las organizaciones, entre ellas los espacios educativos (Ariso, 2019). El logro de la plenitud humana tiene más que ver con la autenticidad de recorrer el proceso que con la facilidad o la satisfacción que aportan unos resultados, pues la virtud -en este caso, intelectual- no está determinada por el resultado alcanzado, sino que tiene que ver con el contraste entre los diferentes puntos de partida y los logros conseguidos. En esto, el conocimiento es esencialmente un acto de libertad, que tiene mucho que ver con la libertad conquistada o el carácter moral forjado por la persona (González-Ayesta, 2010). Frente a esto, la última objeción señalada arriba, hecha por Lackey (2009), desvincula el conocimiento del mérito o del logro personal al señalar casos en los que se da en un sujeto auténtico conocimiento, pero no se le puede atribuir mérito o ningún sentido de logro, como ocurre en el conocimiento testimonial, en el que la creencia verdadera que el sujeto posee no depende de su habilidad o de la excelencia que ha desarrollado en sus propias facultades cognoscitivas, sino de las del testigo. Greco (2007) responde a esta objeción desde una epistemología de la virtud afirmando que, incluso en estos casos, el conocimiento que se posee se debe al mérito del sujeto al saber discriminar fuentes fiables de información de otras que no lo son, si bien Lackey responde reiterando su postura. En palabras de Greco: “Testimonial knowledge requires that the *believer* is a reliable *receiver* of testimony. That is, what is important is not so much that the testifier is reliable, or that the believer knows that he is, but that the believer herself is reliable in the way that she receives and evaluates testimony. This will plausibly involve reliable capacities for discriminating



reliable sources of testimony form unreliable sources” (2007: 63), y esta capacidad es digna de mérito.

Si el conocimiento es, en efecto, un tipo de logro, es necesaria la apropiación de los procesos, a veces digitalmente mediados, que conducen a él, por la que el sujeto se identifica con ellos tras haberlos recorrido con una conciencia inobjetiva y después, quizá, reflexionado críticamente sobre ellos. La imposibilidad de que los sistemas de IA colaboren en la mejora cognitiva aparece, por tanto, cuando el proceso de razonamiento no puede ser revisado retrospectivamente y asumido “en un proceso de apropiación, o bien de eliminación posterior”, como también sucede respecto a las biotecnologías, como indica Ferrer (2005-2006: 178), bien por tratarse de sistemas opacos o por un uso deficiente -autocomplaciente- de sistemas que sí son transparentes. En el primer caso, su mecanismo resulta inasumible mientras el sujeto está bajo la acción de su influjo, como ocurre con la ingesta de un psicofármaco durante el tiempo que dura su efecto. En el segundo, no es de hecho asumido por razones de tipo psicológico (autocomplacencia, pereza intelectual, ingenuidad) y sociológico (funcionalismo, orientación a los resultados, productividad, hipercompetitividad, control) que están operando en su manejo y que más adelante se revisarán. En cualquier caso, un nuevo terreno de opacidad se yuxtapone al espacio subjetivo y modifica la experiencia consciente posterior, específicamente la cognoscitiva.

La opacidad necesariamente afecta a la conciencia humana, por tratarse de una conciencia *inadecuada* (Millán-Puelles, 1967), esto es, una conciencia que no se identifica con el sujeto en el que se da, y apunta a la dimensión de facticidad o fatalidad de la vida humana -su carácter de subjetividad *reiforme*, dice Millán-Puelles-, que también se expresa, como indica Ortega y Gasset (2015), en la circunstancia que le viene impuesta o en su origen (Arendt, 2013): la cuna en la que se nace, la lengua materna, los factores genéticos, los hábitos de vida de la madre, etc. De este origen el sujeto se puede distanciar, al revisarlo, pero resulta ineliminable, y por ello tiene consecuencias irreversibles en la propia configuración de la persona, como también mencionan Bostrom y Roache (2007). Sin embargo, el hecho de que haya “prácticas tradicionales de moldeamiento del individuo” –“*such as taking folic acid supplements, eating healthily, and abstaining from taking drugs during pregnancy... the environment in which very young children are raised... , language-learning*” (Bostrom & Roache, 2007: 21)- en las que no se cuenta con su voluntad, y con resultados irreversibles para él, no debe convertirse en argumento para extender la dimensión de facticidad del sujeto hasta copar las operaciones más características de su libertad, frente a lo que Bostrom y Roache sugieren para defender la conveniencia de la intervención biotecnológica, en concreto, de la

ingeniería genética para el mejoramiento humano. En el caso que nos ocupa, preservar la libertad característica de la operación cognoscitiva requiere hacer un uso de los sistemas de IA que asegure la presencia y participación del sujeto en el proceso y su dominio o apropiación, como tarea por la que el sujeto, como individuo *activo*, dota de significado al proceso a través de una construcción *personalizada* de usos de los sistemas de IA –algunos inéditos– que transforma sus prácticas cotidianas. Esto implica que cada sujeto genera propuestas alternativas de uso o usos novedosos y no previstos de las tecnologías de la computación.

Por tanto, si la inteligencia artificial está por fundar nuevos espacios de intencionalidad que se sumen a la intencionalidad natural como relación clave del conocimiento, es necesario que no genere, a su vez, nuevas opacidades como dimensiones añadidas de facticidad incontrolable por parte del sujeto que, frente a lo esperado, empobrecerían las capacidades cognoscitivas humanas en lugar de aumentarlas.

#### 4. LOS FACTORES PSICOLÓGICOS Y SOCIOLÓGICOS EN JUEGO

Si acudimos a los factores psicológicos y sociológicos que están operativos en la manera de usar los sistemas de apoyo a la decisión y que parecen contribuir a un uso pasivo de ellos, puede observarse lo siguiente. Por una parte, respecto a los factores sociológicos, se aprecia que, si bien en principio estos sistemas parecen estar disponibles para un uso a voluntad, cabe dudar de que así sea cuando el procedimentalismo se imponga. Ya hay evidencias contundentes de esto en muchos campos de la práctica, invadidos por una presión social para reconstruir sus procesos desde determinados usos de la tecnología marcados por un contexto hipercompetitivo de control, en que la productividad y la eficiencia aparecen como criterios de calidad definitivos. En medicina, por ejemplo, la sensación de ser un mero funcionario, aplicador burócrata de medios para lograr cualesquiera que sean los fines que al paciente se le antojen para satisfacer sus preferencias y deseos se ha extendido, y genera el conocido efecto *burnout* entre los médicos, que se refiere a una sensación de ausencia de logro que lleva a la desmoralización y desencanto y, en último término, a la indiferencia y desinterés de los médicos que, frustrados, no experimentan ya su trabajo como auténtica profesión dirigida a un bien claramente definido hacia el que sentir vocación por su carácter intrínsecamente valioso y digno de ser perseguido. Sienten haber pasado de un ejercicio autónomo de la profesión a una heteronomía que siempre resulta en el beneficio de alguna instancia externa al proceso médico. Como indican Curlin y Tollefsen, los médicos se viven “como un engranaje eficiente en una vasta industria de atención sanitaria

estructurada burocráticamente” (2021: XIII), establecida según un “modelo de provisión de servicios” que se rige por criterios de mercado. El resultado es que “los profesionales médicos llegan a trabajar en un ámbito altamente regulado en el que el ejercicio del juicio clínico y la prudencia no es ni posible ni deseable” (2021: 3). La sensación de ausencia de logro que es característica de este desgaste laboral tiene mucho que ver con un uso empobrecedor de las tecnologías de asistencia para la práctica profesional.

Por otro lado, entre las razones psicológicas que contribuyen a una mayor pasividad del sujeto en el uso de los Sistemas de apoyo a la decisión, que tiende a asumir la decisión sugerida por ellos, se encuentran, por un lado, la pereza que provoca un uso no autónomo del entendimiento a la que Kant se refiere como origen de la pérdida de libertad, que permea la condición humana, si bien se intensifica o se agudiza de manera especialmente notoria bajo las condiciones sociológicas examinadas arriba. Pero, de manera más interesante, se encuentran, por otro lado, la ingenuidad y el convencimiento, cada vez más extendido, de la grave limitación cognoscitiva de la condición humana, que no solo se vive como intrínseca, estructural o constitutivamente insuperable, como es propio de la conciencia -inadecuada- de una subjetividad reiforme, sino de alguna manera como inhabilitante. Desde la comprensión funcionalista de la mejora cognitiva y del conocimiento humano como sistema procesador de datos que ha estado operativa en la forma de entender la IA, el sujeto se vive a sí mismo como desbordado, como incapacitado para realizar una de las actividades que tradicionalmente se le consideraban más propias: conocer, deliberar, tomar decisiones informadas. Esta concepción del conocimiento provoca una vívida experiencia de carencia, una vivencia aguda de limitación que lleva a una excesiva desconfianza en la propia capacidad cognoscitiva y, a su vez, a una confianza excesiva en la tecnología, pues se observa que, de hecho, entendido el conocimiento de esa manera, efectivamente las máquinas lo hacen mucho mejor que el ser humano: estas son mucho más eficientes para la tarea de procesamiento de datos que la frágil e inconstante capacidad humana, y puede hablarse entonces de una auténtica disimetría que no sólo es de grado, si se piensa en una superinteligencia (Urban, 2015).

Frente a esta tendencia, no cabe sino corregir la noción de conocimiento humano que se presupone en el diseño de sistemas de IA y que informa la práctica profesional, reclamando el carácter prudencial de esta y, en consecuencia, el carácter libre y liberador del conocimiento humano, tanto como *resultado*, como en tanto que *actividad* distintivamente humana. El uso de sistemas de IA y de asistentes tecnológicos (*Reflection machines*, *Personal Assistants that Learn* y otras propuestas que vengan)

contribuirá a una mejora cognitiva efectiva en la medida en que estos colaboren a potenciar y extender estos radicales humanos de una manera que no ocurriría sin ellos.

#### REFERENCIAS

- Arendt, H. (2013). *Sobre la revolución*. Madrid: Alianza.
- Ariso, J. M. (2019). Can a culture of error be really developed in the classroom without teaching students to distinguish between error and anomalies? *Educational Philosophy and Theory*, 51(10), 1030-1041. <https://doi.org/10.1080/00131857.2018.1519699>
- Bennet, A. & Bennet, D. (2008). The decision-making process in a complex situation. En F. Burstein & C. W. Holsapple (Eds.), *Handbook on Decision Support Systems 1* (pp. 3-20). Berlin/Heidelberg: Springer.
- Bernal Guerrero, A. (1994). *Pedagogía de la persona. El pensamiento de Víctor García Hoz*. Madrid: Escuela Española.
- Bonczek, R. H., Holsapple, C. W. & Whinston, A. B. (1981). *Foundations of decision support systems*. Cambridge (Massachusetts): Academic Press.
- Bonjour, L. (2002). Internalism and Externalism. En P. K. Moser (Ed.), *The Oxford Handbook on Epistemology* (pp. 234-265). New York: Oxford University Press.
- Bostrom, N. & Roache, R. (2007). Ethical issues in Human Enhancement. En J. Ryberg, T. Petersen & C. Wolf (Eds.), *New Waves in Applied Ethics* (pp. 120-152). Londres: Palgrave Macmillan.
- Bostrom, N. & Sandberg, A. (2009). Cognitive enhancement: methods, ethics, regulatory challenges. *Science and Engineering Ethics*, 15, 311-341. <https://doi.org/10.1007/s11948-009-9142-5>
- Carr, N. (2011). *The shallows: What the Internet is doing to our brains*. Nueva York: Norton.
- Celik, I., Dindar, M., Muukkonen, H. & Järvelä, S. (2022). The promises and challenges of artificial intelligence for teachers: a systematic review of research. *TechTrends*, 66, 616-630.
- Chavarría Alfaro, G. (2015). El posthumanismo y los cambios en la identidad humana. *Reflexiones*, 94(1), 97-107.
- Cinel, C., Valeriani, D. & Poli, R. (2019). Neurotechnologies for Human Cognitive Augmentation: Current State of the Art and Future Prospects. *Frontiers in Human Neuroscience*, 13. <https://doi.org/10.3389/fnhum.2019.00013>
- Clark, A. (2003). *Natural-Born Cyborgs: Minds, Technologies, and the Future of Human Intelligence*. Nueva York: Oxford University Press.

- Cornelissen, N. A. J., van Eerdt, R. J. M., Schraffenberger, H. K. & Haselager, W. F. G. (2022). Reflection machines: increasing meaningful human control over Decision Support Systems. *Ethics and Information Technology*, 24(19). <https://doi.org/10.1007/s10676-022-09645-y>
- Curlin, F. & Tollefsen, C. (2021). *The Way of Medicine: Ethics and the Healing Profession*. Indiana: University of Notre Dame Press.
- Ferrer, U. (2005-2006). Posthumanismo y dignidad de la especie humana. *Contextos*, 45-48, 171-188.
- García-Gutiérrez, J., Gil Cantero, F. & Reyero García, D. (2017). El sujeto ético en los estudios universitarios de educación: humanismo, poshumanismo y democracia. *Bordón*, 69(4), 19-33. <https://doi.org/10.13042/Bordon.2017.690402>
- Gervilla Castillo, E. (2003). Pedagogía del esfuerzo y cultura del placer. *Revista Española de Pedagogía*, 224, 97-114.
- González-Ayesta, C. (2010). El amor a la verdad en Tomás de Aquino. *Revista Española de Filosofía Medieval*, 17, 37-46. <https://doi.org/10.21071/refime.v17i.6143>
- Gozálvez-Pérez, V. (2011). Educación para la ciudadanía democrática en la cultura digital. *Comunicar*, 36, 131-138. <https://doi.org/10.3916/C36-2011-03-04>
- Greco, J. (2007). The Nature of Ability and the Purpose of Knowledge. *Philosophical Issues*, 17, 57-69. <https://10.1111/j.1533-6077.2007.00122.x>
- Kristjánsson, K., Fowers, B., Darnell, C. & Pollard, D. (2021). Phronesis (Practical Wisdom) as a Type of Contextual Integrative Thinking. *Review of General Psychology*, 25(3), 239-257. <https://doi.org/10.1177/10892680211023063>
- Lackey, J. (2009). Knowledge and credit. *Philosophical Studies*, 142(1), 27-42. <https://10.1007/s11098-008-9304-3>
- Macchiarola, V., Martini, C., Montebelli, A. E. & Guazzone, J. (2015). Una política pública universal de inclusión digital. Cambios emergentes, obstáculos y desafíos. *Revista Iberoamericana de Educación*, 68(2), 83-106.
- Mackay, J. M., Barr, S. H. & Kletke, M. G. (1992). An Empirical Investigation of the Effects of Decision Aids on Problem-Solving Processes. *Decision Sciences*, 23(3), 648-672. <https://doi.org/10.1111/j.1540-5915.1992.tb00410.x>
- Marina, J. A. (2003). Educación para el esfuerzo. *Aula de Innovación Educativa*, 120, 12-14.
- Medina Rubio, R. (1989). La educación como un proceso de personalización en una situación social. En V. García Hoz (Dir.), *Tratado de educación personalizada* (vol. II, pp. 13-41). Madrid: Rialp.

- Merritt, S. M., Ako-Brew, A., Bryant, W. J., Staley, A., McKenna, M., Leone, A. & Shirase, L. (2019). Automation-Induced Complacency Potential: Development and Validation of a New Scale. *Front. Psychol.* 10(225). <https://doi.org/10.3389/fpsyg.2019.00225>
- Millán-Puelles, A. (1967). *La estructura de la subjetividad*. Madrid: Rialp.
- National Forum on Education Statistics (2006). *Forum Guide to Decision Support Systems: A Resource for Educators*. Washington, D.C.: National Center for Education Statistics.
- Noor, A. K. (2015). Potential of Cognitive Computing and Cognitive Systems. *Open Engineering*, 5(1), 75-88. <https://doi.org/10.1515/eng-2015-0008>
- Ortega y Gasset, J. (2015). *¿Qué es filosofía? y otros ensayos*. Madrid: Alianza.
- Persson, I. & Savulescu, J. (2008). The perils of cognitive enhancement and the urgent imperative to enhance the moral character of humanity. *Journal of Applied Philosophy*, 25(3), 162-177. <https://doi.org/10.1111/j.1468-5930.2008.00410.x>
- Platón (1992). Gorgias. En Platón, *Diálogos* (J. Calonge Ruiz, E. Acosta Méndez, F. J. Olivieri y J. L. Calvo, Trads., Vol. II). Madrid: Gredos.
- Roco, R. (2007). Mirando entre los intersticios de la brecha digital en la educación superior. *Calidad en la educación*, 26, 337-372. <http://dx.doi.org/10.31619/caledu.n26.244>
- Sosa, E. (2007). *A virtue epistemology: Apt belief and reflective knowledge* (Vol. 1). Oxford/New York: Oxford University Press.
- Swartz, R., Costa, A. L., Beyer, B. K., Reagan, R. & Kallick, B. (2008). *El aprendizaje basado en el pensamiento*. UE: SM.
- Urban, T. (2015). The AI Revolution: The Road to Superintelligence. En *Wait but Why*. Recuperado de: <https://waitbutwhy.com/2015/01/artificial-intelligence-revolution-1.html>
- Vásquez Rocca, A. (2009). Sloterdijk, Habermas y Heidegger: Humanismo, posthumanismo y debate en torno al *Parque Humano*. *Eikasia. Revista de Filosofía*, 26, 1-22.
- Voinea, C., Vica, C., Mihailov, E. & Savulescu, J. (2020). The Internet as Cognitive Enhancement. *Science and Engineering Ethics*, 26, 2345-2362. <https://doi.org/10.1007/s11948-020-00210-8>
- Walton, E. (2012). Using Literature as a Strategy to Promote Inclusivity in High School Classrooms. *Intervention in School and Clinic*, 47(4), 224-233. <https://doi.org/10.1177/1053451211424604>
- Wang, H. & Wang, S. (2022). Teaching Tip: Improving Student Performance by Introducing a No-Code Approach: A Course Unit of Decision Support Systems. *Journal of Information Systems Education*, 33(2), 127-134.

Zohny, H. (2015). The myth of cognitive enhancement drugs. *Neuroethics*, 8(3), 257–269. <https://doi.org/10.1007/s12152-015-9232-9>