

*To think (about) the future: Questions about  
the teaching of Philosophy and Artificial Intelligence*

*Pensar (en) el futuro:  
Cuestiones sobre la enseñanza de  
la Filosofía y la Inteligencia Artificial*



ENRIQUE ÁLVAREZ VILLANUEVA  
*Universidad de Oviedo*

*Revista Paideia 116 (2021),  
págs. 59-70. ISSN: 0214-7300*

RECIBIDO: ABRIL 2021  
APROBADO: MAYO 2021

## RESUMEN

El presente artículo tendrá como objetivo analizar críticamente el advenimiento de la Inteligencia Artificial a gran escala en el mundo de la educación, y más en particular en la enseñanza de la Filosofía. Para tal fin, se presentarán dos apartados: el primero analizará las implicaciones a corto y medio plazo, reseñando avances que ya existen o que se implementarán en los próximos años. La Filosofía aquí sufrirá para conservar su lugar en el currículum educativo. El segundo apartado formulará algunas predicciones a más largo plazo, ensayando una visión más optimista donde quizá la automatización a gran escala permita disponer de más tiempo que dedicar a esta materia en los planes de estudio.

**Palabras clave:** Inteligencia artificial, educación, Filosofía, aprendizaje inteligente, currículum educativo.

## ABSTRACT

This paper intends to critically analyze the advent of Artificial Intelligence on a large scale in the world of education, and more particularly in the teaching of Philosophy. To achieve this goal, the paper is split into two sections: the first will analyze the short- and medium-term implications, highlighting advances that have already been made or that will be implemented in the coming years. In this setting, Philosophy will struggle to preserve its place in the curriculum. The second section will formulate several longer-term predictions, attempting a more optimistic vision where perhaps large-scale automation will allow more time to devote to this subject in the curriculum.

**Keywords:** Artificial intelligence, education, Philosophy, smart learning, curriculum.

## Introducción

La Filosofía, como el arte o la literatura, ha sido siempre una disciplina que goza al mismo tiempo del respeto y el desdén del gran público. Todo el mundo habla con admiración de Aristóteles, pero pocos han leído sus obras. Peor aún, cuando se trata de dotar a la Filosofía de recursos para su enseñanza, enseguida surgen voces que claman por la necesidad de aprender cosas “más útiles” para un mundo cada vez más competitivo.

Eva Teba (2021) describe en el monográfico sobre la LOMLOE publicado recientemente por el Colegio Oficial de Doctores y Licenciados en Filosofía y Letras y en Ciencias de Madrid los objetivos que la UNESCO plantea para la educación del futuro. Estos objetivos incluyen el desarrollo personal de los educandos, la búsqueda del bien común o la necesidad de conseguir una educación integradora capaz de transformar el futuro. Todos estos objetivos deben conjugarse con una cada vez mayor incertidumbre y automatización en todos los ámbitos, que cada vez tiene más preeminencia debido a su rentabilidad económica y el convencimiento de que abrazar las posibilidades de la inteligencia artificial —en adelante, IA— es algo que conviene hacer cuanto antes, toda vez que se considera inevitable su ubicuidad en pocos años.

El presente trabajo tendrá como objetivo analizar críticamente el advenimiento de la IA a gran escala en el mundo de la educación, y reflexionar sobre cómo esta impactará a la enseñanza futura de la Filosofía. Para tal fin, se articularán dos apartados; el primero analizará las implicaciones a corto y medio plazo, reseñando avances que ya existen o que se implementarán en los próximos años. El panorama planteado aquí es bastante pesimista, ya que la Filosofía es una disciplina que se adapta con dificultad a la ideología detrás de este fenómeno. El segundo apartado formulará algunas predicciones a más largo plazo, reflexionando sobre sus implicaciones para la Filosofía como materia de educación y como disciplina *per se*, ensayando una visión más optimista donde quizá la automatización a gran escala permita disponer de más tiempo que dedicar a esta materia en los planes de estudio.

## 1. El corto plazo: innovaciones técnicas y horizonte incierto

El hecho de que la tecnología puede ser muy útil para democratizar el conocimiento y hacerlo más cómodo y asequible es algo evidente. Internet, por ejemplo, ha permitido que muchos estudiantes puedan acceder a muchísima información sin necesidad de desplazarse a una biblioteca y limitarse a los recursos que esta pudiera ofrecerles. Al mismo tiempo, es preciso también darse cuenta de que los servicios que la tecnología ofrece no son perfectos ni están exentos de externalidades negativas que pueden quedar ensombrecidas bajo la luz de la eficacia y la comodidad.

### 1.1. La inteligencia artificial entra en el aula.

Los planes de implementación de sistemas educativos inteligentes (*smart*) llevan sobre la mesa desde que Sidney Pressey, y especialmente B. F. Skinner, el padre del conductismo, diseñasen los primeros sistemas automáticos de evaluación que proporcionaron indicaciones a los estudiantes sobre sus errores hace unos 50 años. Mucho después, en 1997, Malasia lanzó su *Smart Education Project*, aunque la propuesta no alcanzó una madurez y difusión notables hasta el advenimiento de refinamientos tecnológicos, como el 5G, la inteligencia artificial o el Internet de las cosas (*Internet of Things*), más de 10 años después (Lee, 2020, p. 302). Gigantes tecnológicos como Microsoft o Google, entre muchos otros, han observado la tendencia creciente de los gobiernos a apostar por modelos educativos cada vez más imbricados con la tecnología, provocando un boom de la inversión en este apartado, hasta el punto de que se prevé que sea un mercado de 6.000 millones de dólares para 2024 (Holmes, Bialik y Fadel, 2019, p. 81).

Desde los primeros diseños de máquinas para la educación, el *leit motiv* fue procurar herramientas a los estudiantes para personalizar su aprendizaje, haciendo posible que estos siguieran el ritmo que mejor se ajustara a sus capacidades y preferencias. La rigidez de los primeros modelos fue flexibilizada gracias a la llegada de la inteligencia artificial, que introdujo la posibilidad de acceder a grandes bases de datos que continuamente se actualizaran con las respuestas de los estudiantes, además de ser capaces de presentar los conocimientos de forma más pedagógica y entretenida. Algunos de estos sistemas siguen un modelo denominado socrático, basado en preguntas pertinentes que hagan reflexionar

a los estudiantes, como el sistema Watson de IBM; otros ensayan entornos más abiertos donde el estudiante explore y manipule elementos que se le proponen para construir su conocimiento, como en el caso de los sistemas Betty's Brain y ECHOES, sistema este último diseñado para trabajar con niños autistas usando una arquitectura FATiMA, que tiene en cuenta factores de respuesta y producción emocional que han dado buenos resultados en niños con esta condición (Holmes et al., 2019, p. 126). Los sistemas más avanzados utilizan chatbots, sistemas que pueden conversar mediante el reconocimiento del lenguaje natural y la producción de voz al estilo de Alexa o Siri, además de reconocimiento facial, que permite percibir si los estudiantes se aburren o están estresados. Los proyectos a futuro son mucho más ambiciosos, introduciendo elementos como la realidad virtual o los profesores holográficos, con la intención de convertir el aula en un “ambiente de aprendizaje inteligente” (*smart learning environment*) (Lee, 2020, p.311).

Por supuesto, la implantación de estos sistemas no está exenta de problemas. El primero de ellos es que, aunque la inteligencia artificial promete ser un sistema mucho más eficiente e innovador, lo cierto es que no hay resultados claros de la efectividad de estos sistemas comparados con el modelo de educación tradicional. Lejos de eso, el modelo Summit Learning, un proyecto pionero de aprendizaje *smart* desarrollado por Facebook, ha recibido muchas críticas por ser aburrido, monótono y por eliminar el contacto con otros seres humanos. El sistema de Facebook no ha sido aún probado extensamente, pero en lugares como China, donde propuestas similares son utilizadas, los estudiantes dan muestra de preferir el modelo tradicional, ya que la continua conexión y vigilancia que estos sistemas ejercen causan notables niveles de ansiedad entre los estudiantes (Holmes et al., 2019, p. 175). En el horizonte temporal cercano, otra consecuencia es que los alumnos deberán adaptarse a la forma en que estos sistemas funcionan, usando pantallas y auriculares, aislándose de otros estudiantes y generando contenidos escritos y orales de una forma que el sistema pueda procesar, lo cual va en contra de la prometida adaptación al estudiante en el que las personas *van primero* que las máquinas, y más en consonancia con la idea de Langdon Winner (1977) de que la sociedad se adapta a la tecnología antes que esta a aquella.

Un factor adicional es el hecho de que son estos sistemas los que eligen el contenido que se les enseña a los estudiantes, ya que, en teoría, la inteligencia artificial es buena cribando y decidiendo qué es mejor estudiar en cada mo-

mento para cada alumno. Esto elimina parte de la libertad tanto del alumno para decidir qué le interesa aprender como del profesor para seleccionar los materiales según su criterio.

Al permitirles operar sin supervisión, varios sistemas de inteligencia artificial han mostrado ya en repetidas ocasiones sesgos xenófobos y sexistas que han adquirido de la información vertida en la nube por millones de usuarios sin haber sido filtrada (Wolf, Miller y Grodzinsky, 2017, Bartneck, Lütge, Wagner y Welsh, 2020, p. 35). Exponer a los estudiantes a recibir su educación sin un filtrado lo suficiente sofisticado podría tener implicaciones muy negativas. Incluso con estos filtros, una paradoja aparece, y es que pese a que una de las ventajas que se barajan en estos sistemas es la adaptación a las preferencias y necesidades de cada estudiante, los parámetros que las máquinas manejen necesariamente tendrán como horizonte regulador una media de rendimiento considerada adecuada. Si los objetivos que enfrentan todos los estudiantes son comunes, y están mediados por las demandas de una sociedad del futuro cada vez más globalizada y centrada en la productividad económica (Cardon, 2018), ¿realmente podemos esperar una personalización real en su educación atendiendo a criterios no económicos, sino de desarrollo de la personalidad y la individualidad del estudiante?

## 1.2. ¿Encaja la Filosofía en este planteamiento?

Por el momento, además, estos sistemas son solo efectivos en materias que puedan ser sistematizadas de forma efectiva, especialmente las ciencias más formales como las Matemáticas o la Física. Es difícil concebir cómo las Humanidades, y especialmente la Filosofía, disciplina de segundo nivel que requiere de un análisis extensivo de otras categorías del saber para desarrollar su actividad propia, pueden encajar aquí.

Al inspeccionar la literatura que reseña la introducción de la IA en la educación, los principios rectores que hallamos son cosas como “aprender a aprender”, “aprender a pensar críticamente”, “considerar cuestiones éticas”, etc., y una mención explícita a que estas cuestiones son mucho más importantes, por ejemplo, que aprender que la mitocondria es la central energética de la célula (Holmes et al., 2019, p.11). En este sentido, se valora mucho más la adquisición de “conceptos centrales” (*core concepts*) que puedan relacionarse de forma inter-

disciplinar y que faciliten la comprensión de forma sinérgica. Un ejemplo de esto sería el concepto de la “división del trabajo” en Economía, que podría aplicarse, por ejemplo, a la Fisiología animal, entendiendo los distintos órganos como elementos que se organizan para sacar el mayor partido de las capacidades de cada miembro del grupo individualmente. El *Center for Curriculum Re-design* —una organización global sin ánimo de lucro cuyo principal objetivo es la mejora de la educación para el siglo XXI—, al igual que el programa *SMART* propuesto por el ministerio de educación de Corea del Sur, entre otras muchas propuestas, aspiran a introducir en el currículum de manera explícita cuestiones como el bienestar o la educación cívica y global. Lo que no está del todo claro es cómo va a conjugarse esto con la propuesta de una mayor automatización del aprendizaje impartido por sistemas de IA, y especialmente cómo es posible desarrollar tales ideas poniendo el énfasis en conceptos centrales y en la interdisciplinaridad, y no en la *metadisciplinaridad*, en un enfoque de análisis de las disciplinas que es propio de las Humanidades, y, en especial, de la Filosofía. Entendemos así que la interdisciplinaridad es un enfoque interesante de aprendizaje si se pretende crear redes semánticas con el fin de dominar conceptos difíciles en disciplinas como la Biología o la Física, pero esto poco tiene que ver con el meta análisis filosófico, que tiene por objetivo sacar conclusiones de segundo nivel y que precisa no de examinar las relaciones de conceptos de las distintas categorías del saber, sino de cuestiones acerca de esas categorías para generar planteamientos que no pueden responderse solo con el conocimiento de los conceptos de las mismas.

Conjugando esto con los datos de desaparición acelerada de trabajos que requieren una formación media (*middle-skill jobs*), como el manejo de maquinaria industrial o las tareas de clasificación de productos en ventas a favor de la automatización, es difícil pensar cómo en un panorama en el que la Filosofía se ve ya como algo accesorio (uno de los motivos de la retirada de la Ética en la ESO en España ha sido precisamente el no “tener un hueco” en un currículum ya saturado), esta vaya a encontrar acomodo cuando la automatización sea mucho más ubicua y agresiva, y la necesidad de adaptación constante de los empleados sea la norma.

El creciente malestar por la precariedad laboral, el mayor coste de la educación, el menor acceso a becas y oportunidades y el pesimismo imperante de cara al futuro está siendo también aprovechado por nuevos actores para enarbolar una suerte de “postmodernidad educativa”, donde los grandes paradigmas

de la escuela tradicional son puestos en cuestión por nuevos métodos que prometen ser más adecuados para el siglo XXI. Métodos alternativos a la educación tradicional, como la pedagogía Waldorf o Montessori, llevan ya muchas décadas entre nosotros, pero aquí nos referimos específicamente a ideas como la llamada Universidad Google, que ofrece cursos de 4 meses con un coste de 300 dólares que, en teoría, equivaldrían a un grado de 4 años, al menos en lo que se refiere a obtener competencias para desarrollar un puesto de trabajo (Pastor, 2020). Esta propuesta, lejos de ser una curiosa excepción, parece ir en consonancia con los nuevos modelos corporativos, en los que se espera que el trabajador tenga una continua formación y complemente sus competencias con cursos constantes. Esto difícilmente casa con un enfoque humanístico, ya que lo imperativo es la adquisición de competencias técnicas, no de capacidades reflexivas de corte filosófico.

## **2. El largo plazo. ¿El retorno del *otium cum dignitate*?**

Max Tegmark (2017, p. 125) ilustra la común preocupación de las personas sobre el futuro del mundo del trabajo tras la creciente automatización que vivimos citando una hipotética conversación entre dos caballos que tiene lugar en 1900, imaginada originalmente por el economista Gregory Clark. En ella, un caballo explica a otro que está preocupado por su futuro laboral [sic]. El otro caballo le contesta que, con la invención del motor de vapor, sus antepasados estaban preocupados porque sus servicios ya no serían necesarios para accionar las ruedas en los molinos, habiendo ganado calidad de vida con la dedicación exclusiva al transporte, que además es mucho más comfortable como profesión. Cuando el otro caballo le pregunta si piensa que el motor de combustión interna podría ser un problema de futuro, el equino optimista le responde que está seguro de que, como en el pasado, habrá muchos trabajos nuevos para ellos. Tegmark, a continuación, reseña que de una población estimada de 26 millones de caballos en los Estados Unidos en 1915 se pasó a una de 3 millones en 1960.

La encrucijada en la que nos encontramos ahora tiene ciertas similitudes con la planteada en esta conversación, con la salvedad de que la población humana no tiene visos de dejar de aumentar en el futuro. Con respecto a la cantidad de empleos disponibles, el panorama, sin embargo, es menos esperanzador. La



Universidad de Oxford sostiene que en el futuro cercano un 90% de los trabajos existentes será realizado por inteligencias artificiales. El Nomura Research Institute, además, calcula que el 49% de trabajos actuales desaparecerá en 15 años (Haque, 2020, p. 207). Ante esta situación, la posibilidad de dejar que las máquinas generen la productividad y se implante una renta básica universal e incondicional que permita a las personas consumir para continuar con el sistema económico está encontrado cada vez más adeptos y planteamientos más serios y elaborados (Wehner, 2019, Yang, 2018).

La otra opción es, por supuesto, que esta cuarta revolución industrial genere por sí misma una plétora de nuevos trabajos que vengan a sustituir a los perdidos por la automatización, pero lo cierto es que la ambición que tienen los investigadores sobre el futuro de la tecnología, incluso los más escépticos, pasa por tender a una sustitución de las personas desde las cuestiones más automáticas y repetitivas, como mover cajas en almacenes de grandes superficies, a sustituir a nuestros psicoterapeutas o profesores, como se ha apuntado antes. Muchos investigadores persiguen el objetivo final de crear una inteligencia artificial igual de potente que la humana, con sus mismas capacidades cognitivas, emocionales y de encarnación (*embodiment*) (Burden y Saben-Baden, 2019, p. 17), y se teoriza sobre la cada vez más real posibilidad de que se integren en nuestra sociedad como amigos y amantes (Levy, 2008). Incluso si se lograsen solo la mitad de los objetivos que los más ambiciosos persiguen, tendríamos robots y sistemas de IA lo suficientemente sofisticados como para ocuparse de todas las labores que no necesitasen de las habilidades más difíciles de emular, como la empatía, a un coste mucho menor que el que conlleva emplear a seres humanos (Russell, 2019, pág. 119).

Si este escenario se hace realidad, es de esperar que haya cambios muy notables en el sistema educativo. Todavía será necesario formar a los estudiantes en las competencias que les sean útiles para su desempeño laboral en el futuro, pero el cada vez menor número de puestos de trabajo disponibles hará que buena parte del enfoque educativo futuro pase por dar a estos estudiantes la capacidad para vivir en un mundo en el que, por primera vez en la historia, la mayoría de la gente no *tiene que* buscar un trabajo con el que mantener su nivel de vida a toda costa, y puede usar su acceso a más tiempo y a la asistencia económica del Estado para buscar actividades que aporten un mayor valor añadido a su vida, y que le ayuden a vivir en un mundo muy diferente del actual (Yang, 2018, p. 103).

Pese a que estas líneas debaten un futuro que no está para nada garantizado, todavía hay algunos motivos para no ser totalmente pesimistas con respecto a la Filosofía en un futuro lejano, y es que el interés por las cuestiones éticas y de concepto relativas a las nuevas tecnologías es cada vez más común, debido al potencial transformador de estas (Wallach y Allen, 2009, Coeckelbergh, 2021). Así, no es disparatado pensar que la posibilidad de estar menos condicionados por la continua adaptación a los requerimientos del mercado de trabajo, sumado a la no necesidad de desarrollar trabajos puramente de subsistencia y el verse enfrentado con el debate perenne y social sobre los retos éticos de cada momento podría suponer de algún modo un escenario propicio para el desarrollo del ocio valioso del que Aristóteles hablaba en su *Ética nicomaquea* (117b), que después se repetiría a lo largo de la historia del pensamiento de la mano de muchos autores, como el *otium cum dignitate* de Cicerón o de Montaigne. Este ocio no tendría nada que ver con los pasatiempos banales, sino con la reflexión y la meditación sobre cuestiones importantes y complejas, lo que requiere tener tiempo y energía disponibles y una mente bien educada. La mejor manera de practicar este *otium* sería desarrollando las habilidades de pensamiento crítico desde la escuela, y quizá esta perspectiva pudiera encontrar acomodo en un currículum menos saturado de contenidos que ahora se entienden como “más prácticos”.

Quizá, por tanto, una escuela menos constreñida por el tiempo y los objetivos de corte económico esté en una mejor disposición de cultivar el humanismo tras sobrevivir a un duro invierno más a corto plazo. Por supuesto, todo esto depende de muchos factores que han dilucidarse en el futuro cercano. Por lo tanto, y por el momento, la misión fundamental será hacer que la Filosofía no desaparezca de la educación antes de que su momento llegue finalmente.

## REFERENCIAS BIBLIOGRÁFICAS

- Aristóteles (1985). *Ética nicomaquea- Ética eudemia*, Madrid, Gredos.
- Bartneck, C., Lütge, C., Wagner, A. y Welsh, S. (2020). *An Introduction to Ethics in Robotics and IA*, Cham: Springer.
- Burden, D. y Savin-Baden, M. (2019). *Virtual Humans Today and Tomorrow*, Nueva York: CRC Press.
- Cardon, D. (2018). *Con qué sueñan los algoritmos. Nuestras vidas en el tiempo de los big data*, Madrid: Dado ediciones.
- Coeckelbergh, M. (2021) *Ética de la Inteligencia Artificial*, Madrid: Cátedra.
- Haque, E (2020). *The Ultimate Modern Guide to Artificial Intelligence*. Londres: Amazon Inc.
- Holmes, W., Bialik, M., Fadel y Ch. (2019). *Artificial Intelligence in Education. Promises and Implications for Teaching & Learning*. Boston: Center for Curriculum Redesign.
- Lee, R.S. (2020). *Artificial Intelligence in Daily Life*. Singapur: Springer.
- Levy, D. (2008). *Love and Sex with Robots*, Wiltshire: Cromwell Press Ltd.
- Pastor, J. (6 de agosto de 2020) Universidad Google: por 300 dólares y en seis meses obtendrás un certificado que ellos equiparan al de una carrera universitaria. *Xataka*. Recuperado de: <https://www.xataka.com/empresas-y-economia/universidad-google-300-dolares-seis-meses-obtendras-certificado-equivalente-a-carreras-universitarias-cuatro-anos>
- Russell, S. (2019) *Human Compatible. AI and the Problem of Control*, Londres: Penguin.
- Teba Fernández, E. (2021). Los futuros de la educación. *Colegio Oficial de Docentes. Colegio Oficial de Doctores y Licenciados en Filosofía y Letras y en Ciencias de Madrid*, 294, 50-52, Consultado en: <https://www.cdlmadrid.org/wp-content/uploads/2021/01/022021.pdf>
- Tegmark, M. (2017). *Life 3.0*, Westminster: Penguin.
- Wallach, W. y Allen, C. (2009) *Moral Machines. Teaching Robots Right From Wrong*, Nueva York: Oxford.

Wehner, B. (2019) *Universal Basic Income and the Reshaping of Democracy. Towards a Citizens' Stipend in a New Political Order*, Cham: Springer.

Winner, L. (1977). *Autonomous Technology: Technics-out-of-control as a Theme in Political Thought*. Cambridge: MIT Press.

Wolf, M.J, Miller, K., Grodzinsky y F.S. (2017). Why We Should Have Seen That Coming. Comments on Microsoft's Tay "Experiment" and Wider Implications, *ACM Computers & Society*, 47, (3), 54-64.

Yang, A. (2018) *The War on Normal People. The Truth About America's Disappearing Jobs and Why Universal Basic Income is our Future*, Nueva York: Hachette.