Frank Pasquale

Las nuevas leyes de la robótica

Defender la experiencia humana en la era de la IA

Galaxia Gutenberg

Serie Interespecies

Cuarto libro de la serie «Interespecies», dirigida por Jorge Carrión, que se propone abordar las claves culturales, sociológicas, tecnológicas y científicas de nuestra época.

Títulos publicados:

Solo quedamos nosotros, Jaime Rodríguez Z.
Atlas del eclipse, Reinaldo Laddaga
La fe en la inteligencia artificial. Los algoritmos predictivos y el futuro de la humanidad, Helga Nowotny

En preparación:

Maneras de ser. Más allá de la inteligencia humana, James Bridle

FRANK PASQUALE

Las nuevas leyes de la robótica

Defender la experiencia humana en la era de la IA

Prólogo de Jorge Carrión

Epílogo actualizado del autor

Traducción de Juan Trejo

Galaxia Gutenberg

Título de la edición original: *New Laws of Robotics: Defending Human Expertise in the Age of AI*
Traducción del inglés: Juan Trejo

Publicado por
Galaxia Gutenberg, S.L.
Av. Diagonal, 361, 2.º 1.ª
08037-Barcelona
info@galaxiagutenberg.com
www.galaxiagutenberg.com

Primera edición: enero de 2024

Preimpresión: Gama, SL
Impresión y encuadernación: Sagrafic
Depósito legal:
ISBN: 978-84-19738-85-1

1

Introducción

La apuesta por el avance tecnológico crece día tras día. Si combinamos las bases de datos de reconocimiento facial con los microdrones, cada día más económicos, tendremos como resultado un poder anónimo y global con una capacidad de ataque de una precisión y eficacia sin precedentes. Pero lo que puede matar también puede curar: los robots podrían expandir enormemente el acceso a la medicina si invertimos más en su investigación y desarrollo. La economía está dando miles de pequeños pasos hacia la automatización de los contratos, del servicio al cliente e incluso de la gestión. Todos esos progresos alteran el equilibrio entre las máquinas y los seres humanos en la disposición de nuestra vida cotidiana.

Evitar las peores consecuencias de la revolución de la inteligencia artificial (IA) y capitalizar todo su potencial depende de nuestra capacidad para cultivar la sabiduría en relación a ese equilibrio. Con ese objetivo, este libro propone tres argumentos pensados para mejorar nuestras vidas. El primero es empírico: hoy por hoy, la IA y la robótica básicamente complementan, más que reemplazan, el trabajo humano. El segundo propone un valor: en muchos campos, tenemos que mantener el *statu quo*. Y el último es un juicio político: nuestras instituciones de gobierno están capacitadas para alcanzar esos objetivos. La principal premisa de este libro es la siguiente: hoy en día disponemos de los medios para canalizar las tecnologías de automatización en lugar de vernos atrapados o transformados por ellas.

Para muchos estas ideas responden al sentido común. ¿Por qué escribir entonces todo un libro para defenderlas? Porque entrañan ciertas implicaciones sorprendentes que cambiarán nuestra

manera de organizar la cooperación social y gestionar los conflictos. Por ejemplo, en el presente, son muchas las economías que favorecen el capital por encima del trabajo y a los consumidores por encima de los productores. Si lo que deseamos es una sociedad justa y sostenible tenemos que corregir esas tendencias.

Dichas correcciones no serán fáciles. Los omnipresentes consultores financieros cuentan una historia sencilla sobre el futuro del trabajo: si una máquina puede grabar e imitar lo que tú haces, serás reemplazado por ella.[1] El discurso sobre el desempleo generalizado tiene a los legisladores atados de manos. Imaginan a los trabajadores humanos como un recurso prescindible en comparación con un *software* más poderoso, con los robots y con el análisis predictivo. Con suficientes cámaras y sensores, prosigue ese argumento, los gestores pueden simular tu «doble mediante bases de datos»: un holograma o robot que realizaría tu trabajo igual de bien, a cambio de una pequeña parte de tu sueldo. Esa visión genera una cruel alternativa: crea robots o sé reemplazado por ellos.[2]

Otra historia es posible y, de hecho, parece más plausible. Prácticamente en todos los ámbitos de la vida los sistemas robóticos pueden hacer que el trabajo resulte más valorado, no menos. Este libro cuenta la historia de médicos, enfermeras, profesores, asistentes de salud en el hogar, periodistas y otras personas que trabajan con especialistas en robótica y científicos informáticos, en lugar de servir dócilmente como fuentes de datos para su futura sustitución. Sus relaciones cooperativas prefiguran el tipo de avance tecnológico que podría propiciar una mejor asistencia sanitaria y educación, mejoras en todos los sentidos para nosotros, sin dejar por ello de llevar a cabo un trabajo significativo. También demuestran que la ley y las políticas públicas pueden ayudarnos a alcanzar una paz y una prosperidad inclusiva en lugar de iniciar una «carrera contra las máquinas».[3] Pero sólo podremos lograrlo si actualizamos las leyes de la robótica que condicionan nuestra visión del progreso tecnológico.

Las leyes de la robótica de Asimov

En 1942, en el relato «Círculo vicioso», el escritor de ciencia ficción Isaac Asimov creó tres leyes para aquellas máquinas que podían percibir su entorno, procesar información y después actuar.[4] El relato habla del «Manual de robótica, 56.ª edición», que indica:

1. Un robot no hará daño a un ser humano ni, por inacción, permitirá que un ser humano sufra daño.
2. Un robot debe cumplir las órdenes dadas por los seres humanos, a excepción de aquellas que entren en conflicto con la primera ley.
3. Un robot debe proteger su propia existencia en la medida en que esta protección no entre en conflicto con la primera o con la segunda ley.

Las leyes de la robótica de Asimov han tenido una enorme influencia. Parecen claras como el agua, pero no son fáciles de aplicar. ¿Puede un dron autónomo atacar a una célula terrorista? La primera mitad de la primera ley («Un robot no hará daño a un ser humano») parece prohibir tal acción. Pero un soldado podría invocar al instante la segunda mitad de la primera ley (está prohibida la «inacción» que podría permitir que «un ser humano sufra daño»). Para decidir qué mitad de la ley debería aplicarse tenemos que tener en cuenta otros valores.[5]

Las ambigüedades no acaban en el campo de batalla. Consideremos, por ejemplo, si las leyes de Asimov pueden aplicarse a los coches automatizados. Los vehículos sin conductor prometen eliminar varios miles de accidentes de tráfico todos los años. Así que el problema puede parecer sencillo a primera vista. Por otra parte, provocaría que cientos de miles de conductores profesionales perdiesen su trabajo. ¿Ese detalle hará que los gobiernos prohíban o ralenticen la adopción de los coches sin conductor? Las tres leyes de Asimov no aclaran nada en ese sentido. Tampoco tienen gran cosa que decir sobre una reciente demanda de los evangelistas de los coches sin conductor: que los peatones aprendan a actuar de un modo que facilite el funciona-

miento de los coches automatizados, e incluso penalizarlos si no lo hacen.

Esa clase de ambigüedades, y otras muchas más, constituyen la razón de por qué los estatutos, las regulaciones y los casos judiciales relacionados con la robótica y la IA en nuestro mundo son mucho más detallados que las leyes de Asimov. A lo largo de este libro analizaremos gran parte de ese nuevo panorama legal. Pero antes de hacerlo quiero introducir cuatro nuevas leyes de la robótica para alentar nuestras futuras investigaciones.[6] Están dirigidas a la gente que construye robots, no a los propios robots.[7] Y aunque son más ambiguas que las de Asimov, constituyen un mejor reflejo de cómo habría que legislar hoy en día. Como los legisladores no pueden anticipar todas las situaciones que las autoridades tienen que afrontar, a menudo facultan a diferentes agencias con estatutos ampliamente redactados. Las nuevas leyes de la robótica deberían suponer algo parecido, articular principios amplios al tiempo que delegan autoridad específica a reguladores con mucha experiencia en ámbitos tecnológicos.[8]

Nuevas leyes de la robótica

A partir de esos objetivos, presentamos las cuatro nuevas leyes de la robótica que serán analizadas más adelante en este libro:

1. Los sistemas robóticos y de IA deberán servir de complemento a los profesionales, no reemplazarlos.[9]

Enfrentarse a la proyección de un posible desempleo a causa de la tecnología genera debates populares sobre el futuro del trabajo. Algunos expertos han vaticinado que prácticamente todos los puestos de trabajo están destinados a desaparecer por los avances tecnológicos. Otros indican posibles obstáculos en el camino a la automatización. La pregunta para los legisladores es: ¿cuáles de esas barreras hacia la robotización tienen sentido y cuáles merecen ser analizadas y eliminadas? Los cortadores de carne robóti-

cos tienen sentido; el cuidado cotidiano robotizado nos da un respiro. ¿Son los reparos, en última instancia, una reacción de tipo ludita o reflejan una sabiduría más profunda sobre la naturaleza humana? Las leyes sobre las licencias impiden, por el momento, que se comercialicen aplicaciones capaces de analizar síntomas de salud como si se tratase de una práctica médica homologada. ¿Es esa una buena política?

Este libro se detiene en ese tipo de ejemplos y reúne argumentos tanto empíricos como normativos para ralentizar o acelerar la adopción de la IA en diferentes campos. Son muchos los factores que importan, relacionados con puestos de trabajo y jurisdicciones. Pero un principio organizador común es la importancia del trabajo significativo para la autoestima de las personas y la gobernanza de las comunidades. Un plan para la automatización priorizaría innovaciones que sirviesen de complemento a aquellos trabajos que tienen, o deberían tener, un carácter vocacional. Máquinas que realicen trabajos peligrosos o degradantes, al tiempo que se asegure de que las personas que actualmente realizan esos trabajos fuesen compensadas por su labor y se les ofrezca la transición para desempeñar otros papeles sociales.

Esa postura equilibrada decepcionará tanto a los tecnófilos como a los tecnófobos. Del mismo modo, el énfasis en el control alejará tanto a aquellos que se oponen a la «interferencia» en los mercados laborales como a aquellos que detestan la «clase formada por gerentes profesionales». En la medida en que las diferentes profesiones equivalen a un sistema económico de castas, que privilegia injustamente a unos trabajadores sobre otros, sus sospechas están justificadas. Sin embargo, es posible suavizar la estratificación si se promueven objetivos superiores para las profesiones.

La clave que se esconde en la esencia de la profesionalización es empoderar a los trabajadores para que tengan voz en cómo se organiza la producción, al tiempo que se les imponen deberes para promover el bien común.[10] Al promover la investigación, tanto en los departamentos universitarios como en las oficinas, los profesionales cultivan la *experiencia distribuida*, aliviando las clásicas tensiones entre la tecnocracia y las normas populares. No deberíamos desmantelar o desactivar profesiones, como aspiran a hacer

demasiados defensores de la innovación disruptiva. Más bien, la automatización humana requerirá el fortalecimiento de comunidades de expertos que ya existen y la creación de otras nuevas.

Una buena definición de profesión tiene que ser amplia y debería incluir a muchos trabajadores sindicados, en especial cuando se trata de personas que utilizan tecnología peligrosa. Por ejemplo, los sindicatos de profesores han protestado por el proceso de «prueba y error» mediante sistemas automatizados y han promovido los intereses de sus estudiantes en otros contextos. Los sindicatos que defienden la profesionalización –facultando a sus miembros para que protejan aquello a lo que se dedican– deberían tener un papel destacado a la hora de conformar la revolución de la IA.

A veces resultará difícil demostrar que un proceso centrado en lo humano es mejor que uno automatizado. Los fríos datos económicos simplifican los complejos procesos críticos. Por ejemplo, los programas de aprendizaje automatizado no tardarán en predecir, basándose en un procesamiento tosco del lenguaje natural, si la propuesta de un libro tiene más probabilidades de llegar a ser un *best seller* o no. Desde una perspectiva puramente económica, dichos programas podrán escoger mejor los libros o los guiones que los editores o los directores. Aun así, los encargados de las industrias creativas tendrán que defender sus conocimientos. Los editores desempeñan un papel importante en la industria de la edición, se valen de su capacidad de juzgar, encontrar y promover trabajos que el público tal vez no sabe (en el presente) que quiere, pero que necesita. Lo mismo puede decirse de los periodistas; a pesar de que la generación automática de textos pueda crear copias que maximicen la publicidad, ese vacuo triunfo nunca podrá reemplazar lo que es capaz de aportar un auténtico y esforzado punto de vista humano. Las escuelas profesionales en las universidades clarifican y reexaminan los estándares de los medios de comunicación, de las leyes, la medicina y otros muchos campos, evitando que caigan en parámetros lo bastante sencillos como para ser automatizados.

Incluso en ámbitos que parecen más susceptibles de encajar en el imperativo de la automatización, en áreas como la logística,

limpieza, agricultura o minería, los trabajadores desempeñarán un papel esencial en una larga transición hacia la IA y la robótica. Reunir o crear los datos necesarios para la IA será una labor muy exigente para algunos. Las regulaciones pueden lograr que sus puestos de trabajo resulten más gratificantes y que sigan bajo su control. Por ejemplo, las leyes de privacidad europeas ayudan a que los conductores no consientan el tipo de vigilancia de 360° o el control que actualmente oprimen a los camioneros en Estados Unidos.[11] Esto no quiere decir que esas actividades, que son peligrosas, no estén monitorizadas. Los sensores pueden detectar problemas en los reflejos de los conductores. Pero existe una enorme diferencia entre sensores pensados específicamente para evitar fallos de seguridad y una grabación constante de vídeo y audio. Conseguir un equilibrio entre una vigilancia inquietante, e incluso degradante, y otra sensata y específica será crucial en una amplia gama de campos.

También podemos diseñar transiciones tecnológicas que incluyan a los seres humanos o, como mínimo, les ofrezcan una oportunidad. Por ejemplo, Toyota ha diseñado coches con una destacada participación de las máquinas, desde el modo conductor (que requiere una monitorización mínima por parte del chófer) al modo guardián (que se centra en el sistema de computación del coche para evitar accidentes, mientras una persona conduce el vehículo).[12] Los aviones disponen de piloto automático desde hace décadas, pero los transportes de mercancías suelen contar como mínimo con dos personas en cabina. Incluso los ocasionales pasajeros de esos vuelos se sienten agradecidos de que los evangelistas de la automatización sustitutiva no tengan prisa por deshacerse de los pilotos.[13]

Cabe señalar, por otra parte, que el transporte de mercancías es uno de los casos más sencillos para la IA. Una vez elegido el destino, no hay discusión posible sobre la dirección del viaje. En otras áreas laborales de servicio sucede justo lo contrario: los clientes o usuarios pueden cambiar de opinión. Los alumnos de una clase pueden estar demasiado alterados en un hermoso día de primavera como para machacar las tablas de multiplicar. Una persona de clase alta puede llamar a su diseñador de interiores

preocupada porque el color escogido para las paredes del salón es demasiado atrevido. Un entrenador personal puede dudar si nota que su clienta está demasiado cansada como para correr un minuto más en la cinta. En todos esos casos, la comunicación es la clave, al igual que en habilidades humanas como la paciencia, la reflexión y el criterio.[14]

Así, si miles de entrenadores personales equipados con Google Glass graban todas sus sesiones, es posible que alguna divina base de datos de gesticulaciones y ojos en blanco, lesiones y éxitos pudiese dictar la respuesta óptima a alguien que lo está pasando mal en un gimnasio. Pero tan sólo empezar a imaginar cómo construir semejante base de datos –que indicase qué es bueno o malo y hasta qué punto– requiere comprender el papel esencial que las personas desempeñarán a la hora de construir y mantener un futuro plausible para la IA y la robótica. La inteligencia artificial seguirá siendo artificial porque siempre será un producto construido a partir de la colaboración humana.[15] Por otra parte, los avances más recientes en IA han sido creados para llevar a cabo tareas específicas en lugar de para ocupar puestos de trabajo o desempeñar papeles sociales.[16]

Son muchos los ejemplos de tecnologías que consiguen que el desempeño del trabajo resulte más productivo, más gratificante, o ambas cosas. Como ha señalado la Agencia para la Italia Digital: «La tecnología no suele reemplazar por completo la figura de un profesional sino tan sólo algunas actividades específicas».[17] Los estudiantes de Derecho de hoy en día apenas pueden creer que los abogados preinternet tuviesen que repasar polvorientos tomos para asegurar la viabilidad de un caso; el *software* de investigación ha facilitado ese proceso y ha ampliado mucho el abanico de fuentes disponibles para los alegatos. Lejos de simplificar las cosas, las ha convertido en algo mucho más complejo.[18] Pasar menos tiempo hojeando libros y más tiempo llevando a cabo el trabajo intelectual de sintetizar casos ha sido una ventaja evidente para los abogados. La automatización puede aportar una eficacia similar a la de miles de otros trabajadores, sin que ello implique un desplazamiento generalizado de la mano de obra. No se trata simplemente de una observación. Es un objetivo propio de la política.[19]

2. Los sistemas robóticos o la IA no tienen que falsificar lo humano.

Desde los tiempos de Asimov al vertiginoso mimetismo de *Westworld*, la posibilidad de crear robots humanoides ha resultado atractiva, aterradora y excitante a partes iguales. Algunos amantes de los robots aspiran a encontrar la mezcla perfecta de huesos de metal y piel sintética que pueda superar el «valle inquietante»: el rechazo que provocan los robots humanoides cuando se parecen demasiado a los humanos, a pesar de no acabar de recrear por completo los rasgos, los gestos y su comportamiento. Los programas de aprendizaje automatizado, que ya han conseguido dominar el arte de crear imágenes de «personas falsas» y voces sintéticas convincentes, seguramente no tardarán en convertirse en algo común.[20] Mientras los ingenieros se esfuerzan para afinar esos algoritmos, surge una pregunta de más amplio calado: ¿queremos vivir en un mundo en el que los seres humanos no saben si están tratando con personas o con máquinas?

Existe una diferencia esencial entre humanizar la tecnología y la falsificación de las características propias de los humanos. Los principales expertos europeos en ética han comentado que «tienen que existir límites (legales) al modo en que la gente puede ser conducida a creer que están tratando con seres humanos cuando, en realidad, están tratando con algoritmos y máquinas inteligentes».[21] Los legisladores ya han aprobado leyes para la «identificación de bots» en contextos propios de internet.

A pesar del creciente contexto ético, hay subcampos de la IA –como los de la computación afectiva, que analiza y simula emociones humanas– dedicados a dificultar cada vez más la distinción entre humanos y máquinas. Esos proyectos de investigación podrían culminar en la creación de androides avanzados como los de la película de Steven Spielberg *I.A.*, indistinguibles de los seres humanos. Los expertos en ética debaten si esos robots humanoides deberían ser siquiera construidos. Pero ¿y si es mejor no construirlos en absoluto?

En hospitales, escuelas, comisarías e incluso fábricas, el beneficio que podría entrañar otorgarle aspecto físico humano al *soft-*

ware sería mínimo y, en cambio, la pérdida podría ser enorme. La carrera para replicar a los humanos podría convertirse, muy fácilmente, en el preludio para sustituirlos. Es posible que algunas personas prefiriesen esa clase de reemplazos en la vida privada y la ley debería respetar dicha autonomía en el ámbito de la intimidad. Pero la idea de una sociedad dedicada a promover el reemplazo en los puestos de trabajo, en la esfera pública y en otros tantos ámbitos, sería una locura. Implicaría confundir el avance de la humanidad con su abolición.

Es posible que esta postura inquiete o confunda a los tecnófilos: rechazar no simplemente la sustancia sino también la premisa, no sólo las leyes de Asimov sino también la amplia literatura sobre el futuro de la tecnología. Espero poder justificar esta visión conservadora reflexionando, capítulo a capítulo, sobre los pasos concretos que tenemos que dar para llegar a un mundo de ciencia ficción en el que los robots sean indistinguibles de los humanos. Esa transición conllevaría una vigilancia generalizada de los seres humanos, crear robots que pudiesen engañar o provocar que los seres humanos tratasen a las máquinas como iguales. No parece una perspectiva deseable.

La voz o el rostro de otro ser humano exige respeto y preocupación; las máquinas no provocan esa clase de reacciones en nuestra conciencia. Cuando los chatbots engañan a los incautos haciéndoles creer que están interactuando con humanos, sus programadores actúan como falsificadores, pues recrean los rasgos de la existencia humana para elevar el estatus de sus máquinas. Cuando la falsificación de dinero alcanza una masa crítica, la divisa real pierde valor. Lo mismo puede decirse de las relaciones humanas en sociedades en las que se permite que las máquinas imiten libremente las emociones, el habla y la apariencia de los seres humanos.

Falsificar lo humano es un peligro muy específico cuando las empresas y los gobiernos desean colocar una cara amistosa al frente de sus servicios. Los Asistentes de Google han cautivado al mundo empresarial con secretarias falsas que conciertan citas, replicando inquietantemente incluso los típicos «hum» y «eh» que se intercalan en cualquier conversación telefónica. Esa clase de

muletillas conversacionales disfrazan el poder de una empresa como Google con la vacilación o la deferencia propias de un discurso humano sin pulir. Encubren una *robocall* haciéndola parecer una consulta humana. Para quienes reciben las llamadas, resulta demasiado fácil imaginar el posible abuso: una avalancha de llamadas por parte de *call centers* robotizados.

Falsificar lo humano no es sólo un simple engaño, es también algo injusto, pues otorga al falsificador el beneficio de aparentar que tiene un interés personal sin tenerlo realmente. Como veremos en un caso tras otro –robots profesores, soldados, servicio de atención al cliente y más–, la insatisfacción y angustia que provocan las imitaciones fallidas de lo humano no son meramente el resultado de una tecnología imperfecta. Son el reflejo de una inteligente precaución respecto hacia dónde apunta la propia tecnología.

3. Los sistemas robóticos y la IA no deben fomentar la carrera armamentística de suma cero.

El debate sobre los «robots asesinos» es un punto central en la cuestión de la ética en relación con las leyes internacionales. Una coalición global de organizaciones civiles está obligando a las naciones a comprometerse a no desarrollar sistemas de armas letales autónomos (SALA). Varios factores obstaculizan ese encomiable propósito a favor de las restricciones tecnológicas. Los líderes militares desconfían de sus homólogos en países rivales. Pueden ocultar descubrimientos en IA militar, adquiriendo un mayor poder a pesar de negar públicamente cualquier avance en ese sentido. Las potencias emergentes pueden imponerse, invirtiendo en industria militar para equipararla con sus nuevos estatus económicos, en tanto que los poderes militares dominantes exigen más recursos para mantener su ventaja relativa. Ese es tan sólo uno de los muchos motivos por los que puede iniciarse una carrera armamentística. Cuando la IA y la robótica entran en juego crece la amenaza de quedar por detrás de los rivales, dado que las tecnologías emergentes prometen ser mucho más precisas, ubicuas y de rápida implementación.

Índice