



Contener a Silicon Valley y otros empeños difíciles

por **Antonio Diéguez**

El verdadero peligro no es la IA: es la élite que la controla. Agita el miedo al apocalipsis no para que se desarrolle una regulación, sino para que se le deje actuar sin ella.

Al hombre medio se le ha hecho hoy su propia vida menos transparente que lo que la suya era al hombre en otros tiempos. La técnica cuya misión es resolverle al hombre problemas se le ha convertido de pronto en un nuevo y gigantesco problema.

Ortega y Gasset, J. Meditación de la técnica, 1939.

Entender la tecnología, es decir, entender la tecnosfera en la que vivimos, es una exigencia ineludible, entre otras razones porque cada vez es más evidente que el futuro de la tecnología, y muy especialmente el de la inteligencia artificial, va ligado al futuro de la democracia e incluso al futuro de la civilización. La ciencia y la tecnología, como admitía el politólogo Francis Fukuyama, se han convertido en el principal motor de la historia. El filósofo de la tecnología Langdon Winner hizo mucho por convencernos hace décadas de que lejos de ser neutrales,

como habitualmente se cree, los artefactos tienen política. Hoy podemos decir con muchos más datos para apoyarlo que los algoritmos tienen política que ha sido inducida en ellos por sus creadores, lo que de forma efectiva es tanto como decir por los directivos de las grandes tecnológicas de Silicon Valley.

Desde los comienzos de la modernidad, y así quedó de manifiesto en las diferentes utopías que se forjaron, el futuro ha sido un asunto de preocupación social y su posible configuración ha marcado las agendas de los movimientos políticos. El problema es que hoy ese futuro es más opaco que nunca. Se ha vuelto aún más impredecible que antes debido al desarrollo de la tecnología, y, pese a ello, nos preocupa sobre todo el medio y el largo plazo, un espacio temporal para el que las deliberaciones democráticas y los análisis científicamente fundados son particularmente miopes. Como cabía esperar en una situación así, la charlatanería ha encontrado

un campo propicio y, por ejemplo, empieza a ser difícil tomarse en serio mucho de lo que se dice y se publica sobre el futuro de la IA. Abunda una subliteratura aparentemente especializada que es mejor evitar si no se quiere perder el tiempo con elucubraciones que inducen a la confusión. Por eso, conviene hacer una selección de aquellos textos que puedan ofrecer alguna guía, y, entre los más recientes, he elegido tres que me parecen interesantes acerca del debate en torno a las implicaciones de la IA. Estos tres libros son: *Inteligencia artificial. Guía para seres pensantes*, de Melanie Mitchell (Capitán Swing), *La república tecnológica. Poder duro, pensamiento débil y el futuro de Occidente*, de Alexander C. Karp y Nicholas W. Zamiska (Tenos), y *Frenar a Silicon Valley*, de Gary Marcus (Shackleton Books). Cada uno de ellos representa, como veremos, un aspecto central de ese debate.

Si hemos de hacer caso a lo que se anuncia con insistencia desde los centros de opinión en tecnología, estamos empezando a librar la partida final, el juego en el que ya se decide no solo la transformación sustancial de las condiciones materiales y culturales de nuestra existencia (algo que se da ya por descontado que la IA producirá, aunque la magnitud es asunto controvertido), sino incluso nuestra supervivencia como especie o nuestra conversión en algo muy superior. Si una superinteligencia artificial es creada en las próximas décadas y esta adquiere autonomía en sus fines, ya no podrá ser controlada por ningún ser humano y la disparidad que cabe esperar entre los fines que ella desarrolle autónomamente y los de nuestra especie será tal que esto conduciría casi con seguridad a nuestra extinción. El asunto así planteado es de tal gravedad que no es extraño que se haya extendido la ansiedad por saber si hay base sólida para aceptar estas ideas o si se trata de un catastrofismo infundado, y de ahí que proliferen los artículos, los libros y las opiniones sobre el asunto. Lo cierto es que la inquietud crece en la población y, según las encuestas, cada vez hay más recelos entre los ciudadanos, aunque no solo por estas razones, ante el despliegue de la IA en la vida cotidiana. Cunde sobre todo el miedo al deterioro del mercado laboral, de nuestras condiciones de vida y el miedo al reemplazo. Según alguna encuesta que cita el libro de Marcus (p. 188), el 68% de los encuestados consideró que la IA debería ser tratada como una tecnología “increíblemente poderosa y peligrosa”. Del entusiasmo y la esperanza ligados a las promesas de la IA antes del *deep learning*, se ha pasado con rapidez al temor y la desconfianza.

El problema no es tanto la IA considerada en sí misma (ni siquiera necesariamente una futura inteligencia artificial general, si es que fuera posible). Tampoco lo es, al menos por el momento, quién y cómo la usa (aunque haya en esto mucho de lo que preocuparse también). Lo que en este momento resulta más inquietante es más bien quiénes y cómo la están gestando y con qué objetivos. Estamos presenciando con nitidez los efectos nocivos de dejar la investigación de vanguardia y las directrices generales del desarrollo



Melanie Mitchell
INTELIGENCIA ARTIFICIAL. GUÍA PARA SERES PENSANTES
Traducción de María Luisa Rodríguez Tapia
Madrid, Capitán Swing, 2024, 360 pp.



Alexander C. Karp y Nicholas W. Zamiska
LA REPÚBLICA TECNOLÓGICA: PODER DURO, PENSAMIENTO DÉBIL Y EL FUTURO DE OCCIDENTE
Traducción de Francesc Pedrosa
Barcelona, Editorial Tenos, 2025, 323 pp.



Gary Marcus
FRENAR A SILICON VALLEY. CÓMO LAS BIG TECH SE APROVECHAN DE NOSOTROS Y CÓMO LA INTELIGENCIA ARTIFICIAL PUEDE EMPEORARLO
Traducción de Marc Figueras
Barcelona, Shackleton Books, 2025, 255 pp.

tecnológico venidero en las manos de directivos de empresas privadas que han conseguido establecer una fructífera alianza de apoyo mutuo con el poder político ahora establecido en la Casa Blanca, hasta el punto de que Trump les ha cedido a estas empresas el desarrollo de sistemas militares que hasta el momento eran asunto exclusivo del Pentágono. Es como si los misiles nucleares más potentes fueran propiedad de empresas privadas y ellas conservaran el control sobre ellos e incluso tuvieran en sus manos los maletines con los botones nucleares.

Todo parece encaminado a que el futuro de la humanidad lo decidan apenas una decena de personas, los directivos de las grandes compañías de IA, pero resulta que el futuro que nos presentan es el de un más que probable apocalipsis, el de un inevitable final de nuestra especie. Porque hay aquí algo que no conviene descuidar: si pasamos la prueba del riesgo existencial de una superinteligencia artificial no alineada con nuestros valores, el futuro brillante que habría detrás de ese momento, al que llaman “Singularidad”, no será un futuro para nuestra especie, sino para una especie poshumana de la que poco sabemos, excepto que supuestamente será muchísimo mejor que la nuestra. Ellos hablan del futuro de la humanidad, pero bajo ese concepto incluyen de manera principal a los innumerables poshumanos de existencia probablemente virtual que habrá en el futuro.

Sin embargo, pese a confesar que lo que están haciendo va dirigido a la sustitución del ser humano por otra entidad surgida de la unión con las máquinas, no quieren que nadie les moleste en su tarea, por si acaso consiguen salvarnos durante un tiempo (tampoco demasiado, porque es importante no retrasar al poshumano, que es lo que tiene auténtico valor). Infunden miedo para que confiemos en ellos, ya que solo ellos conocen la mejor salida posible,

y para que así les dejemos prepararla en completa libertad, aunque esto signifique que la libertad de los demás les importe muy poco. Tratan, en suma, de evitar el control de la tecnología fomentando el temor y la sensación de impotencia (“no se le pueden poner puertas al campo”, “no se puede impedir el progreso”), de modo que puedan asumir ellos todo el control. Por eso Europa les estorba y se ha convertido para alguno en particular en una obsesión. Su normativa legal para regular la IA es un mal ejemplo para el mundo y es necesario acabar con ella.

Peter Thiel, el cofundador de PayPal y CEO de Palantir, es probablemente el que ofrece hoy un discurso más alucinatorio. Defiende en sus conferencias que el Anticristo llegará en la forma de un gobierno o un Estado mundial que intente controlar la tecnología. No deja de ser una ironía que el empresario que ha desarrollado la más poderosa empresa de vigilancia jamás habida considere un mal apocalíptico que pretendan vigilarle a él.

Por primera vez en la historia del capitalismo, la élite empresarial de la industria más potente y exitosa, cuando habla de progreso, habla realmente de un final, de nuestro final, y es lo que anhelan sin tapujos. Quieren que llegue lo antes posible la Singularidad, porque desean el mundo que habrá después y que ellos confían en ver gracias a los avances en la medicina antienvjecimiento y a sus precauciones para que la superinteligencia artificial no nos sea hostil. Pero si llega la Singularidad en los términos que proclaman, ningún ser humano podrá beneficiarse, porque al cabo de poco no quedará ninguno para hacerlo. Como en algunas visiones religiosas del fin de los tiempos, se nos presenta el final como un verdadero comienzo, solo que un comienzo en el que ya no estará nuestra especie. El progreso, según esto, es desaparecer para dejar nuestro lugar a algo que ellos estiman como superior.

Hay siempre un buen número de personas que encuentran interesante escuchar que el fin del mundo está cerca. Al parecer, ese discurso las saca durante un tiempo de una vida aburrida o de problemas cotidianos demasiado difíciles como para bregar con ellos todo el rato. El discurso apocalíptico, este anuncio de algo parecido a una segunda venida de Jesús, cuenta de hecho con una alta popularidad en los Estados Unidos, mucho más alta que en Europa. Según un estudio del Pew Research Center publicado en 2022, el 39% de los estadounidenses cree que “estamos viviendo los últimos tiempos”, una creencia extendida sobre todo en algunas iglesias protestantes. Uno de cada diez encuestados cree que la segunda venida de Jesús ocurrirá durante su vida.

Ahora bien, creer que el apocalipsis está cerca no es una posición muy adecuada para lograr una buena orientación para la acción política y social, y lo que justamente ahora más necesitamos es ánimo sereno y una visión equilibrada del desarrollo tecnológico. Muchos pensamos, además, que es demasiado pronto para dar por finalizado al ser humano.

En tal contexto, todo lo que nos dicen los grandes magnates tecnológicos acerca de que su gran preocupación y objetivo prioritario es la seguridad de la IA es sencillamente intranquilizador. En la carrera desenfrenada en la que se ha convertido la investigación en IA, los que intenten prestar atención a la seguridad y a las consecuencias peligrosas de los sistemas de IA tienen las de perder. Sus competidores menos escrupulosos sacarán antes sus productos al mercado. Las cuestiones de ética y de seguridad están en la boca de todos, pero no tienen ninguna relevancia real en los resultados. A menudo son un simple recurso para lograr mejor imagen ante el público. Nadie se las toma muy en serio.

¿Cómo es posible, cabe preguntarse ante todo esto, que nadie haga nada frente a quienes nos aseguran que trabajan para nuestra pronta obsolescencia y desaparición? ¿Cómo es posible que muchos sigan asumiendo que su petición de manos libres y ausencia de regulación es el mejor camino para un futuro próspero? No se me malinterprete. No estoy diciendo que debamos creernos esos anuncios apocalípticos. Lo que debería preocuparnos, en mi opinión, no es tanto que este discurso catastrofista se cumpla (las posibilidades son remotas por el momento), como el hecho mismo de que se produzca y los efectos que él mismo está causando.

La economista y expresidenta del Instituto de Innovación italiano Francesca Bria ha llamado con buenas razones a esta élite empresarial “el complejo tecnológico autoritario”. Y basta con leer una entrevista a Elon Musk, o un texto de Peter Thiel o Curtis Yarvin, para entender al instante que es una absoluta temeridad el que estas personas sean las que están tomando las decisiones que marcan buena parte de la economía actual y que orientarán el futuro de la tecnología.

Aunque estos discursos apocalípticos causan gran impacto, su objetivo principal, como decimos, además de publicitar ante los inversores su inmenso poder, es que el temor lleve a la parálisis y que esa parálisis se convierta en completa libertad de acción para las élites tecnológicas. No obstante, buscan también desviar la atención de otros problemas reales. Entre ellos, la aplicación creciente de estas máquinas en la toma de decisiones importantes en sustitución del ser humano. Como bien señalan Dan Hendrycks, Mantas Mazeika y Thomas Woodside (en “An overview of catastrophic AI risks”):

A medida que los retos de la sociedad se vuelven cada vez más complejos y acelerados, y que las IA se vuelven cada vez más inteligentes y rápidas, es posible que les cedamos cada vez más funciones por comodidad. En tal situación, la única solución viable a las complejidades y retos que plantean las IA podría ser depender aún más de ellas. Este proceso gradual podría conducir finalmente a la delegación de casi todo el trabajo intelectual, y eventualmente físico, a las IA. En un mundo así, las personas podrían tener pocos incentivos para adquirir conocimientos y cultivar habilidades, lo que podría conducir a un estado de debilitamiento.

Ese es el peligro real ante el que estamos. Un peligro indisoluble del enorme poder político y económico que las grandes empresas tecnológicas están adquiriendo. Y lo cierto es que muchas promesas positivas que se hicieron se ven ya claramente como incumplidas. La IA no ha generado el aumento espectacular de productividad que se decía ni ha ahorrado tanto trabajo, y para lo que está sirviendo más bien es para usos militares, para ejercer una vigilancia creciente sobre los ciudadanos, para amenazar la información veraz, para favorecer la delincuencia común y para dificultar algunas tareas que se hacían suficientemente bien sin ella.

El mensaje de estos libros es una buena ayuda para formar una opinión sobre la verborrea desorientadora y angustiosa en la que se han convertido las noticias sobre la IA. Los tres libros dan una señal de que las cosas no van bien, pero disienten en los análisis del problema y, sobre todo, en lo que puede hacerse al respecto.

Inteligencia artificial. Guía para seres pensantes es una excelente introducción al campo de la inteligencia artificial escrita por una de sus más acreditadas investigadoras, Melanie Mitchell, catedrática Davis de Complejidad en el Instituto de Santa Fe y Premio Herbert Simon en 2020. Mitchell sabe bien de lo que habla, puesto que lleva años trabajando en zonas de gran interés en la IA, como el razonamiento analógico, los algoritmos genéticos y los autómatas celulares. El libro dedica la mayor parte de su contenido a explicar de forma detallada, pero muy clara, la historia y los fundamentos de la inteligencia artificial como ámbito de investigación, desde la vieja inteligencia artificial simbólica a las actuales redes neuronales artificiales, a las que, como es lógico, presta mayor atención, así como a las diferentes modalidades de su aprendizaje. Particularmente interesantes son los apartados que dedica a las debilidades que las redes neuronales profundas manifiestan en el reconocimiento de imágenes.

El lector encontrará aquí reflexiones sensatas y bien articuladas sobre preguntas clásicas, como si puede pensar una máquina, y sobre otras más recientes, como la posibilidad real de que acontezca la Singularidad y la creación de una superinteligencia artificial dañina; y si tiene inclinaciones filosóficas, como es mi caso, disfrutará en especial de la parte quinta, dedicada a la cuestión del significado y de la comprensión. Una frase que resume bien la posición de Mitchell en estas discusiones es la siguiente: “En mi opinión, se ha prestado demasiada atención a los riesgos de la IA superinteligente y muy poca tanto a la falta de fiabilidad y transparencia del aprendizaje profundo como a su vulnerabilidad a los ataques” (p. 162). Las capacidades de abstracción conceptual y de razonamiento por analogía propias del ser humano siguen estando, según Mitchell, fuera del alcance de nuestras máquinas, entre otras razones porque son capacidades que requieren la posesión de sentido común y la comprensión del significado, y nadie sabe cómo conseguir que las máquinas tengan ambas cosas. Mitchell aventura incluso que las

“Si Karp y Zamiska quieren poner a Silicon Valley en la vanguardia de la **defensa de los Estados Unidos** y su cultura, Marcus cree, por el contrario, que lo que hay que hacer es **contener a Silicon Valley.**”

limitaciones que nuestro cuerpo impone a nuestra inteligencia (como las emociones y prejuicios de probable origen evolutivo) son fundamentales para conseguir una auténtica inteligencia general. Por eso, ya casi al final del libro, su vaticinio no alegrará a los tecnoentusiastas: “La IA actual —escribe— está muy lejos de la inteligencia general, y no creo que la ‘superinteligencia’ de las máquinas esté en el horizonte. Si la IA general se hace algún día realidad, estoy segura de que su complejidad será equiparable a la de nuestro cerebro” (p. 350). Lo que a Mitchell le preocupa es más bien que estemos delegando decisiones cada vez más importantes en las máquinas.

El libro de Karp y Zamiska *La república tecnológica* se ha convertido en poco tiempo en un éxito de ventas y ha generado una intensa discusión en redes sociales al haber sido resumido por Palantir, la empresa que dirige Karp, en veinte puntos en forma de manifiesto. De hecho, lo sustancial del libro está ahí, de modo que, leído el manifiesto, puede uno casi ahorrarse la lectura del libro, que, la verdad sea dicha, no añade demasiado al asunto. El *Wall Street Journal* lo ha descrito como “un grito de corazón que critica a la industria tecnológica por haber abandonado su tradición de ayudar a Estados Unidos y a sus aliados”. Es una buena descripción, porque, en efecto, el libro es una llamada ideológica a la asunción de responsabilidades en los usos militares y geoestratégicos por parte de las grandes industrias tecnológicas de Silicon Valley. Los autores creen que hay que defender a Occidente de las amenazas que lo cercan y la alta tecnología, con la IA a la cabeza, no puede ya mirar para otro lado. Debe implicarse en la guerra (por ahora, sobre todo económica, política y cultural). Al igual que en 1945 el informe de Vannevar Bush *Ciencia, la frontera sin fin* puso las bases teóricas

del reclutamiento de la tecnociencia norteamericana al servicio de la Guerra Fría, el libro de Karp y Zamiska busca hacer algo parecido adaptándolo a nuestro tiempo. Ahora el rival principal no es el comunismo soviético sino el capitalismo de Estado chino y sus aliados. Ellos lo dicen sin recovecos:

La élite de ingenieros de Silicon Valley tiene la obligación de participar en la defensa de la nación y en la articulación de un proyecto nacional —qué es este país, cuáles son sus valores y qué es lo que defendemos— y, por extensión, de preservar la perdurable, aunque frágil, ventaja geopolítica que Estados Unidos y sus aliados en Europa y otros lugares han mantenido sobre sus adversarios (p. 17).

Por si había dudas, añaden páginas después:

Nuestro argumento esencial es que, en esta nueva era de la IA avanzada, que ofrece a nuestros adversarios geopolíticos la oportunidad más convincente desde la última guerra mundial para desafiar nuestra posición global, deberíamos volver a esa tradición de estrecha colaboración entre la industria tecnológica y el gobierno (p. 37).

Y ya para los menos espabilados aclaran: “Si un marine estadounidense pide un fusil mejor, debemos construirlo. Y lo mismo puede decirse del software” (p. 182).

Es interesante el dato de que tanto Karp como su colega en la dirección de Palantir, el ya mencionado Peter Thiel, se formaron como filósofos, no como ingenieros o científicos. Desafortunadamente, la filosofía, que no es tan inútil como muchos creen, está teniendo un papel extraño y peligroso en todo esto. En algunos breves párrafos del libro se critica la filosofía posmoderna, pero no para volver a la razón ilustrada, sino para realzar el papel que debería volver a tener la religión. Los ideales de la Ilustración quedan lejos del libro.

Gary Marcus, científico cognitivo y empresario, es bien conocido entre los especialistas por sus críticas a las promesas desmedidas de los líderes tecnológicos acerca de la inminencia de una inteligencia artificial general (AGI). Su libro con Ernest Davis, *Rebooting AI*, fue un aldabonazo para muchos que se dejaban llevar por los discursos entusiastas emanados de Silicon Valley. En dicho libro argumentaban que las máquinas carecen de sentido común y de una comprensión del mundo y, por lo tanto, no poseen un genuino modelo cognitivo de la realidad. En este nuevo libro Marcus se enfrenta directamente con algunos de los responsables de los desarrollos recientes de la IA y, sobre todo, de los mensajes que proclaman. En una entrada reciente en Substack explicaba:

El discurso apocalíptico es un caso especial de exageración; Sam [Altman] y Dario [Amodei] usan esa carta una y otra vez, y básicamente equivale a decir “nuestra tecnología es tan aterradora que podría arruinar la sociedad, así que

denos más dinero para que podamos hacer esa tecnología cada vez con más rapidez”. ¿Actuaría así una persona ética y honesta? Si de verdad pensaras que tu tecnología podría destruir la sociedad, ¿te apresurarías a desarrollarla más rápidamente? ¿O te concentrarías en cómo evitar el daño? [...] A menudo parece que el trabajo principal de Amodei es el mismo que el de Altman: recaudar dinero para sistemas costosos de operar y hasta ahora no rentables, en gran medida mediante publicidad exagerada.

De esto va el libro que reseñamos aquí. Su título en inglés (*Taming Silicon Valley*) es quizá menos contundente que el que se le ha dado en español, porque *taming*, más que frenar, es ‘contener’, ‘domesticar’ o ‘controlar’. Pero, en todo caso, su mensaje es muy claro. Si Karp y Zamiska quieren poner a Silicon Valley en la vanguardia de la defensa de los Estados Unidos y su cultura, Marcus cree, por el contrario, que lo que hay que hacer es contener a Silicon Valley, dado el rumbo irresponsable que ha tomado (no a la IA, que puede ser hecha de otra manera y que —eso al menos nos asegura— no va a matarnos a todos).

En las primeras páginas encontramos una buena declaración de intenciones: “He escrito este libro desde la perspectiva de alguien que ama a la IA y que desea desesperadamente que triunfe, pero también como alguien que está desilusionado y profundamente preocupado por el rumbo que están tomando las cosas” (p. 14). Y las cosas que le preocupan son fundamentalmente cuatro: (1) la IA generativa es defectuosa y comete errores o genera alucinaciones, (2) las grandes empresas tecnológicas no están desarrollando una IA responsable y no lo harán si no se las controla, (3) la IA generativa está muy sobrevalorada y los empresarios exageran sus capacidades para conseguir financiación, y (4) se está generando una oligarquía en torno a la IA con demasiado poder. El libro constituye un esfuerzo encomiable por infundir algo de esperanza frente a los discursos catastrofistas o abiertamente apocalípticos que mencionábamos al principio. Y, en mi opinión, lo consigue. Es una llamada convincente a no caer en el derrotismo en este asunto. “Podemos impulsar una IA mejor —escribe—, más confiable y segura y que sea buena para la humanidad, en lugar de la tecnología prematura que tenemos ahora, lanzada a toda prisa, y que está haciendo que algunas personas se vuelvan muy ricas a la vez que amenaza el sustento de muchas otras. [...] Pero eso significa poner controles y contrapesos a la tecnología, no confiar ingenuamente en que todo saldrá bien” (pp. 28-29). En el libro se señalan diversas medidas que podrían ayudar al control de la actual IA y que, aunque difíciles, no son de imposible aplicación. Esperemos que tenga razón. Algunas están ya en la normativa europea, que después de todo no lo hizo tan mal como algunos dicen. ~

ANTONIO DIÉGUEZ es catedrático de filosofía de la ciencia y de lógica en la Universidad de Málaga. En 2024 publicó *Pensar la tecnología* (Shackleton Books).