



INTELIGENCIA ARTIFICIAL, naturalmente

*Un manual de convivencia entre humanos
y máquinas para que la tecnología
nos beneficie a todos*

Nuria Oliver

“Pensamiento para la sociedad digital” Número 1



GOBIERNO
DE ESPAÑA

VICEPRESIDENCIA
TERCERA DEL GOBIERNO
MINISTERIO
DE ASUNTOS ECONÓMICOS
Y TRANSFORMACIÓN DIGITAL

SECRETARÍA DE ESTADO
DE DIGITALIZACIÓN
E INTELIGENCIA ARTIFICIAL

ontsi | observatorio
nacional de las
telecomunicaciones
y de la SI
red.es

Edición de textos

MÓNICA RINA GONZÁLEZ SALOMONE

Edición y coordinación técnica

OBSERVATORIO NACIONAL DE LAS TELECOMUNICACIONES
Y DE LA SOCIEDAD DE LA INFORMACIÓN

Este libro está basado en dos publicaciones por parte de la autora: reproduce secciones de su discurso de toma de posesión en la Real Academia de Ingeniería y el capítulo “Erudit@s digitales” en el libro “Los nativos digitales no existen”.

El contenido de este documento es responsabilidad exclusiva de la autora y las opiniones expresadas en él no representan necesariamente la posición oficial de **red.es**

Reservados todos los derechos. Se permite su copia y reproducción por cualquier medio siempre que se mantenga el reconocimiento de su autor, no se haga uso comercial de las obras y no se realice ninguna modificación de las mismas.



INTELIGENCIA ARTIFICIAL, naturalmente

Un manual de convivencia entre humanos
y máquinas para que la tecnología nos
beneficie a todos

Nuria Oliver

Dedicatoria:
A mi familia, naturalmente

Prólogo

Este libro inaugura la nueva serie de publicaciones del ONTSI titulada “Pensamiento para la sociedad digital”, cuyo propósito es posibilitar que autoras y autores relevantes acerquen a un público amplio información y reflexiones en torno a los principales desafíos -sociales, económicos, culturales y políticos- que representan las tecnologías emergentes en el mundo de hoy.

No es casualidad que el primer número de la serie trate sobre Inteligencia Artificial, al constituir precisamente esta el eje sobre el que gira la nueva revolución socio-tecnológica. Es cierto que hay muchos aspectos de la IA que resultan indescifrables para una mayoría de población no tecnológica; sin embargo, no por ello deja de ser indispensable -más bien al contrario- el intento de democratizar el conocimiento acerca de su génesis e impactos, sus oportunidades y riesgos, su estatus actual y evolución previsible. Y estas son justo las cuestiones que aborda este libro, partiendo de la convicción de que el futuro de la IA dependerá de que exista o no una apropiación activa de sus beneficios y una gobernanza responsable de sus riesgos por parte del conjunto de los agentes institucionales, económicos y ciudadanos. Apropiación y gobernanza que arrancan, siempre, por la disposición de un conocimiento y capacidades que alimenten el criterio para crear, utilizar, proponer y decidir.

Tampoco es casualidad que sea la doctora Nuria Oliver la encargada de conducirnos por los entresijos de este mundo complejo y apasionante. La doctora Oliver es una de las más reconocidas especialistas a nivel internacional sobre la materia, además de una mujer altamente comprometida con el uso de las tecnologías para el bien común, tal y como acredita su dilatada y exitosa trayectoria en este campo.

Sí resulta, en cambio, completamente inesperado que esta obra vea la luz en medio de la terrible pandemia que nos ha traído la COVID-19, llegada -a nuestro país y al conjunto del planeta- en un abrir y cerrar de ojos, pero que va a cambiar nuestras vidas para siempre. Curiosamente, esta realidad no solo no quita vigencia al contenido del libro que aquí se presenta sino que le concede aún mayor actualidad. Porque, ahora más que nunca, el abordaje de los retos sanitarios (información, detección, prevención, tratamiento, fabricación de insumos, búsqueda de una vacuna eficaz, coordinación...), económicos (prestación de servicios, teletrabajo, gestión de ayudas a las empresas...), educativos (educación en línea, coordinación de equipos docentes y familiares), socioculturales (entretenimiento, solidaridad, cocreación, vínculos personales, acompañamiento emocional...), y un largo etcétera, están encontrando en las tecnologías en general, y en la inteligencia artificial en particular, un aliado de primera magnitud.

Pero, de la misma manera, en el contexto actual quedan expuestos crudamente los dilemas y requerimientos éticos en torno a la IA (uso eficaz de los datos sin vulnerar los derechos), el modo en que las brechas digitales son consecuencia -y a la vez causa- de desigualdades sociales

que se expresan hoy con más claridad que en el pasado, aquellos déficits que pudieron haber sido mejor atendidos (monitorización de la salud y bienestar de nuestros mayores y enfermos crónicos, interoperatividad entre los servicios sanitarios y sociales...) y la importancia crucial de invertir en una mayor soberanía tecnológica, sin dejar de reconocer por ello que el desafío que afrontamos es a la par local y global, porque a los virus no los detiene ninguna frontera geográfica.

Con todo, tal vez la mayor enseñanza que nos brinda este libro para los tiempos que vivimos es que urge romper las tradicionales dicotomías entre el desarrollo tecnológico y el desarrollo social. La actual crisis ha abierto un territorio donde se mezclan algoritmos y emociones, la lucha cuerpo a cuerpo con la batalla virtual, las ciencias con las humanidades y las diversas formas de inteligencia entre sí con un único propósito: defender y proteger la vida, poniéndola en el centro de nuestras prioridades. Es este un territorio en el que el valor lo produce la integración, en el que todo es imprescindible y pierden sentido los muros entre tecnología y humanismo para conjugar, así, un nuevo humanismo tecnológico carente de retórica y lleno de verdad. Todos estamos conectados pero, igualmente, todo está -y debe estar- holísticamente conectado. Esta es la principal lección que nos toca aprender en estos momentos difíciles que alumbran por igual sacrificio y esperanza.

David Cierco Jiménez de Parga
Director General de Red.es

Índice

Prólogo	04
Introducción	08
[Capítulo 1] Amor a primera vista. Por qué investigo en tecnología basada en el comportamiento humano	11
[Capítulo 2] El principio. Un paseo personal por la historia de la IA	24
[Capítulo 3] Aquí y ahora. La IA ya transforma nuestras vidas	49
[Capítulo 4] Los retos. Aprovechar la IA exige conocer sus limitaciones	77
[Capítulo 5] Hacia una IA por y para todos. La gobernanza del planeta digital	110
Conclusiones	143
Referencias	145

Introducción

Vivimos en un mundo tecnológico. Nuestro trabajo, entretenimiento, salud, transporte, educación, economía y comunicación dependen y se ven enriquecidos por la tecnología. Sin embargo, somos muy pocos –y aún menos mujeres– quienes creamos tecnología y la utilizamos como herramienta para resolver problemas.

Durante más de 25 años he investigado en cómo modelar el comportamiento humano usando técnicas de Inteligencia Artificial (IA). He trabajado con habitaciones, ropas, coches y móviles inteligentes. He inventado sistemas capaces de reconocer comportamientos o características humanas, como expresiones faciales [1], actividades [2], interacciones [3], maniobras de la conducción [4], la apnea del sueño [5], el riesgo crediticio [6], los puntos calientes de crimen en las ciudades [7] o incluso el aburrimiento [8]. He construido sistemas interactivos e inteligentes en ordenadores y teléfonos móviles [9 y 10]. He sentido en primera persona la felicidad profunda que te invade cuando lo que no era más que una idea –a veces incluso un tanto loca– se convierte en una realidad que puede ayudar a millones de personas.

He sido arte y parte, testigo y partícipe, del progreso tecnológico, de la presencia cada vez más relevante y ubicua

de la tecnología en nuestras vidas, de la dependencia que hemos desarrollado hacia ella.

La inspiración y el motor de mi trabajo han sido, durante toda mi carrera, preguntas con una clara aplicación social. La persona, en sentido individual y colectivo, ha sido y es el elemento central en todos mis proyectos: tecnología dotada de inteligencia por y para la sociedad, tecnología capaz de entendernos como paso previo a ayudarnos. Sin embargo, el impacto que esa misma tecnología está teniendo ahora, ya, en nuestras vidas no siempre es positivo, y por eso siento preocupación. Me pregunto si no nos encontramos ante una crisis social de base tecnológica.

Al mismo tiempo, no tiene sentido aspirar a frenar el desarrollo tecnológico: explorar lo desconocido y empujar el estado del arte forma parte de la esencia del ser humano. Además, necesitamos la tecnología para sobrevivir como especie, superando retos tan inmensos como el cambio climático, la sostenibilidad del planeta, el envejecimiento de la población y la prevalencia cada vez mayor de las enfermedades crónicas.

La Inteligencia Artificial (IA) está abandonando velozmente la esfera de la ciencia ficción. Los sistemas enriquecidos con IA forman hoy parte de nuestra vida cotidiana, y tendrán un papel mucho más relevante en el futuro. El potencial de la IA

para transformar a fondo la sociedad, en prácticamente todos los ámbitos, es inmenso. Pero corremos el riesgo de que un porcentaje muy elevado de la ciudadanía quede al margen de esta transformación. Es más, la metamorfosis no será necesariamente positiva para el conjunto de la sociedad si no trabajamos activamente para que así sea, exigiendo que los avances contribuyan de verdad al progreso, a la igualdad, a la prosperidad... A un mundo mejor para todos, no solo para unos pocos.

Por ello este primer libro de la colección *Pensamiento para la Sociedad Digital* está centrado en la Inteligencia Artificial, haciendo un breve recorrido por su historia, describiendo su impacto actual y planteando los retos que presenta desde diferentes perspectivas. Las últimas páginas esbozan mi visión del futuro. Una visión que no puede ser sino esperanzadora.

7

[CAPÍTULO 1]

Amor a primera vista

Por qué investigo en tecnología basada en el comportamiento humano

Entendí realmente el poder de la Inteligencia Artificial con mi proyecto de fin de carrera. Había escrito un programa para detectar automáticamente coches en vídeos de autopistas, mi primer programa para que un ordenador hiciese algo “inteligente”; de pronto vi, con sorpresa, que conseguía no solo detectar los coches, sino además seguirlos. ¡Qué sensación de empoderamiento! Me di cuenta del valor de la tecnología para ayudarnos a abordar problemas complejos, a hacer tareas de manera más eficiente. Me inspiré para seguir creando tecnología que nos entienda y nos ayude aún más. Por eso he dedicado mi vida profesional a la investigación en Inteligencia Artificial, o IA, y más concretamente al modelado computacional del comportamiento humano.

Soy ingeniera de telecomunicaciones. He aprendido y trabajado en lugares donde se forjaba la era digital que hoy envuelve nuestras vidas, y en los principales focos de creación de las tecnologías

inteligentes. Ahora investigo desde Alicante, donde nací y vivo con mi familia. Trabajo para varias compañías y organizaciones. Soy *Chief Data Scientist* –algo así como investigadora principal de datos– en Data-Pop Alliance, una ONG dedicada al Big Data y la IA para el bien social; y *Chief Scientific Advisor* en el Vodafone Institute, un laboratorio de ideas basado en Berlín. También colaboro con numerosas instituciones, incluyendo ELLIS, que busca fomentar la investigación excelente en Inteligencia Artificial en Europa; y OdiselA, que aspira a contribuir a un uso ético de la Inteligencia Artificial. He sido hasta hace poco la primera directora de Investigación en Ciencias de Datos a nivel mundial de Vodafone.

Mi próximo reto es crear desde cero una unidad ELLIS en Alicante, dedicada a la investigación excelente en Inteligencia Artificial, y conectada con una veintena de otras unidades ELLIS en diferentes países europeos. Una red de excelencia para conseguir atraer, retener e invitar a la próxima generación de talento excelente en investigación en Inteligencia Artificial a quedarse en Europa.

Durante un fascinante período en mi etapa formativa realmente sentí que estábamos inventando el futuro: realidad aumentada, ropa inteligente, coches inteligentes, informática afectiva, tinta electrónica, objetos conectados... Estas aplicaciones hoy convertidas en realidad –o a punto de serlo– fueron algunas de mis áreas de investigación en los años noventa.

Por supuesto, no crecí pensando que me dedicaría a esto. Mi primer contacto con las telecomunicaciones fue durante la Semana Santa de mi último año de bachillerato. Sentía la presión de tener que elegir una carrera, y no sabía cuál. Siempre me ha apasionado la ciencia; Leonardo da Vinci, Marie Curie, Albert Einstein... me atraía enormemente la figura del inventor/investigador. Desgraciadamente no podía preguntar a esos gigantes del conocimiento en qué consistía su trabajo, ni cómo llegaron a él. Pero un amigo de mi hermano, Miguel Vallés, que hacía “Teleco” en Madrid y había venido a Alicante a pasar la Semana Santa, me habló con tanta pasión de las telecomunicaciones, de sus aplicaciones prácticas y de la vida en un colegio mayor, que me contagió. Decidí estudiar Ingeniería de Telecomunicación en la Universidad Politécnica de Madrid (UPM).

Fue el primer paso hacia la realización de muchos de mis sueños. Siempre estaré agradecida a Miguel por contarme su experiencia, por inspirarme, por ayudarme –sobre todo en el aterrizaje en Madrid– y por su amistad generosa e inquebrantable todos estos años.

Entendí pronto, incluso antes de saber a qué área específica me dedicaría, el grandísimo potencial de la tecnología para mejorar la sociedad. Mi pasión era y sigue siendo la investigación, así que mi idea fue hacerme investigadora en algún tema tecnológico. Así fue como, siendo estudiante, descubrí la Inteligencia Artificial. Fue amor a primera vista.

La imaginación es el límite

Mi primera publicación científica la presenté siendo estudiante de tercero o cuarto de carrera –no recuerdo bien–, en un congreso en Roma, tutelada por mi gran profesora, hoy amiga, Carmen Sánchez. Era un trabajo sobre redes neuronales, uno de los primeros modelos de Inteligencia Artificial inspirado en la manera en que aprende el cerebro humano. Poco después hice el proyecto fin de carrera, tan decisivo para mí, dentro del Grupo de Tratamiento de Imágenes de la UPM, liderado por el profesor Narciso García. Ese trabajo para identificar automáticamente coches en vídeos, con técnicas de visión por ordenador, dio fuerza y forma a mi primer gran sueño: estudiar un doctorado en Inteligencia Artificial en Estados Unidos.

Pude hacerlo realidad gracias a una beca de la Fundación Obra Social la Caixa, y a que me aceptaron en siete universidades estadounidenses –todas a las que apliqué– incluyendo Stanford, Caltech, Carnegie Mellon y el Instituto Tecnológico de Massachusetts (MIT). Me sentía muy afortunada. Tras un proceso de decisión obligatoriamente difícil opté por el legendario Media Lab del MIT. Comencé así mi carrera científica bajo la dirección del profesor Sandy Pentland, uno de los investigadores en ciencias de la computación más citados del mundo, de quien sigo aprendiendo y a quien tan agradecida estoy.

He de reconocer que el MIT impone. Yo aterricé con cierto temor a carecer de la formación necesaria para sobrevivir en un entorno académico tan exigente, pero resultó que la educación recibida en la

Escuela de Telecomunicación de la UPM era, en muchas áreas, incluso más completa y profunda –aunque menos experimental– que la de mis compañeros en el MIT. Me sentí como pez en el agua.

Fueron años de trabajo intenso, pero muy divertido y enriquecedor. Prácticamente vivía en el laboratorio, trabajando todos los días –fines de semana incluidos– con gran dedicación y pasión. Compartimos experiencias inolvidables: las demos –el lema del Media Lab es *demo or die*–, los eventos con los patrocinadores, el primer desfile del mundo de ropa inteligente, las carreras de coches de fórmula Dodge...

Fue un periodo de aprendizaje constante, de gran creatividad y estimulación intelectual, y de exposición a una cultura extremadamente positiva que ya forma parte de mi identidad. Una cultura que fomenta el asumir riesgos, que considera las equivocaciones un regalo de la vida para aprender y donde no hay más límite que tu propia imaginación. Conservo de entonces amistades intensas con personas extraordinarias, autoras de profundas contribuciones a la ciencia y a la sociedad. También fueron años duros de trabajo sin descanso, a miles de kilómetros de mi familia y con un clima inhóspito en invierno y en verano.

Estuve en el Media Lab desde finales de agosto de 1995 hasta junio del año 2000, una etapa dorada para este laboratorio justo antes de la crisis de las *puntocom*. Hice el doctorado en poco tiempo, teniendo en cuenta la duración media entonces de los doctorados en el Media Lab. Fue una época de gran optimismo respecto a la tecnología,

un periodo de creatividad y de definición de algunas de las áreas tecnológicas que son hoy día una realidad.

Por ejemplo, mi primer proyecto en el MIT fue LAFTER [1], uno de los primeros sistemas en el mundo de reconocimiento de expresiones faciales en tiempo real. Ahora lo hacemos con un móvil, pero en 1995 reconocer expresiones era casi ciencia ficción. También, en 1997 organizamos el primer desfile de ropa inteligente del mundo, y trabajé en un coche capaz de detectar y además predecir las maniobras de los conductores. Otro de mis proyectos, para museos, reconocía en tiempo real los cuadros y mostraba en unas gafas de realidad aumentada –término que apenas había salido de los laboratorios– un vídeo explicativo sobre ellos, de nuevo una aplicación que ahora no sorprende pero que hace dos décadas era experimental.

Dejé el MIT para iniciar mi carrera profesional en los laboratorios de investigación de Microsoft, investigando con Eric Horvitz y Mary Czerwinski. Durante este periodo me di cuenta de que si mi sueño –¡otro!– era conseguir que los ordenadores nos entendieran –como paso necesario previo a ayudarnos–, debía enfocar mi trabajo al que es el ordenador más personal: el móvil. Por ello desde 2005 he investigado casi exclusivamente en proyectos relacionados con móviles, acercándome de nuevo a las telecomunicaciones.

Un perfil poco común

Nunca pensé que podría regresar a España, aunque siempre lo deseé.

Durante mucho tiempo la vuelta fue solo un sueño más. Pero hace trece años Telefónica me ofreció la oportunidad de regresar como Directora Científica –la primera mujer– en Telefónica I+D, en Barcelona, y fue como si la diosa Fortuna de nuevo llamara a mi puerta.

El reto profesional era grande: definir la visión de un área de investigación nueva para la compañía; identificar, atraer y liderar talento; publicar y patentar. También, aprender a ser directora y a desarrollar mi capacidad de liderazgo, en femenino. En el plano personal debíamos comenzar de cero en una ciudad donde no conocíamos a nadie. Un nuevo capítulo en nuestras vidas. Un gran desafío y, a la vez, una oportunidad maravillosa de contribuir al progreso científico en España.

En efecto la vuelta fue intensa desde todos los puntos de vista. Gracias a la red de becarios de La Caixa, y también a través del colegio de mis hijos, conocimos a grandes personas que son hoy amigos y amigas. Pronto nos sentimos integrados. En cuanto al plano profesional, mi papel fue ser agente de cambio.

Hoy en día los datos –lo que llamamos *Big Data*– y la Inteligencia Artificial son elementos estratégicos para la mayoría de las empresas de telecomunicaciones, pero no era así hace más de una década. Mi objetivo era crear y liderar en Telefónica I+D las áreas en que se centra mi investigación: la Inteligencia Artificial, el análisis de Big Data, la personalización, la interacción persona-móvil. Eran áreas emergentes para una compañía tradicionalmente enfocada, como

todas las compañías de telecomunicaciones, en redes y sistemas de comunicación.

Me siento muy orgullosa del trabajo realizado durante casi nueve años en Telefónica I+D. Obtuvimos numerosos premios y nominaciones a mejor artículo científico; decenas de patentes; y proyectos de investigación que dieron lugar a nuevos productos e incluso a compañías. También contribuimos a la creación del área de Big Data e Inteligencia Artificial a escala global y creamos el área de Big Data para el Bien Social, entre otros logros.

Sé que mi perfil como investigadora y directora de investigación, con impacto a nivel científico y en la sociedad, es poco común en España. Aquí puede costar entender que tenga una carrera científica de primer nivel sin estar en una universidad, o en un organismo público de investigación. Eso ocurre porque la investigación industrial en inteligencia artificial en EE. UU. y España es diferente. EE.UU. cuenta con potentísimos laboratorios de investigación en Inteligencia Artificial en las empresas tecnológicas –Microsoft, Facebook, Google, Apple, Amazon, IBM–, mientras que en España no hay tradición de que se investigue en este campo, fundamentalmente porque no hay grandes tecnológicas españolas. Defiendo que es importante que haya investigación tanto en el sector público como en el privado, y en colaboración. Muchos de mis proyectos de investigación han sido en colaboración con universidades.

Trabajo global, vida local

Dicen que la vida es cíclica, y sin duda en mi caso así ha sido. Hace cuatro años decidimos mudarnos a Alicante para estar junto a mi familia. Gracias a las telecomunicaciones hemos podido convertir en realidad lo que parecía imposible. Tanto mi marido –arquitecto de software de Microsoft– como yo trabajamos desde casa: la tecnología nos ha permitido encontrar un equilibrio entre una vida profesional intensa y global, y una vida personal no menos intensa, pero más local. Hay intangibles sumamente valiosos en la vida, como crecer cerca de tus abuelos, tíos, primos... Mi marido, que es de origen alemán, y yo nos sentimos afortunados de poder ofrecer esa experiencia vital a nuestros hijos.

Estando en Alicante decidí dejar Telefónica para embarcarme en tres nuevas aventuras: Data-Pop Alliance, Vodafone y el Vodafone Institute. Aunque recientemente he dejado Vodafone, sigo investigando a través de Vodafone Institute y de Data-Pop Alliance, donde abordamos cómo el Big Data puede ayudarnos en áreas relativas a los Objetivos de Desarrollo Sostenible, como la salud pública, la seguridad física, la igualdad de género, la inclusión financiera o la educación.

También intento tener impacto en mi región, dando charlas en las universidades, al público en general y, en especial a adolescentes, para inspirarles y acercarles –sobre todo a las chicas– hacia las carreras tecnológicas. Además, a través de la nueva unidad ELLIS que espero poder crear en breve, tengo la oportunidad de atraer talento investigador excelente a la Comunidad Valenciana, conectado con una comunidad de investigadores en otros países de Europa.

La divulgación científico-tecnológica me parece esencial, así que dedico parte de mi tiempo a colaborar con medios de comunicación y a apoyar eventos científico-tecnológicos para todos los públicos. Asesoro a varias universidades, a la compañía Mahindra Comviva, a la Fundación Gadea Ciencia, a la Comisión Europea, al Foro Económico Mundial y a los Gobiernos valenciano y español, sobre tecnología y especialmente IA y Big Data.

Con la perspectiva de más de dos décadas desde que descubrí la Inteligencia Artificial en la UPM, siento hoy un gran entusiasmo por ver convertidas en realidad muchas de las ideas a las que contribuí hace un cuarto de siglo. Me apasiona poder ayudar con mi trabajo a mejorar la calidad de vida de las personas y de nuestro planeta gracias a la IA.

No obstante, a pesar del impacto positivo de la IA en nuestra sociedad, también siento preocupación. El desarrollo de la Inteligencia Artificial por y para la sociedad requiere no solo una inversión ambiciosa en la investigación, innovación y adopción de la IA en el tejido empresarial, sino también la definición y cumplimiento de marcos éticos y regulatorios apropiados. Es además imprescindible una profunda reforma educativa a todos los niveles, combatiendo activamente la falta de equilibrio, de diversidad geográfica, demográfica e institucional en la expansión de una tecnología capaz de transformar profundamente nuestra forma de vida. De mi pasión por la IA, y de mi sueño de usarla para el beneficio de todos, nace este libro.



TECNOLOGÍA PARA LAS PERSONAS (Y NO AL REVÉS)

Mi investigación ha contribuido a aumentar la presencia de la IA en nuestras vidas. He impulsado el reconocimiento automático en imágenes de vídeo de expresiones faciales [1] y de interacciones humanas [3], dos habilidades necesarias para que los ordenadores entiendan nuestro estado emocional, y para automatizar tareas de vigilancia, entre otras aplicaciones. También he trabajado en la identificación automática de las actividades que se llevan a cabo en una oficina [2], para mejorar el diseño de espacios de trabajo inteligentes; en la predicción de las maniobras de la conducción [4], -para aumentar nuestra seguridad al volante-, y en el modelado automático de características humanas como la personalidad, para desarrollar tecnología que nos comprenda y pueda, en consecuencia, ayudarnos.

Uno de mis objetivos ha sido diseñar tecnología que se adapta a nosotros, y no al revés. He propuesto modelos de recomendación -de compras, ocio, amigos- que incorporan información de contexto, como el día de la semana o la localización, porque nuestras necesidades varían en función del momento y el lugar [15]. He diseñado algoritmos para optimizar la selección de resultados [16], y para que las recomendaciones incluyan opiniones de expertos.

Así ha surgido el concepto de la sabiduría de los expertos, que complementa a la sabiduría de las masas [17]—basada en acciones y opiniones de personas similares a ti—, y uno de los fundamentos de los algoritmos de recomendación.

En varios proyectos he buscado convertir el móvil en un asistente para la salud y el bienestar. *HealthGear* [5] detecta automáticamente la apnea del sueño; *MoviPill* [9] ha mejorado la adherencia al tratamiento de los enfermos crónicos en un 60%, al convertir en un juego social la toma correcta de la medicación.

En el ámbito financiero la unión del móvil y la IA puede igualmente ser fuente de mejoras. Unos 1.700 millones de personas en el mundo no tienen cuenta bancaria pero sí acceso a un móvil, lo que convierte a esta herramienta en una de las llaves para lograr la inclusión financiera, es decir, el acceso de todos a productos y servicios financieros de calidad. En *MobiScore* [6] diseñamos un sistema para la inferencia de riesgo crediticio en economías en desarrollo a partir de patrones de uso del móvil, y permitir así el acceso a crédito a al menos una parte de quienes carecen de cuenta corriente. Al mismo tiempo hemos investigado los factores que llevan a la adopción del dinero móvil en África, un fenómeno que fomenta la inclusión financiera, ya que permite realizar transacciones a personas sin

cuenta bancaria, pero con móvil (no necesariamente un smartphone) [19].

En Telefónica y en Vodafone creé un área de investigación para analizar grandes cantidades de datos agregados y (pseudo)anonimizados con fines sociales positivos. A esta tarea se dedica Data-Pop Alliance. Aplicar técnicas de IA al estudio de este tipo de datos procedentes de la red de telefonía móvil nos abre muchas puertas. Nos permite desde entender mejor el comportamiento en las ciudades, por ejemplo, para detectar automáticamente puntos calientes de crimen [7], hasta estimar el impacto de desastres naturales, como terremotos e inundaciones [45], y modelar la propagación de enfermedades infecciosas como la malaria [21].



[CAPÍTULO 2]

El principio

Un paseo personal por la historia de la IA

Lo hemos asumido: las máquinas nos superan a nosotros, sus creadores, incluso en tareas asociadas a la capacidad de estrategia y a la intuición. ¿Cómo lo han conseguido? ¿Significa que pueden *pensar* como nosotros? A todo esto, ¿cómo pensamos nosotros? El esfuerzo por crear cerebros no biológicos parte de la reflexión sobre el pensamiento y el aprendizaje humanos, e incluye ambiciosas visiones futuristas; pruebas para detectar inteligencia; y procesadores y algoritmos capaces de analizar, dar sentido y aprender a partir de cantidades ingentes de datos, entre otros elementos. El éxito, en última instancia, curiosamente depende de la capacidad de otorgar a los errores el valor (matemático) que se merecen.

Aunque la Inteligencia Artificial (IA) parezca algo novedoso, el desarrollo de máquinas capaces de pensar o dotadas de algunas capacidades humanas ha cautivado nuestro interés desde la antigüedad. Los primeros autómatas –robots antropomorfos– que imitaban movimientos humanos fueron contruidos hace milenios.

Según la Ilíada, Hefesto –el dios griego del fuego y la forja– creó dos mujeres artificiales de oro con “sentido en sus entrañas, fuerza y voz” que lo liberaban de parte de su trabajo, es decir, creó robots para que lo ayudaran, lo cual lo convierte en todo un adelantado a su tiempo. Heron de Alejandría en el siglo I escribió *Automata*, donde describe máquinas capaces de realizar tareas automáticamente, como estatuas que sirven vino o puertas que se abren solas. Otros ejemplos incluyen autómatas con fines religiosos, como las figuras mecánicas de los dioses en el Antiguo Egipto, operadas por sacerdotes para sorprender a la multitud; y también lúdicos: las famosas cabezas parlantes y autómatas de la Edad Media, el Renacimiento y el siglo XVIII.

Más allá de la automatización, el ser humano siempre ha sentido curiosidad por explicar y entender la mente humana para, entre otros motivos, construir una mente artificial. Hace más de 700 años, Ramon Llull –beato y filósofo mallorquín patrón de los informáticos– describió en su *Ars Magna* (1315) la creación del *Ars Generalis Ultima*, un artefacto mecánico capaz de analizar y validar o invalidar teorías utilizando la lógica: un sistema de Inteligencia Artificial.

La creadora, en el siglo XIX, del primer algoritmo destinado a ser procesado por una máquina –en otras palabras, el *primer programa informático*– fue la matemática Ada Byron (Lovelace). Su visión de que las máquinas podrían servir para algo más que para hacer cálculos matemáticos convirtió a Byron en la primera

persona en proponer el uso de la máquina analítica de Babbage para resolver problemas complejos. La máquina de Babbage, no obstante, con sus previstos treinta metros de largo por diez de ancho, nunca llegó a construirse. Se la considera el primer diseño de un computador de propósito general, pero los obstáculos técnicos y la falta de respaldo político –en parte por miedo a un posible uso bélico– no permitieron convertirla en realidad.

Una conversación inteligente

El mito y la ficción literaria respecto a la Inteligencia Artificial empezaron a materializarse a partir de los años cuarenta del siglo XX, con los primeros ordenadores.

El genial matemático e informático inglés Alan Turing es considerado el padre de la Inteligencia Artificial. Habló de ella en el legendario artículo *Computing Machinery and Intelligence*, publicado en 1950. Es en este trabajo donde propone la famosa prueba de Turing, ilustrada en la Figura 1, para determinar si un sistema artificial es inteligente.

El ejercicio consiste en que un humano (C en la figura), conocido como el interrogador, interacciona vía texto con un sistema al que puede hacer preguntas. Si el humano no logra discernir cuándo su interlocutor es una máquina (A en la figura), y cuándo otra persona (B en la figura), entonces el sistema supera la prueba de Turing: es inteligente.

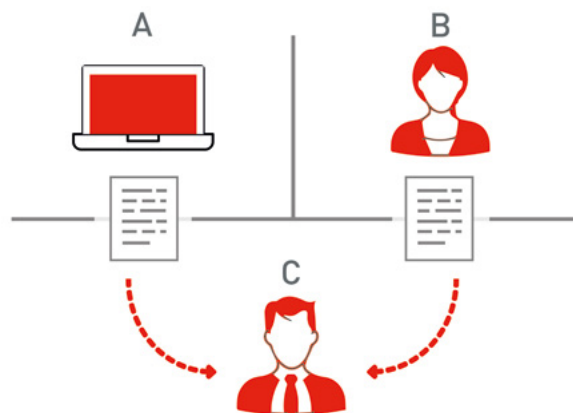


Fig.1 La prueba de Turing (fuente: Wikipedia).

La década fundacional para la Inteligencia Artificial fue la de los cincuenta. En 1951 el profesor Marvin Minsky –a quien tuve el honor de conocer en el MIT– construyó la primera red neuronal computacional como parte de su doctorado en la Universidad de Princeton. Se trataba de una máquina con válvulas, tubos y motores que emulaba el funcionamiento de neuronas interconectadas, y lograba simular el comportamiento de ratas que aprenden a orientarse en un laberinto. La máquina, con sus 40 neuronas, fue uno de los primeros dispositivos electrónicos construidos con capacidad de aprender.

Apenas cinco años más tarde, en 1956, tuvo lugar la mítica convención de Dartmouth (New Hampshire, EE. UU.), en la que participaron figuras legendarias de la informática como John McCarthy, Marvin Minsky, Claude Shannon, Herbert Simon y Allen Newell, todos ellos ganadores del premio Turing, el más prestigioso en computación, equivalente al Nobel –galardón que, además, ganó Simon–.

Dartmouth marca un hito porque es en este encuentro donde se define la Inteligencia Artificial y se establecen las bases para su desarrollo, identificando preguntas clave que incluso hoy día nos sirven de mapa conceptual a los investigadores en esta área (Ver recuadro *Las siete cuestiones fundacionales de la IA*).

Es en Dartmouth donde McCarthy acuña el término de Inteligencia Artificial, para referirse a “la disciplina dentro de la Informática o la Ingeniería que se ocupa del diseño de sistemas inteligentes”, esto es, sistemas con la capacidad de realizar funciones asociadas a la inteligencia humana como percibir, aprender, entender, adaptarse, razonar e interactuar imitando un comportamiento humano inteligente.

¿Algún gato en la imagen?

McCarthy quiso diferenciar la Inteligencia Artificial del concepto de cibernética, impulsado por Norbert Wiener –también profesor del MIT–, y en el que los sistemas inteligentes se basan en el reconocimiento de patrones, la estadística, y las teorías de control y de la información. McCarthy, en cambio, quería enfatizar la conexión de la Inteligencia Artificial con la lógica. Esta diferencia dio lugar a dos escuelas distintas dentro del desarrollo de la IA, como explico más adelante.

Para conseguir que un ordenador aprenda, por ejemplo, a identificar gatos en imágenes, podemos usar distintas

estrategias. La aproximación basada en la estadística y en el reconocimiento de patrones requiere mostrar al ordenador miles de fotos con gatos y sin gatos –llamamos a estos ejemplos *datos de entrenamiento anotados*–; así, proporcionamos a los algoritmos de reconocimiento de patrones la información que necesitan para *aprender* a identificar automáticamente los patrones recurrentes en las fotos con gatos, *versus* en las fotos sin gatos. Una vez estos algoritmos han sido entrenados con suficientes ejemplos, serán capaces de determinar si hay o no un gato en las fotos nuevas que les sean presentadas.

La aproximación basada en la lógica conllevaría la definición de una taxonomía de los animales; dentro de estos, de los mamíferos; dentro de estos, de los felinos; y dentro de estos, de los gatos, describiendo sus características. Este conocimiento, programado en el ordenador, tendría que ser lo suficientemente rico, preciso y flexible como para permitir a la máquina localizar en las imágenes las características que definen a un gato, y detectar su presencia.

En un guiño del destino, la propuesta intelectual de Wiener –basada en datos y estadística– se ha convertido en la dominante en la Inteligencia Artificial, pero utilizando la terminología de McCarthy. Sin embargo, no adelantemos acontecimientos.



Fig. 2 Participantes de la convención de Dartmouth, incluyendo a Marvin Minsky (en el centro del grupo, con gafas), Claude Shannon (primero por la dcha.) y Ray Solomonoff (tercero por la izda.) (Cortesía de la familia Minsky).

Aprendiendo a distinguir entre izquierda y derecha

Entre las más controvertidas y citadas declaraciones realizadas en los albores de la Inteligencia Artificial se cuentan las del psicólogo Frank Rosenblatt, creador del Perceptrón en el Laboratorio Aeronáutico de Cornell. El Perceptrón se presentó a la prensa en 1958, como un programa instalado en un ordenador de IBM, el 704, que por cierto ocupaba una estancia entera.

Según la crónica publicada el 8 de julio de 1958 en el *New York Times*, el Perceptrón habría de convertirse en “el primer ordenador

capaz de pensar como el cerebro humano”, equivocándose al principio, pero “volviéndose más sabio con la experiencia”. Rosenblatt –señala el *Times*– lo describió como el “embrión” de un ordenador en el futuro capaz de “caminar, hablar, ver, escribir, reproducirse y ser consciente de su existencia”.

Fue un apreciable ejercicio de extrapolación, teniendo en cuenta que la única habilidad que el Perceptrón mostró a los medios fue aprender a distinguir entre izquierda y derecha. Al 704 se le introducían dos tarjetas, una con marcas en el lado izquierdo y la otra en el derecho; la computadora empezaba a distinguir una tarjeta de otra al cabo de 50 intentos. Rosenblatt explicó que el avance se debía a un cambio en el programa autoinducido por el propio programa, lo que implica aprendizaje.

Más tarde el Perceptrón se implementó en un dispositivo propio, el Perceptrón Mark 1 (ilustrado en la Figura 4), que se aplicaba al análisis de imágenes. Rosenblatt estaba convencido que la máquina reproducía de manera simplificada el funcionamiento de neuronas que trabajan estableciendo conexiones, llamadas sinapsis, con otras neuronas.

Como puede observarse en la Figura 3, el Perceptrón recibe un conjunto de valores de entrada que multiplica cada valor de entrada por un coeficiente determinado al que llamamos peso, o $W1j$, $W2j$, etc... de weight en inglés en la Figura, que representa la fuerza de la sinapsis con cada neurona adyacente; y produce una salida: 1 si la suma de las entradas moduladas por sus pesos

es superior a un cierto valor, y 0 si es inferior. Las salidas 1 y 0 representan la activación o no de la neurona.

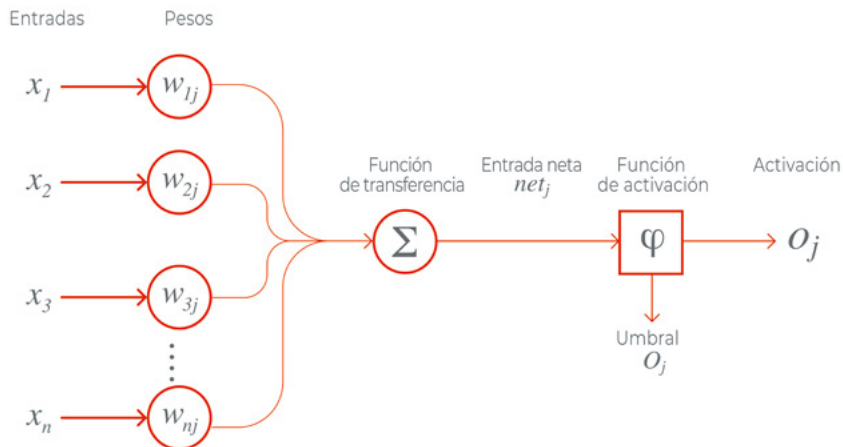


Fig. 3 Perceptrón de Rosenblatt (fuente: Wikipedia).

El modelo está basado en un trabajo anterior de Warren McCulloch y Walter Pitts, que demostraron que un modelo de neurona como el descrito puede representar funciones de OR/AND/NOT. Este resultado era importante porque, como hemos explicado, en los albores de la IA se pensaba que cuando los ordenadores pudieran llevar a cabo operaciones de razonamiento lógico formal, se conseguiría la Inteligencia Artificial.

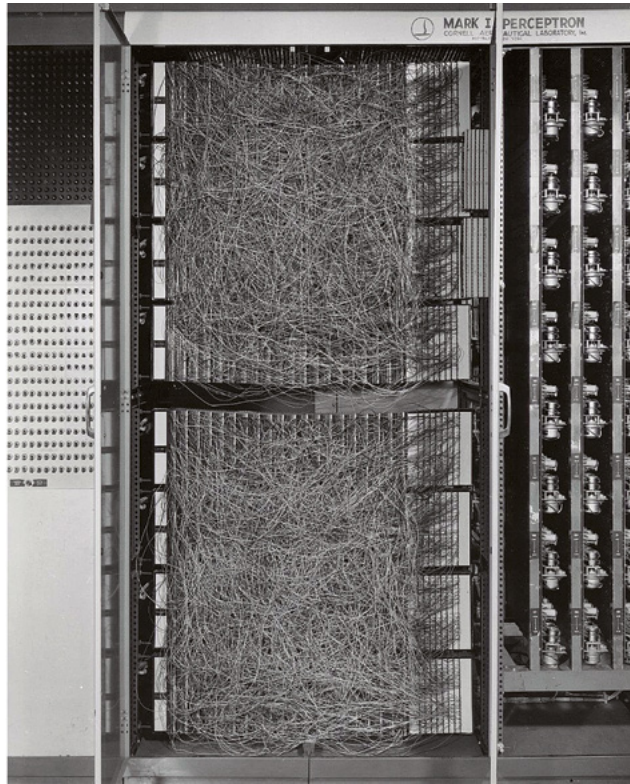


Fig. 4 *Perceptrón Mark I del Laboratorio de Aeronáutica de Cornell*
(fuente: Wikipedia).

Aprender a partir de reglas, o de la experiencia

Desde las dos aproximaciones distintas a la Inteligencia Artificial por parte de Wiener (basada en datos) y McCarthy (basada en la lógica), ha existido cierto enfrentamiento entre dos escuelas de pensamiento en la Inteligencia Artificial: el enfoque simbólico-lógico o *top-down* -originalmente llamado *neat*-; y el enfoque basado en datos, conexionista o *bottom-up* -originalmente conocido como *scruffy*-.

Son abordajes muy distintos conceptualmente. El simbólico, *top-down* –en inglés, de arriba a abajo–, postulaba que las máquinas, para razonar, debían seguir un conjunto de reglas predefinidas y unos principios de la lógica. La idea es programar en la máquina el conocimiento que poseemos los humanos, de forma que después, aplicando las reglas que también han sido enseñadas previamente, el ordenador pueda derivar conocimiento nuevo.

El ejemplo canónico son los llamados sistemas expertos, el primer ejemplo comercial de la Inteligencia Artificial. Hablaremos de ellos más adelante.

Por su parte la escuela *bottom-up* –de abajo a arriba– proponía que la Inteligencia Artificial debía inspirarse en la biología, aprendiendo a partir de la observación y de la interacción con el mundo físico, esto es, de la experiencia. Según este enfoque, si aspiramos a crear IA debemos proporcionar a los ordenadores observaciones de las que aprender. Esto conlleva entrenar algoritmos a partir de miles de ejemplos de lo que queremos que aprendan.

En los sistemas de Inteligencia Artificial suele tomarse como referencia la inteligencia humana. Del mismo modo que la inteligencia humana es diversa y múltiple, la Inteligencia Artificial es una disciplina con numerosas ramas de conocimiento, que se nutren de las dos grandes escuelas de pensamiento *top-down* y *bottom-up*.

La escuela simbólico-lógica incluye, entre otras, áreas como la teoría de juegos; la lógica; la optimización; el razonamiento y la

representación del conocimiento; la planificación automática; y la teoría del aprendizaje.

En la escuela *bottom-up* destacaría la percepción computacional – una de mis áreas de especialidad, que abarca el procesamiento de imágenes, vídeos, texto, audio y datos de otro tipo de sensores–; el aprendizaje automático estadístico –*machine learning*, otra de mis áreas–; el aprendizaje con refuerzo; los métodos de búsqueda –de texto, imágenes, vídeos–; los sistemas de agentes; la robótica; el razonamiento con incertidumbre; la colaboración humano-IA; los sistemas de recomendación y personalización; y las inteligencias social y emocional computacionales.

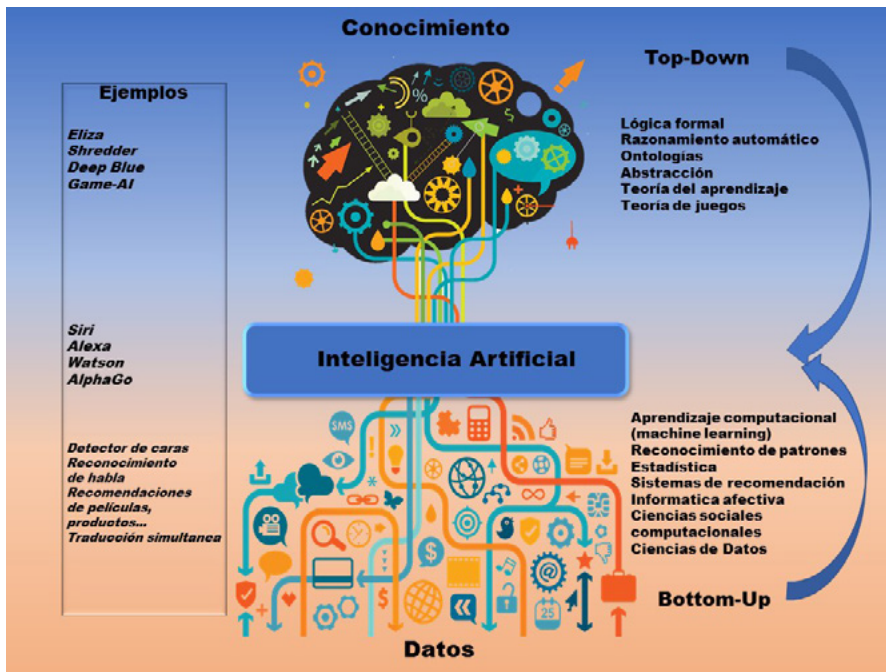


Fig. 5 Ejemplos de sistemas top-down y sistemas bottom-up.

Primera era dorada: los sistemas expertos

Mis trabajos de investigación dentro de la IA se enmarcan en el enfoque *bottom-up*. Inicialmente esta escuela no tuvo mucho éxito práctico, ya que no había disponibles grandes cantidades de datos, ni la capacidad de computación necesaria para entrenar modelos suficientemente complejos como para resultar útiles. Por ello la primera aplicación práctica de la Inteligencia Artificial fue en los años 60, con los sistemas expertos, que pertenecen al enfoque simbólico-lógico.

En 1956, después de la convención de Dartmouth, Herbert Simon predijo que “en veinte años, las máquinas serán capaces de hacer el trabajo de una persona”. Marvin Minsky, por su parte, declaró en 1970 a la revista *Life* que “dentro de tres a ocho años tendremos una máquina con la inteligencia general de un ser humano”. Hasta mediados de los años setenta predominó el optimismo en todo lo relativo a la Inteligencia Artificial y su impacto.

De hecho, el periodo entre 1956 y 1974 suele conocerse como la primera etapa dorada de la Inteligencia Artificial. Fueron los años en que Edward Feigenbaum –uno de los fundadores del departamento de informática de la Universidad de Stanford– lideró el equipo que construyó el primer sistema experto, implementado en LISP, el programa de ordenador desarrollado por McCarthy.

El nombre de este sistema experto era DENDRAL, y fue fruto

del deseo del biólogo molecular Joshua Lederberg, también de Stanford, de disponer de un sistema que facilitara su investigación sobre compuestos químicos en el espacio. DENDRAL ayudaba a los químicos orgánicos a identificar moléculas desconocidas a partir de su espectro de masas, gracias a que le había sido transferido el conocimiento de un prestigioso químico –en concreto Carl Djerassi, creador de la píldora anticonceptiva–. DENDRAL era experto en química porque atesoraba el conocimiento químico y la experiencia de un humano experto en este campo.

Arrecia el ‘primer invierno’

Pero a principios de los 70 llegó el invierno, el primer invierno de la IA. Las ambiciosas expectativas creadas durante las dos décadas anteriores no se cumplieron. Además, en 1969 fue publicado el libro *Perceptrons*, de Minsky y Seymour Papert, a quienes tuve el honor de conocer durante mis años en el MIT. Esta obra contribuyó a desinflar aún más el interés por los modelos *bottom-up* y, en particular, por las redes neuronales.

En *Perceptrons* se demostraba que los perceptrones eran muy limitados porque solo pueden aprender funciones extremadamente simples o, expresado en términos más matemáticos, funciones linealmente separables.

Desgraciadamente, la gran mayoría de los problemas del mundo real son complejos, y no cumplen la condición de ser linealmente separables. En concreto, Minsky y Papert mostraban que los perceptrones no

aprenden ni siquiera la función XOR, ilustrada en la Figura 6, que es la más sencilla de las que no pueden separarse con una línea.

La función XOR (de *exclusive OR* en inglés, o disyunción exclusiva) es una operación lógica que es 1 solo cuando los valores de entrada son distintos (es decir 1 y 0, o 0 y 1), y es 0 cuando los valores de entrada son iguales (1 y 1 o 0 y 0). En términos generales, XOR devuelve un 1 cuando tiene un número impar de valores de entrada que valen 1, y devuelve 0 en caso contrario.

Pese a su sencillez, XOR no es linealmente separable: si representamos los valores (0,0) (1,1) (0,1) y (1,0) en un gráfico, es imposible separar con una recta el (0,0) (1,1) del (0,1) (1,0) para reflejar que XOR devuelve 0 para (0,0) (1,1) y devuelve 1 para (0,1) (1,0). Para poder representar XOR, o cualquier otra función no linealmente separable, se necesitan modelos más complejos que un simple perceptrón.



Fig. 6 Representación de XOR (fuente: Wikipedia).

Por tanto, los investigadores en Inteligencia Artificial se encontraron con limitaciones y dificultades insalvables en la década de los 70. La escasa capacidad de computación de las máquinas impedía procesar grandes cantidades de datos, algo indispensable para entrenar modelos complejos con que abordar problemas reales. Hace apenas

una década, o poco más, que hemos comenzado a solventar este reto.

Dichas limitaciones, combinadas con grandes expectativas incumplidas, dieron lugar a un declive tanto en el interés como en la financiación de la Inteligencia Artificial durante el periodo entre 1974 y 1980. El invierno había llegado.

El inasible sentido común

Pero tarde o temprano llega la primavera. El interés por la Inteligencia Artificial, y los fondos disponibles para su desarrollo, empezaron a aumentar de nuevo a principios de los 80. Durante esa década llegaron al mercado los primeros sistemas expertos, con éxito apreciable. En 1985 el gasto en sistemas de IA en las empresas era de miles de millones de dólares.

Dentro del acercamiento simbólico-lógico, en 1984 nació el primer esfuerzo científico por implementar en una máquina el razonamiento de sentido común, mediante una gigantesca base de datos con todo el conocimiento sobre el mundo que tiene, de media, una persona. Llamado Cyc, hoy en día sigue activo en la compañía Cycorp y atesora decenas de millones de aserciones, reglas o ideas del sentido común aportadas por humanos –por ejemplo, el agua causa humedad y la humedad pudre la comida–, que pueden ser usadas por otros programas.

Sin embargo, de nuevo aparecieron obstáculos. Durante el congreso de 1984 de la Asociación Americana de Inteligencia Artificial, Minsky

y Roger Schank alertaron de que el entusiasmo y la inversión en Inteligencia Artificial conducirían a una nueva decepción. En efecto, en 1987 comenzó el segundo invierno de la Inteligencia Artificial, que alcanzaría su momento más oscuro en 1990.

Mientras tanto la comunidad científica seguía avanzando en las dos escuelas de pensamiento. Uno de los hitos más importantes de la estrategia *bottom-up* y, en particular, del conexionismo, fue el uso del algoritmo de *backpropagation* por parte de David Rumelhart, Geoffrey Hinton y Ronald Williams [11] en 1986.

Entrenando máquinas a partir de datos: el resurgir del conexionismo

Gracias al algoritmo de *backpropagation* es posible entrenar redes mucho más complejas que el *Perceptrón*, con numerosas capas de neuronas ocultas –llamadas así en la jerga– operando entre las capas de entrada y salida y con capacidad, esta vez sí, de modelar problemas complejos. Hoy en día el algoritmo de *backpropagation* es la base de la gran mayoría de modelos de redes neuronales profundas.

El funcionamiento, en términos muy básicos, es el siguiente. Las redes nacen ignorantes, no saben nada sobre el problema que tienen que resolver a partir de los datos que se les van a proporcionar –volviendo al ejemplo de los gatos, no saben si hay o no un gato en la foto–, pero se lanzan y hacen una predicción; esa predicción es cotejada con la realidad, y se mide su grado de error. En función de esta medida se ajustan los pesos en la red, es decir, los coeficientes que deben ser procesados por la neurona.

Se llama *backpropagation* porque se propagan los errores hacia atrás en la red, desde las neuronas de salida (las que están más a la derecha en la Figura 8), a las neuronas de entrada. Por tanto, los errores que comete la red neuronal al entrenarse sirven, gracias al algoritmo *backpropagation*, para determinar los valores de los pesos que lograrían reducir tales errores. Es un proceso iterativo: en cada iteración se van ajustando los pesos en función de los errores cometidos, de forma que estos, y la propia corrección a que se les debe someter, se van reduciendo (ver ejemplo en la Figura 8).

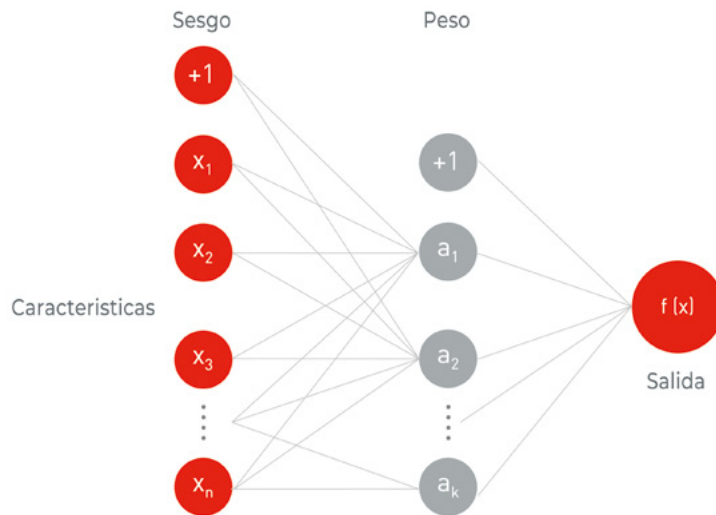


Fig. 8 *Perceptrón multi-capas con una capa oculta (fuente: figura propia).*

Aunque Rumelhart, Hinton y Williams no fueron los primeros en publicar un artículo sobre *backpropagation*, fue su trabajo el que logró calar en la comunidad científica por la claridad con que presenta esta idea.

Igualmente cabe destacar el trabajo de Judea Pearl a finales de los 80, cuando incorporó a la Inteligencia Artificial las teorías de la probabilidad y de la decisión. Algunos de los nuevos métodos propuestos incluyen modelos clave en mi investigación, como las redes bayesianas¹ y los modelos ocultos de Markov², así como la teoría de la información, el modelado estocástico y la optimización. También se desarrollaron los algoritmos evolutivos, inspirados en conceptos de la evolución biológica como la reproducción, las mutaciones, la recombinación de genes y la selección.

En los algoritmos evolutivos se generan soluciones candidatas al problema que se quiere resolver. Cada solución juega el papel de un individuo en una población; se van seleccionando las soluciones de mayor calidad aplicando ciertos criterios predefinidos, y estas soluciones se hacen evolucionar aplicando los conceptos anteriores de reproducción, mutaciones, etcétera.

¹ Una red bayesiana es un modelo gráfico probabilístico que representa una serie de variables y sus dependencias probabilísticas en forma de un gráfico donde los nodos son las variables, y las conexiones entre nodos representan las dependencias entre variables.

² Un modelo oculto de Markov es un modelo estadístico de un sistema dinámico que puede representarse como la red bayesiana dinámica más sencilla.

El objetivo es que, tras un cierto número de generaciones, las soluciones encontradas sean cada vez mejores. La ventaja es que estos algoritmos se pueden aplicar para resolver multitud de problemas. La desventaja es su complejidad computacional, que dificulta su aplicación a muchos problemas reales.

Desde mediados de los años 90, precisamente cuando comencé mi doctorado en el MIT, hasta hoy en día –y especialmente en la última década–, se ha producido un avance muy significativo en las técnicas de aprendizaje estadístico por ordenador basadas en datos (*statistical machine learning*), que pertenecen al enfoque *bottom-up*.

El acceso a cantidades ingentes de datos –Big Data–; la disponibilidad de procesadores muy potentes a bajo coste; y el desarrollo de redes neuronales profundas y complejas, los modelos llamados de *deep learning* [14] (ver Figuras 9 y 10), son los tres factores que han confluído para instalar hoy día a la Inteligencia Artificial en una “primavera perpetua”, en palabras del profesor de la Universidad de Stanford Andrew Ng, con quien también coincidí en MIT.

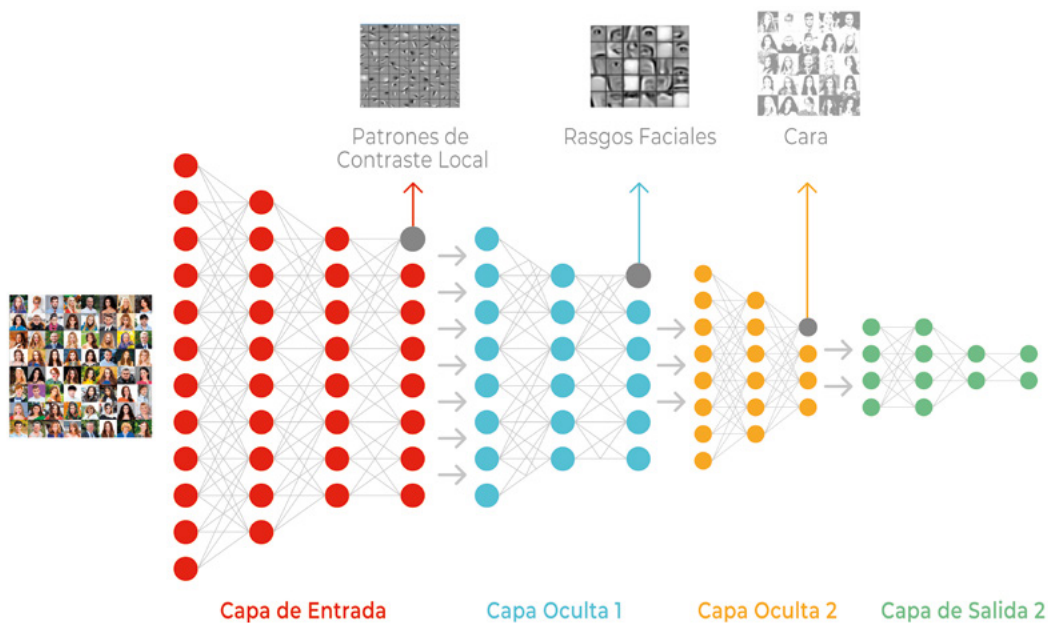


Fig. 9 Ejemplo de red neuronal profunda (Deep neural Network) para el procesamiento y análisis de caras en imágenes
(fuente: <https://cdn.edureka.co/blog/wp-content/uploads/2017/05/deep-neural-network-what-is-deep-learning-edureka.png>).

En los últimos años –como puede observarse en la Figura 7–, con el éxito de los métodos de aprendizaje de *deep learning* se ha producido un fuerte resurgir del acercamiento *bottom-up* y en particular del conexionismo, dentro de la Inteligencia Artificial. Así lo atestigua el hecho de que los pioneros del *deep learning* Yoshua Bengio, Geoffrey Hinton y Yann LeCun hayan recibido en 2019 el premio Turing, el equivalente al Nobel en informática.

La Inteligencia Artificial –no queda ya alguna duda– forma parte de nuestro presente. De ello hablaremos a continuación.

Grafico Completo de Redes Neuronales

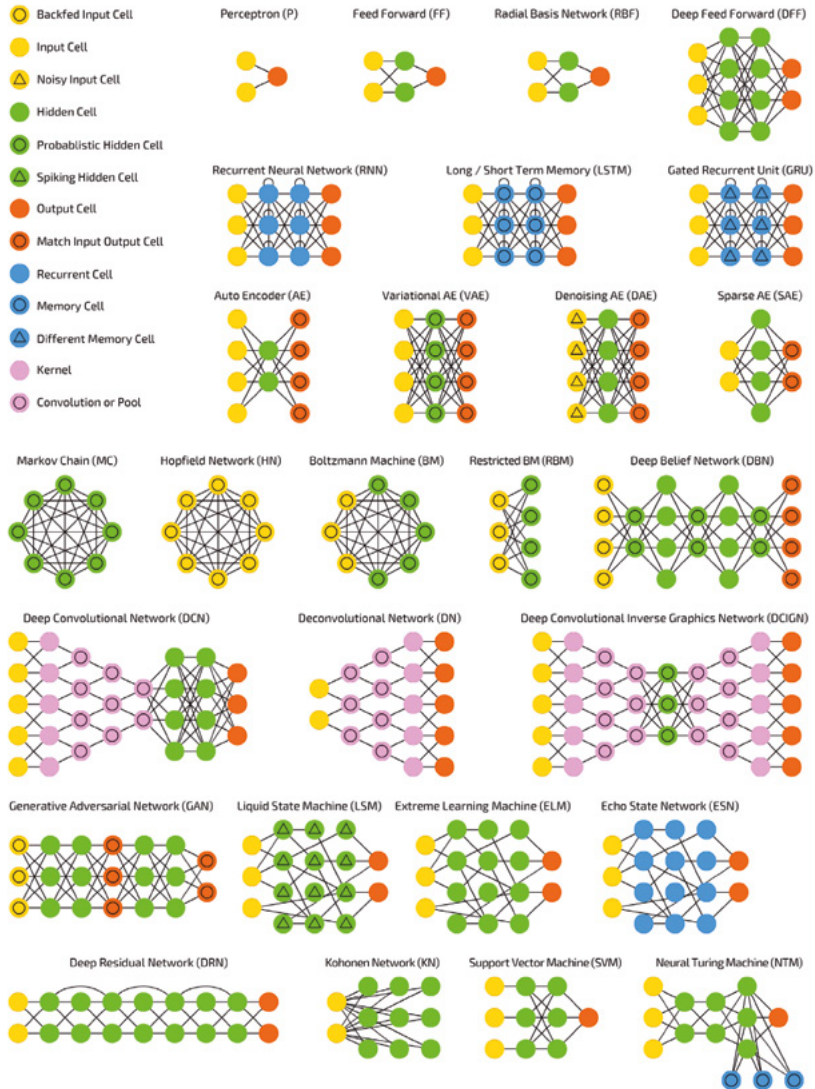


Fig. 10 Ejemplos de diferentes arquitecturas de redes neuronales profundas
(fuente: asimovinstitute.org, Fjodor van Veen).

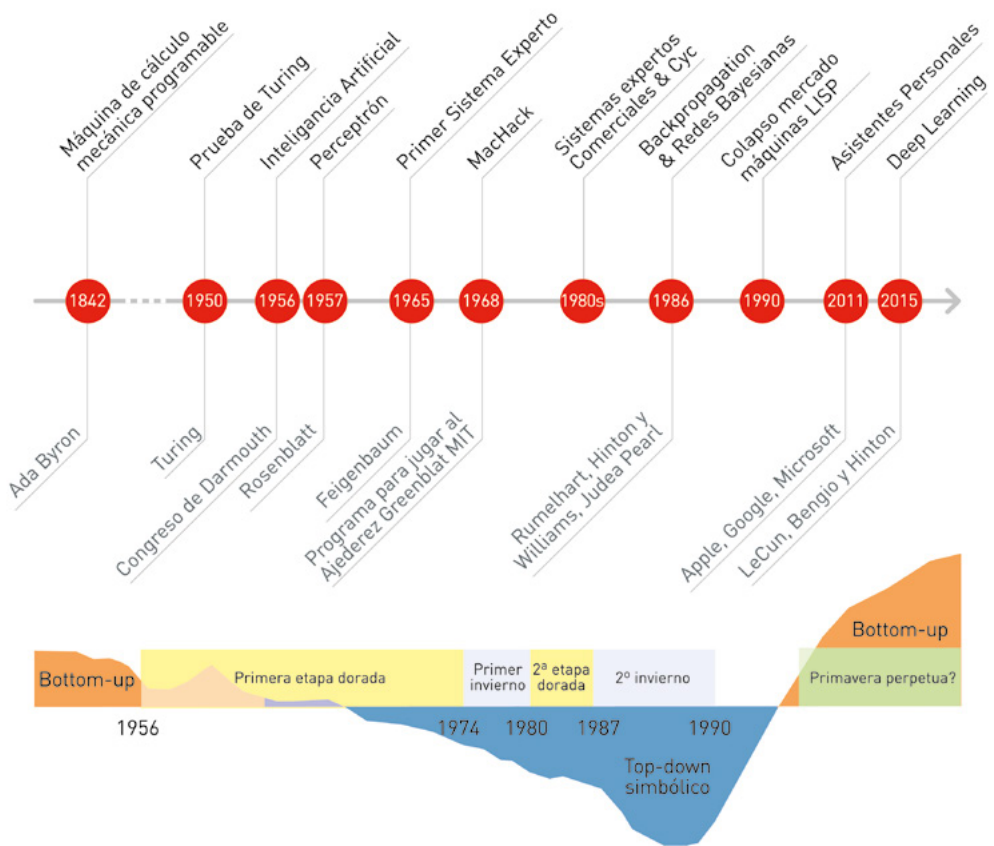


Fig. 7 Cronología de la Inteligencia Artificial. Puede observarse la alternancia en la prevalencia de los modelos simbólico-lógicos (en azul) y los modelos bottom-up (en naranja).



LAS SIETE CUESTIONES FUNDACIONALES DE LA IA

En 1955, investigadores pioneros en Inteligencia Artificial propusieron analizar en la Conferencia de Dartmouth (New Hampshire, EE.UU.) la conjetura de que, en principio, todos los aspectos del aprendizaje humano, y en general de la inteligencia, pueden ser descritos de manera lo bastante precisa como para que una máquina pueda simularlas. Desmenuzaron en siete cuestiones el desafío de lograr una Inteligencia Artificial, muchas de las cuales perduran hoy en día como retos a abordar:

- **Capacidad computacional:** “La velocidad y memoria de los ordenadores actuales podría ser insuficiente para simular muchas de las funciones superiores del cerebro humano, pero el principal obstáculo no es la falta de capacidad de las máquinas, sino nuestra incapacidad para programar (...)”, escribieron John McCarthy, Marvin Minsky, Claude Shannon y Herbert Simon.
- **Lenguaje.** Los humanos, en gran parte, usamos palabras para pensar. ¿Se puede programar a un ordenador para que tenga lenguaje –y pueda, por ejemplo, integrar términos nuevos en frases con significado, como hacemos nosotros–?

- **Capacidad de abstracción en las redes neuronales:**

¿Cómo se puede disponer una red de hipotéticas neuronas para lograr que formen conceptos?

- **¿Es eficiente esta manera de resolver el problema?**

Los expertos echaban en falta un método para dimensionar una operación computacional, para decidir si abordarla o por el contrario buscar otras estrategias.

- **Superarse a uno mismo.** “Probablemente una máquina verdaderamente inteligente llevará a cabo tareas que pueden ser descritas como de superación personal”.

- **Pensamiento abstracto.** “Puede valer la pena hacer un intento de clasificar las abstracciones, y describir métodos por los que las máquinas podrían generar abstracciones a partir de estímulos sensoriales y otros datos”.

- **Aleatoriedad y creatividad.** “Una conjetura atractiva, y sin embargo claramente incompleta, es que la diferencia entre el pensamiento creativo y el pensamiento competente poco imaginativo reside en la introducción de una cierta aleatoriedad”.



[CAPÍTULO 3]

Aquí y ahora

La IA ya transforma nuestras vidas

Si la electricidad impulsó la Segunda Revolución Industrial, e internet y los ordenadores personales la Tercera, la Inteligencia Artificial está provocando la Cuarta. Los sistemas de IA están transformando la medicina, el transporte, el sector energético, nuestra elección de contenidos de cultura y ocio, la economía en su conjunto y por supuesto la ciencia, a pasos agigantados. Pese a sus limitaciones, y a que queda aún muy lejos el sueño de una inteligencia equiparable a la humana, todo apunta a que en poco tiempo el planeta estará envuelto en un sistema circulatorio que lo irrigará capilarmente con Inteligencia Artificial.

“Soy un ser humano. Me asusto cuando veo algo que supera con mucho mi capacidad de comprensión”, declaró el campeón mundial de ajedrez Gary Kasparov tras su derrota contra el programa *Deep Blue*, de IBM, el 11 de mayo de 1997. Era la primera vez que un campeón de ajedrez perdía contra una máquina.

“Enérgica y brutalmente, el ordenador *Deep Blue* de IBM arrebató a la humanidad, al menos temporalmente, el puesto de mejor ajedrecista del planeta”, decía el *New York Times* al inicio de su crónica.

La gran repercusión mediática de este logro de la Inteligencia Artificial, impensable hace solo tres décadas, ha contribuido al exorbitante aumento en el interés por esta disciplina.

Pero hay muchos más hitos. En 2005 un vehículo autónomo desarrollado en la Universidad de Stanford, EE. UU., recorrió autónomamente 210 kilómetros en el desierto y se convirtió así en el primero en superar el DARPA *Grand Challenge*, una carrera de vehículos sin conductor creada solo un año antes por la Agencia de Investigación en Proyectos Avanzados de Defensa estadounidense, más conocida por su acrónimo DARPA. Este hito demostró que la conducción autónoma era posible y fue el punto de partida de los miles de millones de dólares de inversión en la conducción sin conductor.

Poco más tarde, en 2011, el programa de IA *Watson*, de IBM, venció a dos de los campeones humanos del concurso estadounidense de preguntas y respuestas *Jeopardy!*. Ese fue el año en que muchos de nosotros empezamos a hablar con asistentes personales instalados en nuestros teléfonos móviles –*Siri*, *Cortana* y *Google Now*–, programas que permiten a sus usuarios utilizar la voz y el lenguaje natural para hacer preguntas y dar instrucciones, y

que desde 2015 están también en nuestro hogar –*Alexa, Google Home*–.

En 2016 otro enfrentamiento hombre-máquina, con victoria para el contendiente no biológico, conquistó portadas en todo el planeta.

El programa *AlphaGo*, desarrollado por la compañía *DeepMind*, de Google, venció en el juego chino *Go* a uno de los mejores jugadores humanos del mundo, Lee Sedol. Las reglas del *Go* son más simples que las del ajedrez, pero el número de configuraciones a tener en cuenta es mucho mayor. Además, el juego requiere grandes dosis de intuición, por lo que dominarlo parecía del todo imposible para una máquina. *AlphaGo* solo pudo lograrlo recurriendo a su capacidad de aprendizaje, mucho más desarrollada que la de *Deep Blue*.

También aprendió mucho el programa de la Universidad Carnegie Mellon *Libratus* cuando, en enero de 2017, se enfrentó al póker a cuatro de los mejores jugadores del mundo. Al concluir veinte días de torneo las ganancias de *Libratus* superaban en más de 1.700 dólares las de los humanos. Fue una victoria relevante, porque el póker representa un nivel adicional de complejidad: en el ajedrez y en el *Go* el tablero, con todas sus piezas, es visible para ambos jugadores, mientras que en el póker desconocemos las cartas de nuestros contrincantes, que, además, pueden marcarse faroles. Es decir, el póker es un juego de información incompleta y por tanto mucho más difícil de jugar computacionalmente.

Otros hitos recientes se producen en los combates máquina-máquina. En diciembre de 2017 *AlphaZero*, de *DeepMind*, no solo venció al mejor jugador de ajedrez del mundo –que por cierto es un programa de ordenador llamado *Stockfish*–, sino que hizo gala de su capacidad de aprender el juego por sí solo. Ambos programas se enfrentaron en una serie de cien partidas, de las que *AlphaZero* ganó 28 y el resto quedaron en tablas; para lograr la hazaña, al programa de *DeepMind* le bastó conocer las reglas del ajedrez y dedicar cuatro horas a entrenarse, jugando contra sí mismo millones de veces.

No menos relevante es que un sistema de procesamiento de lenguaje desarrollado por Alibaba –la gran compañía china de comercio electrónico– superara en 2018 los resultados de los humanos en la prueba de comprensión lectora de la Universidad de Stanford (EE. UU.). El *Stanford Question Answering Dataset* es un conjunto de cien mil preguntas, que hacen referencia a más de 500 artículos de Wikipedia.

En la salud, el ocio, la seguridad, la economía...

Más allá de estos hitos, que pueden parecer alejados de la aplicación práctica, la Inteligencia Artificial ocupa ya un lugar importante en multitud de esferas de nuestra vida. Convivimos con la IA probablemente sin saberlo. Está presente en los sistemas de búsqueda y recomendación de información, contenido, productos o amigos que utilizamos en nuestro día a día, como Netflix, Spotify,

Facebook, y en cualquier servicio de noticias o de búsqueda en internet. También en las aplicaciones para la cámara del móvil que detectan automáticamente las caras en las fotos; en los asistentes personales de móviles y hogares; en *chatbots* conversacionales; y en las ciudades inteligentes, para por ejemplo predecir el tráfico.

La IA interviene en la compraventa de acciones, la adjudicación de créditos, la contratación de seguros y la fijación de tarifas, entre otras muchas decisiones que marcan el ritmo de los mercados financieros y las empresas. En el ámbito de la salud funcionan ya los sistemas de diagnóstico automático a partir de historiales clínicos, así como programas de análisis de imágenes médicas, para asistir el diagnóstico radiológico, y de ADN, para por ejemplo detectar mutaciones o variantes genéticas asociadas a enfermedades.

La toma de decisiones de las Administraciones Públicas se apoya igualmente en la Inteligencia Artificial, con sistemas de vigilancia, de soporte a decisiones judiciales o de clasificación y jerarquización del alumnado. Numerosas aplicaciones se destinan al ámbito de la seguridad y la defensa, desde en el control de viajeros en las fronteras y la adjudicación de visados, hasta para fabricar armas autónomas.

La industria, y en general los procesos productivos, llevan décadas utilizando robots industriales, al igual que sistemas de planificación y predicción de la demanda o de la producción.

Y, por supuesto, sin el apoyo de la Inteligencia Artificial no podríamos soñar con tener vehículos autónomos, una predicción meteorológica certera a medio plazo ni, en general, avances en numerosas áreas de conocimiento. La IA empieza a convertirse en un actor importante de la investigación científica, interviniendo en modelos físicos de toda clase de fenómenos y procesos, en la predicción de la estructura tridimensional de las proteínas, en el diseño de fármacos... La lista es larga.

No hay duda de que la IA tiene un potencial inmenso para construir una sociedad mejor, y ese es el motor de mi trabajo.

El sustrato físico

Si en la evolución humana el aumento de la capacidad cognitiva va de la mano de cambios biológicos, también la inteligencia de las máquinas depende del sustrato físico en que se implementa. No solo de *software* vive la Inteligencia Artificial. En el desarrollo de la IA, tanto en sus aplicaciones prácticas como en la consecución de los hitos antes descritos, han tenido un papel clave los avances en el *hardware*, en particular en los sistemas de procesamiento y almacenamiento a gran escala, distribuidos y en paralelo.

Sin los potentes procesadores actuales tampoco existirían hoy los complejos modelos de *deep learning* o aprendizaje profundo, basados en redes neuronales con muchas capas de procesamiento de información (ver Figuras 9 y 10). Es el nuevo hardware, a menudo

optimizado para esta tarea específica, el que permite entrenar a los modelos, alimentándolos con grandes cantidades de datos en un tiempo y con un consumo energético razonables.

En los últimos años hemos pasado de utilizar procesadores de propósito general (CPUs y GPUs o *graphics processing units*) a procesadores especializados, optimizados para este tipo de modelos de IA (FPGAs o *field-programmable gate arrays* y ASICs o *application-specific integrated circuits*, como la TPU desarrollada por Google). Las figuras 11 y 12 ilustran la evolución en la capacidad de computación. La Figura 11 muestra la famosa Ley de Moore desde 1971, según la cual el número de transistores que podemos integrar en un circuito por el mismo precio se duplica cada año, o año y medio.

La parte de la izquierda representa la Ley de Moore en CPUs, y la derecha en GPUs. En la Figura 12 se ve dónde se sitúan los distintos tipos de procesadores en función de su flexibilidad y facilidad de uso, versus su eficiencia energética y rendimiento.

El lector interesado puede encontrar un resumen de los principales procesadores utilizados para el aprendizaje automático en [22]. Tampoco podemos obviar que transmitir y procesar las enormes cantidades de datos con que se trabaja hoy día requiere un elevadísimo consumo energético (Ver recuadro ¿Un planeta inteligente poco sostenible?).

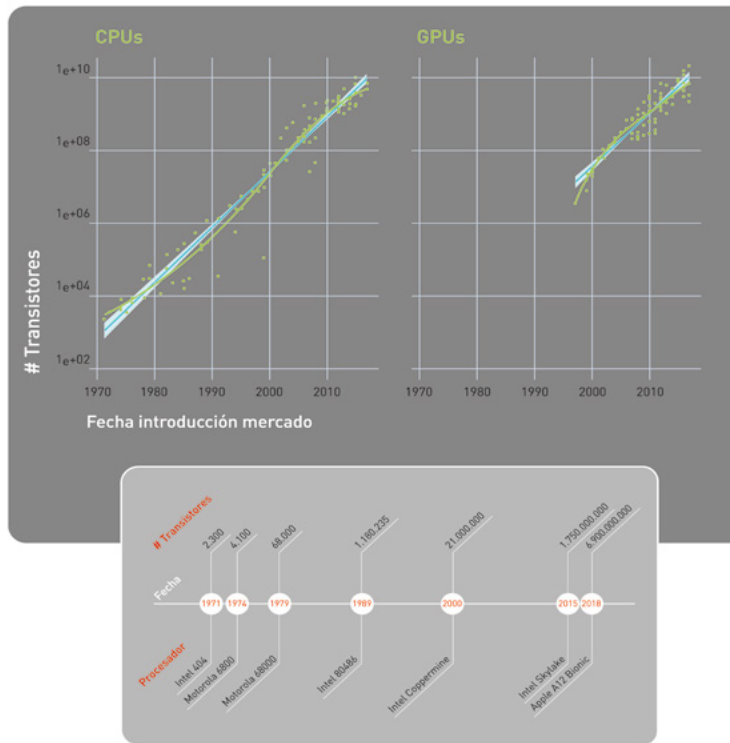


Fig. 11 Ley de Moore desde 1971 para CPUs (izda.) y GPUs (dcha.)
(fuente: Figuras propias a partir de <http://coulmont.com> basado en datos de Wikipedia).

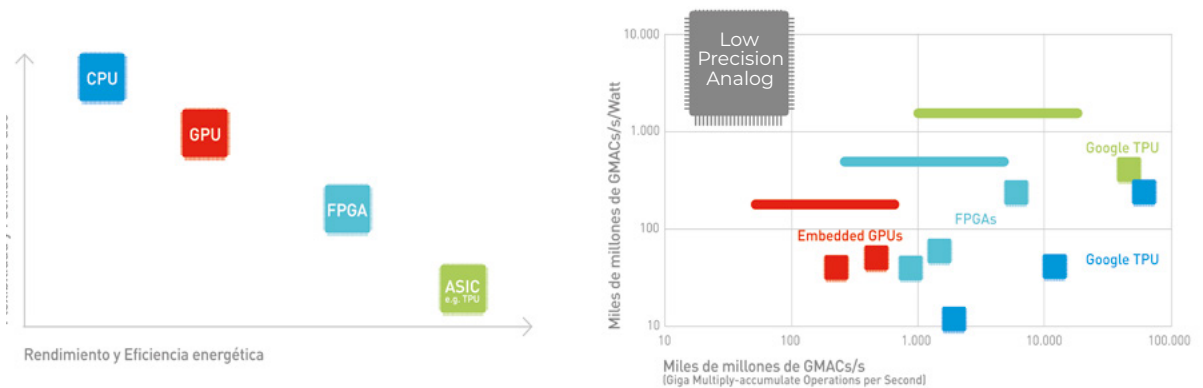


Fig. 12 Arriba: diferentes tipos de procesadores utilizados en sistemas de deep learning; abajo: sendimiento vs eficiencia de distintos tipos de procesadores
(fuente: figuras propias a partir de datos de Chris Rowen, Cognitive Ventures).



¿UN PLANETA INTELIGENTE PERO POCO SOSTENIBLE?

El experto en sostenibilidad de nuevas tecnologías Anders Andrae estima en un estudio reciente [26] que, si no mejoramos drásticamente la eficiencia energética de los procesadores, dentro de solo seis años, en 2025, la industria TIC podría consumir el 20% de toda la electricidad del planeta y emitir hasta un 5,5% de las emisiones de CO₂.

La predicción de demanda de electricidad iría de 200-300 teravatios-hora de electricidad al año, a 1.200-3.000 teravatios-hora en 2025. Los centros de datos podrían producir 1,9 Gigatoneladas de CO₂, lo que supone un 3,2% del total de las emisiones del planeta.

No podemos obviar la demanda energética que conllevaría una implantación masiva de los sistemas de Inteligencia Artificial. Por ello, y siempre desde la perspectiva del desarrollo y adopción de la Inteligencia Artificial para el bien social, hemos de invertir en investigación e implementación de soluciones que sean sostenibles desde el punto de vista medioambiental.

Necesitamos apoyarnos en algoritmos de Inteligencia Artificial para modelar el clima, por ejemplo, o construir modelos predictivos del cambio climático o del consumo energético. Pero también hemos de considerar el impacto medioambiental del desarrollo de la Inteligencia Artificial.

El campeón de ajedrez que desconoce qué es el ajedrez

El despliegue de los hitos de la Inteligencia Artificial durante la última década puede resultar abrumador, incluso inquietante. Basta una pregunta para ganar perspectiva y reconsiderar la situación: ¿sabe *Libratus* que está jugando –y ganando– al póker?

Los sistemas de Inteligencia Artificial suelen dividirse en tres tipos, atendiendo a su grado de competencia. La *IA específica* hace referencia a sistemas capaces de realizar una tarea concreta, como jugar al ajedrez, reconocer el habla, imágenes o procesar texto, incluso mejor que un humano. Pero solamente saben hacer esa tarea. Este es el tipo de Inteligencia Artificial que tenemos hoy en día.

Los sistemas con IA general, en cambio, exhiben una inteligencia similar a la humana: múltiple, adaptable, flexible, eficiente, incremental... Esta sería la aspiración última de la Inteligencia Artificial, y estamos aún muy lejos de hacerla realidad.

La tercera categoría corresponde a los sistemas con súper-inteligencia, término un tanto controvertido referido al desarrollo de sistemas con inteligencia superior a la humana, tal y como propone el filósofo de la Universidad de Oxford, Reino Unido, Nick Bostrom [23].

En la actualidad disponemos de sistemas de Inteligencia Artificial específica, es decir, capaces de realizar automática y autónomamente

una tarea concreta, pero esa y solo esa. Aunque un algoritmo juegue mejor que el mejor de los humanos, es incapaz de hacer cualquier otra tarea. De hecho, tampoco sabe qué es el ajedrez, y tendría dificultades para jugar si introduyéramos cambios en las reglas.

Los sistemas actuales manifiestan un tipo limitado de inteligencia: son incapaces, entre otras cosas, de generalizar y extender a otros ámbitos sus niveles de competencia en una tarea de manera automática, como haría un humano.

Inteligencia Artificial ‘versus’ Inteligencia Humana

El éxito reciente de los sistemas de Inteligencia Artificial quizás esté desviando la atención respecto al grado de avance en cuestiones profundas del área. Hay problemas fundacionales de la Inteligencia Artificial que todavía están por resolver.

Algunos de estos problemas derivan de tres limitaciones básicas, tres aspectos que marcan la diferencia entre los mecanismos de aprendizaje automático disponibles actualmente y los sistemas inteligentes biológicos. Lo explica Jeff Hawkins en un artículo para *IEEE Spectrum* [24] que resumo en los siguientes párrafos.

Los sistemas de aprendizaje biológicos –cerebros– son capaces de aprender rápidamente. Unas pocas observaciones o experiencias táctiles suelen ser suficientes para aprender algo nuevo, a diferencia de los millones de ejemplos que necesitan los sistemas de Inteligencia Artificial actuales.

Además, los sistemas biológicos aprenden de manera incremental, es decir, agregamos conocimiento nuevo sin tener que volver a aprender todo desde cero ni perder conocimiento anterior. Es más, hacemos todo lo anterior de manera continua. Aprendemos conforme interaccionamos con el mundo físico, y nunca dejamos de hacerlo. El aprendizaje rápido, incremental y constante es un elemento esencial que permite a los sistemas biológicos inteligentes adaptarse a un entorno cambiante, y sobrevivir.

La neurona es un elemento clave en el aprendizaje biológico. La complejidad de las neuronas biológicas y sus conexiones es lo que dota al cerebro de la capacidad para aprender. Hoy sabemos que el cerebro tiene plasticidad y que constantemente se están creando nuevas neuronas –el fenómeno llamado *neurogénesis*– y sinapsis –*sinaptogénesis*–. Se estima que diariamente se sustituyen hasta un 40% de las sinapsis en cada neurona, lo que implica que hay en marcha un proceso constante de creación de nuevas conexiones neuronales.

Los sistemas de aprendizaje artificiales no tendrían por qué reproducir exactamente el funcionamiento de las neuronas biológicas, pero esta capacidad para disponer de un aprendizaje rápido, incremental y constante, caracterizado por la destrucción y creación de sinapsis, es esencial.

Otra cuestión clave es que el cerebro procesa la información mediante representaciones distribuidas no densas, o dispersas

(*sparse distributed representations* o SDRs) [25], una terminología referida a que solo un conjunto reducido de neuronas está activo en cada momento del tiempo. Este grupo de neuronas encendidas cambia de un instante a otro en función de lo que haga el ser vivo, pero es en cualquier caso pequeño.

Esta configuración con representaciones distribuidas de la información es robusta respecto a los errores y a la incertidumbre. Además, goza de dos propiedades interesantes: la propiedad del solape, que permite detectar rápidamente si dos percepciones son idénticas o diferentes; y la propiedad de la unión, que permite al cerebro mantener varias representaciones en paralelo.

Por ejemplo, si oímos un animal que se mueve entre unos arbustos, pero no hemos podido verlo, podría ser un conejo, una ardilla o una rata. Dado que las representaciones en el cerebro son dispersas, nuestro cerebro puede activar tres SDRs al mismo tiempo -la del conejo, la ardilla y la rata- sin interferencias entre ellas. Gracias a esta propiedad de poder unir SDRs el cerebro gestiona, opera y toma decisiones aún sin disponer de información completa, es decir, con incertidumbre.

Además, el aprendizaje tiene *cuerpo*. Nuestro cerebro recibe la información de los distintos sentidos, y esta información cambia según nos movemos y actuamos en nuestro entorno.

Un sistema de Inteligencia Artificial no tiene por qué tener un cuerpo físico, pero sí la capacidad de actuar sobre su entorno físico y virtual, y recibir una reacción *-feedback-* a sus acciones. Los sistemas de aprendizaje con refuerzo *-reinforcement learning-* hacen algo similar, y son instrumentales en la consecución de algunos de los hitos previamente descritos, como los logrados por *AlphaZero*.

El cerebro es capaz de integrar la información captada por los distintos sentidos y por el sistema motor, para poder no sólo procesar, reconocer y decidir en función de lo percibido, sino también actuar. Esta integración sensorial-motora es básica en el funcionamiento del cerebro y probablemente deberá también serlo en los sistemas de Inteligencia Artificial.

Finalmente, en el cerebro la información sensorial es procesada por un sistema jerárquico: conforme la información va pasando de un nivel a otro se van calculando características cada vez más complejas y abstractas de lo que se está percibiendo.

Los modelos de aprendizaje profundo también utilizan jerarquías, pero mucho más complicadas de las empleadas por el cerebro humano, con decenas o una centena de niveles y centenas o miles de millones de parámetros. Estas redes neuronales profundas *-modelos de deep learning-* necesitan millones de observaciones para aprender un patrón. A nuestro cerebro, en cambio, le bastan pocos niveles de jerarquías y también, como ya

he dicho, pocos ejemplos para aprender. El cerebro aprende de manera mucho más eficiente que los modelos computacionales de hoy en día.

Estas limitaciones, sin embargo, no están frenando el creciente impacto de la Inteligencia Artificial en nuestras vidas. La IA es ya un elemento integral de la Cuarta Revolución Industrial en la que estamos inmersos.



CONOCIMIENTO PARA TODOS, ACTUALIZADO AL INSTANTE

Se ha afirmado que la Inteligencia Artificial será al siglo XXI lo que la electricidad fue al XX. Aunque comparar revoluciones siempre es un ejercicio arriesgado, lo cierto es que la IA toca prácticamente todos los sectores, es escalable, invisible y permite la actualización constante. Funciona, además, como herramienta para predecir el futuro, y por tanto para ayudar en cualquier proceso de toma de decisiones.

Podemos imaginar escenarios sin precedente en la historia de la humanidad. Escenarios donde, por ejemplo, una red de sistemas de IA podría muy rápidamente incorporar los últimos métodos computacionales en el diagnóstico de una enfermedad, y desplegarlos a toda la población del planeta. El equivalente analógico, que consistiría en una incorporación casi instantánea del conocimiento a todos los médicos del planeta, es simplemente inviable.

La Cuarta Revolución Industrial

En los últimos tres siglos hemos vivido cuatro revoluciones industriales, ilustradas en la Figura 12. La primera tuvo lugar entre los siglos XVIII y XIX en Europa y Norteamérica, y se corresponde con el momento histórico en que sociedades que eran mayoritariamente agrarias y rurales se convirtieron en industriales y urbanas. El principal factor impulsor de esta revolución fue la invención de la máquina de vapor, junto con el desarrollo de las industrias textil y metalúrgica.

La Segunda Revolución Industrial –conocida como la Revolución Tecnológica– ocurrió justo antes de la Primera Guerra Mundial, entre 1870 y 1914, y se corresponde con un crecimiento de las industrias anteriores y el desarrollo de otras nuevas, como la del acero y el petróleo, y con la llegada de la electricidad. Los avances tecnológicos más importantes de esta revolución incluyen el teléfono, la bombilla, el fonógrafo y el motor de combustión interna.

La Tercera Revolución Industrial es la Revolución Digital y hace referencia a la transición de dispositivos mecánicos y analógicos al uso de tecnologías digitales. Comenzó en los años 80 y continúa hoy en día. Los avances tecnológicos clave incluyen los ordenadores personales, internet y el desarrollo de otras tecnologías de la información y las comunicaciones (TICs).

Finalmente, la Cuarta Revolución Industrial se apoya en avances de la Revolución Digital, pero incorpora la ubicuidad de la tecnología digital tanto en nuestra sociedad como en nuestro cuerpo, y la unión creciente entre los mundos físico, biológico y digital.

Los avances tecnológicos que hacen posible esta nueva revolución incluyen la robótica, la Inteligencia Artificial –alimentada con Big Data– la nanotecnología, la biotecnología, la ingeniería genética, el internet de las cosas, los vehículos autónomos, las impresoras en tres dimensiones y la informática cuántica. El término fue presentado y reconocido globalmente durante el Foro Económico Mundial en 2016 por su fundador, el economista alemán Klaus Schwab.



Fig. 13 Las cuatro revoluciones industriales desde el siglo XVIII.

Una IA transversal e invisible

La Inteligencia Artificial tiene un conjunto de características que contribuyen a convertirla en un elemento clave en esta Cuarta Revolución Industrial. Para empezar, es transversal e invisible.

Las técnicas de Inteligencia Artificial, como hemos explicado, pueden utilizarse en un sinnúmero de aplicaciones en medicina, energía, transporte y educación; en la investigación científica, en los sistemas de producción, la logística, los servicios digitales y la prestación de servicios públicos y privados. Y en la gran mayoría de estas aplicaciones, los sistemas de IA consisten en software instalado en el corazón de aplicaciones y servicios. Es decir, son invisibles. Estas dos propiedades, la transversalidad y la invisibilidad, sitúan a la Inteligencia Artificial en el núcleo de la Cuarta Revolución Industrial, con un papel similar al jugado por la electricidad en la Segunda Revolución Industrial.

Otras características de la IA son la complejidad, la escalabilidad y la actualización constante. Los sistemas actuales de Inteligencia Artificial basados en modelos de aprendizaje profundo son complejos, con cientos de capas de neuronas y millones de parámetros. Esta complejidad tiene un doble filo.

Por una parte, dificulta la interpretación de los modelos, lo que, en ciertas aplicaciones, por ejemplo, en medicina o en educación, es un obstáculo limitante. Pero al mismo tiempo es esta complejidad la que permite a la Inteligencia Artificial procesar cantidades ingentes

de datos, y realizar tareas con niveles de competencia superiores a los de los humanos. Es decir, los complejos modelos de aprendizaje profundo dotan a los sistemas de IA de gran escalabilidad.

En muchos casos solo podemos extraer conocimiento y valor del Big Data a través del uso de sistemas de Inteligencia Artificial, ya que los métodos tradicionales no pueden procesar volúmenes de datos tan grandes que, además, son variados, generados a gran velocidad y no estructurados –es decir, son texto, audio, vídeo, imágenes o proceden de sensores–. Para los sistemas de Inteligencia Artificial esto no es un problema gracias a que son altamente escalables, esto es, consisten fundamentalmente en software que puede estar conectado con miles o millones de otros sistemas de IA, dando lugar a una red colectiva.

Hoy en día, sin IA seríamos incapaces de analizar e interpretar las enormes cantidades de texto, imágenes, audio o vídeo que existen. Entre otras cosas, no podríamos buscar información en internet. En astronomía, física, biología, química, meteorología o medicina cada vez generamos más datos, datos a los que, por su complejidad y volumen, tampoco podríamos sacar partido. Algo similar sucede en la economía y las finanzas, en el comercio electrónico o en el transporte, por citar otras áreas. Básicamente, cualquier aplicación que se beneficie del análisis de grandes cantidades de datos no estructurados es susceptible de ser transformada por la IA.

La capacidad de actualizar el software de manera masiva es también una propiedad clave. Combinada con la escalabilidad, la actualización

masiva permite a la IA tener impacto en la vida de cientos o incluso miles de millones de personas en poco tiempo.

Otra propiedad definitoria de la IA es su habilidad para predecir. Los sistemas de Inteligencia Artificial pueden utilizarse para la toma de decisiones automáticas y para predecir situaciones futuras. De hecho, aspiramos a que las decisiones algorítmicas basadas en IA entrenada con datos carezcan de las limitaciones de las decisiones humanas – conflictos de interés, sesgos, intereses propios, corrupción...–, y sean por tanto más justas y objetivas.

Sin embargo, esto no será necesariamente así si ignoramos las limitaciones de las decisiones que toman los algoritmos, como explicaré posteriormente.

Ni ciencia ficción, ni moda

La presencia de la IA en nuestras vidas, y su capacidad para mejorar la sociedad, son innegables. Las grandes potencias mundiales –tanto empresas como gobiernos– han comprendido que el liderazgo en Inteligencia Artificial debe ser no solo económico, sino también político y social.

Por eso en los últimos dos años los gobiernos de más de una veintena de países –incluyendo EE. UU., China, Canadá, Francia, Taiwán, Singapur, México, Suecia, India, Australia y Finlandia– han elaborado estrategias nacionales sobre Inteligencia Artificial, tal y como refleja la Figura 14.

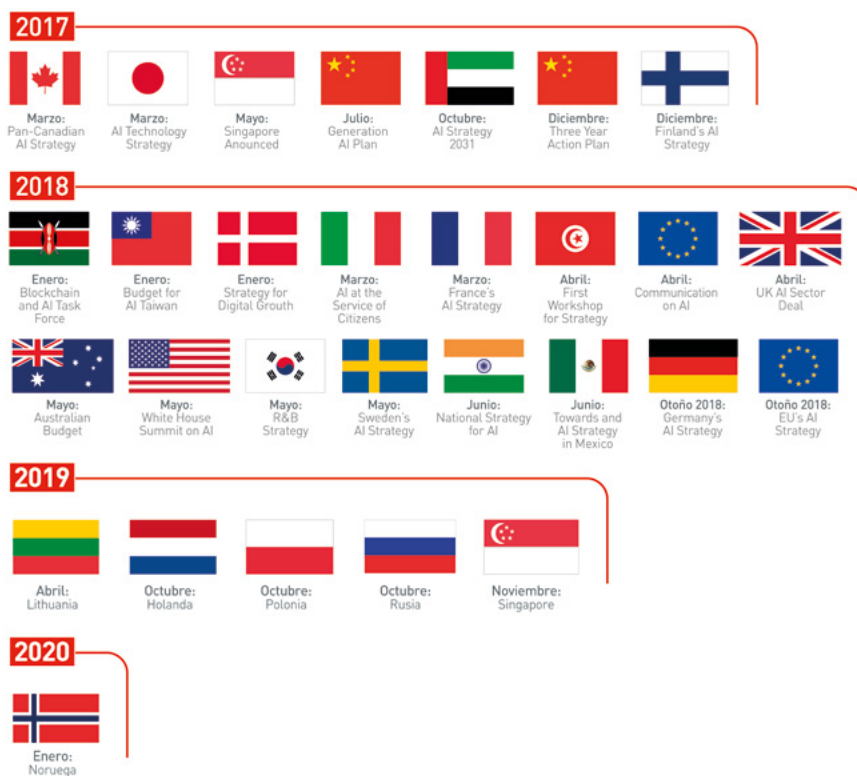


Fig. 14 Línea temporal de las estrategias nacionales sobre Inteligencia Artificial (fuente: figura propia a partir de datos de Tim Dutton).

En Europa, el grupo de expertos sobre IA creado por la Comisión Europea –del que soy miembro reserva– publicó en 2019 un conjunto de guías éticas para la Inteligencia Artificial y unas recomendaciones sobre inversiones y políticas en IA. Esta estrategia europea propone actuaciones conjuntas para una cooperación más estrecha y eficiente entre los Estados miembros, Noruega, Suiza y la Comisión en cuatro ámbitos clave: aumentar la inversión, lograr que haya más datos disponibles (para entrenar a los algoritmos de IA), fomentar

el talento y garantizar la confianza. Además, en febrero de 2020 se espera que la Comisión Europea publique un documento estratégico con recomendaciones de regulación de la Inteligencia Artificial, especialmente en aplicaciones con impacto en la vida de las personas, como la salud o el transporte.

El entonces vicepresidente de la Comisión, el estonio Andrus Ansip, afirmaba en una nota de prensa: “Hemos acordado colaborar a fin de reunir datos –la materia prima de la Inteligencia Artificial– en sectores como la asistencia sanitaria para mejorar el diagnóstico y el tratamiento del cáncer. Tendremos que actuar coordinadamente para lograr el objetivo de inversiones públicas y privadas de al menos 20.000 millones de euros, algo fundamental para el crecimiento y el empleo. La Inteligencia Artificial no es un capricho, es nuestro futuro”.

Esta inversión anual en I+D dedicada a la IA de 20.000 millones de euros anuales en el periodo 2021-2027 es necesaria para reducir la brecha entre la inversión en IA de Europa –de entre 3.000 y 15.000 millones de euros–, y la de Asia y Norteamérica. La Comisión propone un reparto de esfuerzos de inversión entre diferentes actores europeos para conseguir alcanzar la inversión mencionada: 1.000 millones al año serían a cargo de programas de I+D (Horizon Europe y Europa Digital); los Estados miembros invertirían 6.000 millones anuales, parte de los cuales podría proceder de fondos europeos; y 13.000 millones de euros deberían proceder del sector privado.

En febrero de 2020, la Comisión Europea publicó tres documentos estratégicos en el contexto de la Inteligencia Artificial.

En primer lugar, un libro blanco³ sobre Inteligencia Artificial donde la Comisión Europea presenta un marco para el desarrollo de una Inteligencia Artificial confiable, basada en la excelencia y la confianza. Este documento propone la definición de reglas claras en aplicaciones de alto riesgo de la Inteligencia Artificial, incluyendo la salud, el transporte o la policía. El objetivo es garantizar que los sistemas de IA son transparentes, supervisados por humanos y susceptibles de ser evaluados y certificados por autoridades externas. También se incluye la necesidad de asegurar que los datos con los que se entrenan estos sistemas no tienen sesgos y que siempre se respetan los derechos fundamentales. En el caso de aplicaciones de la IA de bajo riesgo, la Comisión Europea propone un esquema de etiquetado voluntario.

En segundo lugar, la Estrategia de Datos⁴ aspira a crear un único mercado europeo para datos que asegure la competitividad de Europa y la soberanía de sus datos.

En tercer lugar, un documento con recomendaciones por parte del grupo de alto nivel de “Business to Government Data Sharing”⁵, que incluye la adopción de medidas legislativas sobre la gobernanza de los datos, el acceso y su reutilización para el interés público.

³ https://ec.europa.eu/info/sites/info/files/commission-white-paper-artificial-intelligence-feb2020_en.pdf

⁴ https://ec.europa.eu/info/strategy/priorities-2019-2024/europe-fit-digital-age/european-data-strategy_en

⁵ <https://ec.europa.eu/digital-single-market/en/news/meetings-expert-group-business-government-data-sharing>

En cuanto a los Estados miembros, numerosos países han publicado sus estrategias nacionales de IA acompañadas de ambiciosos compromisos presupuestarios: Alemania ha comprometido 500 millones anuales entre 2019 y 2025, Finlandia 100 millones anuales a partir de 2019 y Francia 1.500 millones en el periodo 2018-2022.

Además, algunos Gobiernos, como el británico, han creado sus propias organizaciones de Inteligencia Artificial (*Office of AI, Office of Ethics in AI*), con carácter transversal, para maximizar el impacto de la IA y acelerar su desarrollo y adopción.

La OCDE probó en mayo de 2019 una Recomendación sobre Inteligencia Artificial [61] a la que se han adherido 36 países miembros y seis países no miembros.

En España he sido miembro portavoz de un comité de nueve expertos multidisciplinares creado por el Gobierno de España en 2017, para la elaboración de un libro blanco sobre el Big Data y la IA con recomendaciones estratégicas. Debido al cambio de Gobierno aún no contamos con una estrategia nacional, si bien se publicó una estrategia de I+D+i en IA en junio de 2019. Desde un punto de vista presupuestario, y tomando como referencia estrategias y de países de nuestro entorno, sería razonable esperar un compromiso para el desarrollo de la IA a escala nacional con una dotación de 125 millones de euros anuales durante los próximos 5-7 años⁶.

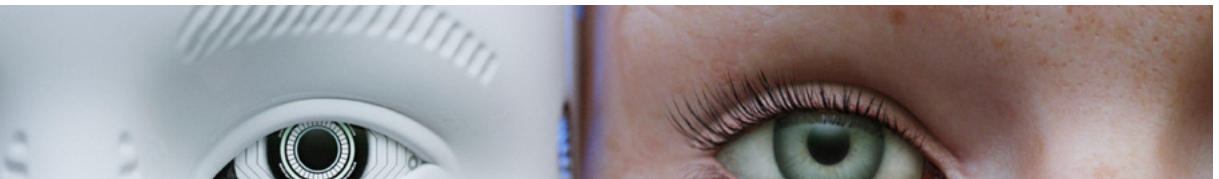
⁶ Dado que el PIB de Francia es aproximadamente el doble del de España y el compromiso francés es de 300 millones de euros anuales.

Curiosamente, dos comunidades autónomas han asumido cierto liderazgo en IA y han publicado sus estrategias: Cataluña y la Comunidad Valenciana⁷.

Sin embargo, como consecuencia de algunas de las características descritas anteriormente, la IA no estará necesariamente distribuida de manera homogénea o justa en la sociedad. Esto significa que la IA plantea retos que debemos afrontar con rigor y ambición, si queremos una Inteligencia Artificial por y para la sociedad. El siguiente capítulo se centra en analizar estos retos, y en proponer soluciones.

⁷ <https://participa.gencat.cat/uploads/decidim/attachment/file/818/Document-Bases-Estrategia-IA-Catalunya.pdf>

http://www.presidencia.gva.es/documents/80279719/169117420/Dossier_cas.pdf/88361b83-0e33-4b49-99c0-ad894ffc0f75



TAREAS PENDIENTES DE LA IA

Los sistemas inteligentes han cosechado un éxito tras otro a lo largo de la última década, y en nuestra vida cotidiana son ya ubicuas –aunque no siempre visibles– las aplicaciones derivadas de la Inteligencia Artificial. Sin embargo, aún hay tareas pendientes:

- **Lenguaje.** - Los sistemas de procesamiento del lenguaje natural aún necesitan incorporar semántica y razonamiento. Es decir, falta conocimiento semántico en los sistemas que procesan e interpretan el lenguaje natural.

- **Incertidumbre.** - Los seres vivos somos capaces de tomar decisiones en entornos de incertidumbre, con falta de información. Pero la incertidumbre es compleja de modelar matemáticamente, lo que da lugar a modelos que no son resolubles computacionalmente.

- **Aprendizaje a corto, medio y largo plazo.** - La gran mayoría de los sistemas de aprendizaje por ordenador son entrenados en una fase inicial, utilizando datos de entrenamiento que sirven para determinar automáticamente los valores de los parámetros de los modelos. Una vez el modelo ha aprendido, se puede aplicar a datos nuevos. Sin embargo, los sistemas inteligentes biológicos aprenden constantemente y con diferentes horizontes temporales –corto, medio y largo plazo–, algo aún no al alcance de los sistemas de aprendizaje computacionales.

- **Causalidad.**- Este concepto es una abstracción para ayudarnos a explicar cómo funciona el mundo, y forma parte del aprendizaje temprano de nuestra y otras especies. Sin embargo, la gran mayoría de los modelos computacionales de aprendizaje estadístico aprenden y detectan correlaciones entre variables y factores, no relaciones de causalidad. La capacidad para inferir automáticamente la causalidad es un área activa de investigación.

- **Contexto.**- Los modelos entrenados con datos aprenden de la información contenida en dichos datos, que no son más que un reflejo parcial de una realidad compleja. Sin información adicional -no captada por los datos-, los modelos pueden llegar a conclusiones erróneas. Por ello es importante entender qué información de contexto puede no estar siendo captada por los datos con que se entrena un determinado modelo.

- **Aprendizaje constante.**- Los seres vivos aprendemos constante, incremental y asociativamente, en lugar de solamente durante la fase de entrenamiento de los modelos.

- **Robustez.**- Los modelos no deben fallar estrepitosamente si se cambian ciertas características en los datos de entrada, como sucede hoy en día. Por ejemplo, hoy podemos engañar a redes neuronales que reconocen objetos en imágenes agregando a la imagen ruido que es imperceptible para los humanos, pero que confunde totalmente a la red neuronal, de manera que en lugar de reconocer que en la foto hay un oso panda, la red neuronal piensa que hay un mono. Este tipo de ataques a los modelos de aprendizaje se conoce como *aprendizaje adversarial* (*adversarial machine learning*).

4

[CAPÍTULO 4]

Los retos

Aprovechar la IA exige conocer sus limitaciones

La Inteligencia Artificial ayudará sin duda a hacer frente a los grandes problemas del siglo XXI. Pero, como todas las herramientas poderosas, su uso requiere sabiduría. ¿Cómo adquirirla? Más aún, ¿cómo lograr que todos la adquiramos? Por paradójico que parezca, ser actor de pleno derecho en la era de las máquinas inteligentes exige potenciar habilidades exclusivamente humanas, como la creatividad. También, entrenar la capacidad de resistir a adictivos cantos de sirena que llegan en forma de notificaciones o *likes*. ¡Ah! Y recuperar esa exótica sensación, hoy casi olvidada, de estar aburridos.

Entramos en el siglo XXI enfrentándonos a no pocos desafíos. Nuestro éxito como especie ha traído consigo una población cada vez más envejecida, el cambio o, mejor dicho, la crisis climática y una acuciante pérdida de biodiversidad a escala planetaria, entre otros grandes retos.

En este contexto, el potencial de la Inteligencia Artificial para ayudarnos a buscar soluciones es inmenso. Sin embargo, los sistemas actuales de IA presentan limitaciones que tendremos que abordar.

Algunos de estos puntos débiles derivan de rasgos distintivos de la Inteligencia Artificial, más allá de los que ya conocemos: transversalidad, invisibilidad, complejidad, escalabilidad, actualización constante y capacidad para predecir. Pero debemos considerar otras características adicionales.

Una es la asimetría. Hoy día los sistemas de Inteligencia Artificial deben ser diseñados y entrenados por personas expertas, y necesitan acceder a grandes cantidades de datos y de computación. Se da, desgraciadamente, una situación de asimetría porque solo una minoría que tiene acceso a datos, capacidad de computación, conocimiento y experiencia puede beneficiarse plenamente de la Inteligencia Artificial. En cambio, la mayoría de la población, incluyendo administraciones públicas, ONGs o pequeñas y medianas empresas, carece de estas capacidades y es, en el mejor de los casos, mera usuaria de la tecnología. No es baladí el detalle de que a la minoría privilegiada pertenecen las grandes empresas tecnológicas que poseen, de hecho, gran parte de los datos.

Minimizar esta asimetría es uno de los grandes desafíos al que nos enfrentamos si queremos garantizar que la Inteligencia Artificial tenga un impacto positivo en toda la sociedad y no solo en una pequeña parte.

Otro gran reto deriva de la capacidad de los sistemas de IA para crear contenidos ficticios que podrían pasar por auténticos. El poder de la IA para generar fotos, textos, audios y vídeos indistinguibles del contenido real –lo que llamamos contenido *sintético o deep fakes*, por estar generados utilizando redes neuronales profundas– está transformando la comunicación, la difusión de la información y la formación de la opinión pública. Controlar la producción y distribución de este contenido artificial equivale a poseer un poder sin precedentes [27].

Además, los sistemas de IA no son invulnerables ante un uso malicioso de los mismos o ante su *hackeo*. De hecho, un área nueva dentro de la IA es la llamada aprendizaje adversarial, donde se entrenan sistemas de IA que engañen a otros sistemas de IA. Cuantas más decisiones de nuestra vida estén parcial o totalmente delegadas a sistemas de IA, dichos sistemas no van a ser *hackeados*.

Si no abordamos estos desafíos, el impacto de la IA no estará necesariamente distribuido de manera homogénea o justa en la sociedad. En este capítulo y en el siguiente describiré brevemente cuatro dimensiones fundamentales que hemos de considerar en el contexto del desarrollo de la IA: su impacto económico; la necesidad de invertir en el desarrollo de capacidades; el valor de la diversidad; y la importancia de definir un marco ético y una nueva gobernanza de los sistemas de IA.

Economía: viento en las velas

Desde un punto de vista económico, en 2020 el mercado de productos, *hardware* y *software*, relacionados con la Inteligencia Artificial se espera que supere los 10.000 millones de dólares, según un estudio de Statista (ver Figura 15). Con un crecimiento sostenido en el tiempo, un análisis macroeconómico de PwC estima que la Inteligencia Artificial generará en el año 2030 más de 15 billones –millones de millones– de dólares a escala mundial. Su efecto se hará sentir en todos los ámbitos de actividad y tanto en el sector público como en el privado.

Aplicada a los procesos de automatización, y como apoyo a la fuerza laboral, la Inteligencia Artificial impulsará la productividad, y de ahí su gran impacto económico. Además, habrá una demanda creciente de productos y servicios enriquecidos con IA por parte de los consumidores.

Norteamérica y China experimentarán los mayores beneficios, dada la concentración de la investigación, la innovación y el desarrollo de la Inteligencia Artificial en estas regiones.

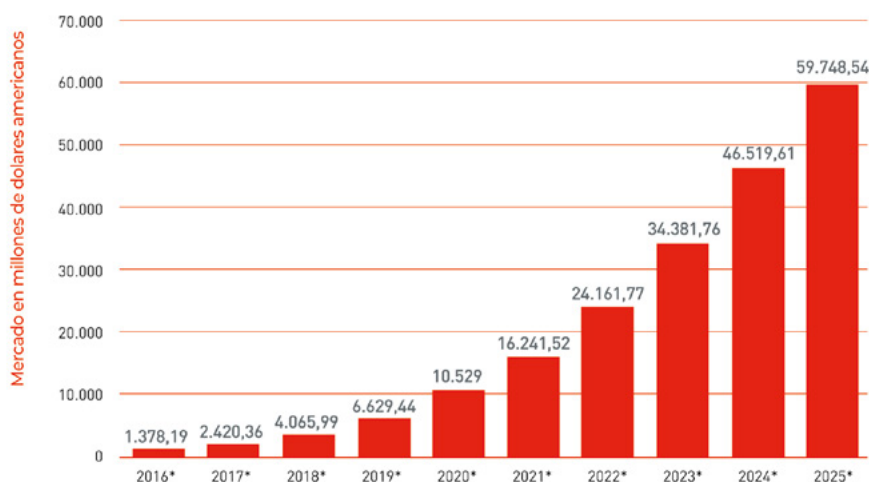


Fig. 15 Estimación del tamaño del Mercado global de la Inteligencia Artificial en millones de dólares americanos (fuente: Statista).

Con respecto a la innovación en IA, un estudio reciente de Asgard y Roland Berger analiza la distribución de jóvenes empresas del sector en todo el planeta. EE.UU. es el líder mundial, con un 40% de todas las empresas emergentes de IA analizadas –alrededor de 1.400 *start-ups*, de las cuales unas 600 están en San Francisco–, seguido de China e Israel. En Europa, Londres es la segunda ciudad del mundo en número de empresas de IA, seguida de París en décima posición.

Otro informe, de Accenture y Frontier Economics, estima el crecimiento económico que diferentes países podrían tener si invirtiesen en IA y consiguiesen incorporar sus beneficios a la economía –en comparación con el crecimiento económico de base, sin la aportación de la Inteligencia Artificial–. El modelo descrito en este informe estima que la Inteligencia Artificial tiene el potencial de duplicar las tasas de crecimiento de los 12 países estudiados, entre los que se encuentran EE. UU., Finlandia, Reino Unido, Alemania, Francia y España.

La Inteligencia Artificial nos permitirá disponer de una medicina de precisión personalizada, preventiva y predictiva; una educación a medida de cada uno, y permanente; ciudades inteligentes; una gestión más eficiente de los recursos; y una toma de decisiones más justa, transparente y basada en la evidencia. Pero las nuevas herramientas vendrán acompañadas de cambios sociales profundos fruto de la Cuarta Revolución Industrial anteriormente descrita, incluyendo una transformación radical del mercado laboral y de las capacidades relevantes para el siglo XXI.

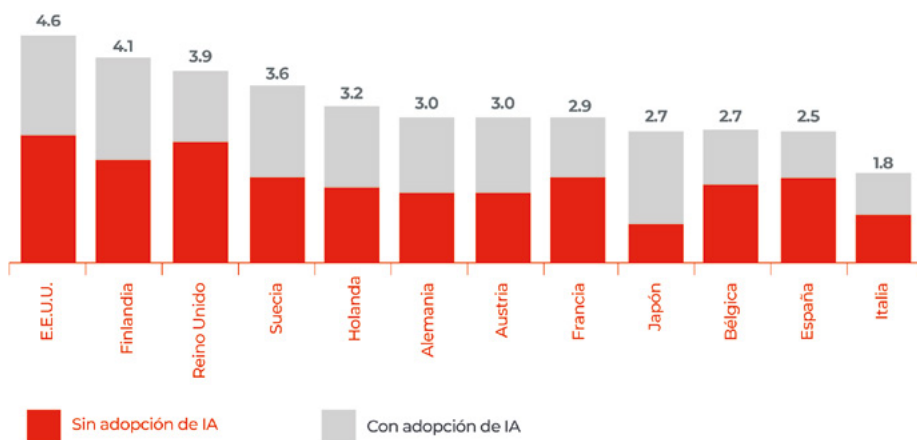


Fig. 16 Estimación del crecimiento del Producto Interior Bruto (PIB) de diferentes economías mundiales (incluyendo la española), en función de su capacidad para adoptar o no la IA (fuente: Accenture y Frontier Economics).



LA ECONOMÍA INTELIGENTE EN ESPAÑA

La economía española podría verse muy beneficiada del uso de sistemas de Inteligencia Artificial en multitud de sectores, entre ellos muchos de valor estructural como la

salud; el transporte; la energía; la agricultura; el turismo; el comercio electrónico; la banca; y la administración pública. Un estudio de Accenture y Frontier Economics [66] prevé que en 2035 el 0,8% del crecimiento del PIB español podrá atribuirse a la IA, un efecto que debería traducirse en más enriquecimiento para la sociedad en su conjunto.

El mercado laboral sufrirá una transformación profunda. En España, la Organización para la Cooperación y el Desarrollo Económico (OCDE) cifra en un 12% los empleos susceptibles de perderse por la automatización. Pero no hay que olvidar que el desarrollo de tecnologías disruptivas con capacidad para transformar la sociedad ha conllevado históricamente la generación de empleo [25].

El informe EPYCE 2018 [67] recoge que el perfil más demandado en España actualmente es el de Científico/a de Datos, seguido de perfiles relacionados con el Big Data e ingeniero/a informático. La tendencia para el futuro, según el mismo estudio, no es muy distinta: el 57% de las profesiones más demandadas en España se enmarcarán en los ámbitos de la ingeniería y la tecnología.

Habría que asegurarse de que los trabajadores/as -hoy en día y en las décadas venideras- cuentan con la preparación necesaria para aprovechar las nuevas oportunidades.

Si lo puede hacer una máquina, lo hará (sola o con nuestra colaboración)

El progreso tecnológico asociado a la Cuarta Revolución Industrial está polarizando el mercado laboral. Mientras aparecen nuevos trabajos bien remunerados y que requieren especialización en áreas tecnológicas, como científico/a de datos, otras profesiones se enfrentan a la automatización parcial o total: taxistas y transportistas, cajeros/as y agentes de viaje, entre muchos otros. Si una tarea puede ser automatizada, lo será, total o parcialmente. Si la tecnología puede aumentar la eficiencia de un proceso, será necesario un menor número de trabajadores humanos para llevarlo a cabo.

En consecuencia, la demanda laboral está experimentando un sesgo en beneficio de habilidades profesionales especializadas, y en perjuicio de ocupaciones rutinarias y mecánicas. Esta tendencia permite pronosticar un cambio completo en la estructura ocupacional, una transformación que probablemente conlleve riesgos para la sociedad si no nos adaptamos.

Según un estudio de McKinsey, un tercio de los nuevos puestos de trabajo creados en EE. UU. en los últimos 25 años pertenecen a disciplinas que no existían anteriormente, en áreas como las tecnologías de la información, la fabricación de hardware, la creación de aplicaciones móviles o la gestión de sistemas tecnológicos.

En la próxima década el empleo se concentrará en funciones cualificadas y con mayor aportación de valor, llegando a tasas de paro inferiores al 3,5% para estos perfiles, frente a un 20% para profesiones de una baja cualificación. Estos datos no implican necesariamente un aumento del desempleo. A escala global, un estudio reciente del Foro Económico Mundial [62] prevé un crecimiento neto de 58 millones de puestos de trabajo en 2022 como efecto de la IA. Es decir, se espera la desaparición de 75 millones de puestos de trabajo, sí, pero también la creación de 133 millones.

La Comisión Europea [63], por su parte, anticipa que hará falta cubrir entre 700.000 y 900.000 nuevos puestos de trabajo tecnológicos a corto plazo.

La clave es que estas nuevas oportunidades laborales serán de naturaleza muy distinta a los puestos que se verán desplazados por la IA. Por ello, ¿estamos preparados, como sociedad, para suplir la demanda de nuevas ocupaciones derivadas de la IA?

Considero que no. Si no transformamos nuestros programas educativos, no lo lograremos. Es de vital importancia que invirtamos en la formación de profesionales cuyo trabajo va a verse afectado por el desarrollo de la Inteligencia Artificial, de manera que puedan seguir contribuyendo a la sociedad.

En continuo aprendizaje

Estamos progresando hacia un modelo de aprendizaje continuo a lo largo de la vida, un modelo en el que cada persona actualiza y diversifica su carrera profesional de manera permanente.

Gracias a la tecnología disponemos de sistemas de formación continua como los llamados *cursos masivos* en línea –MOOC en sus siglas en inglés–, que ofrecen, de manera escalable y económica, oportunidades de aprendizaje a cualquier persona desde cualquier lugar. Deberíamos asegurarnos de que los profesionales pueden aprender tecnologías emergentes en sus ámbitos de actividad, y así continuar cumpliendo una función incluso –y especialmente– si sus áreas de competencia se ven afectadas por la automatización.

Esta necesidad de aprendizaje constante, consecuencia del cambio también incesante causado por el progreso tecnológico, puede ser difícil de gestionar desde un punto de vista emocional. Como bien sabemos, los humanos solemos ser resistentes al cambio, especialmente conforme envejecemos. Deberíamos por tanto contemplar la posibilidad de que algún colectivo sea incapaz de adaptarse a la permanente necesidad de aprendizaje, lo que lo privaría de las herramientas necesarias para contribuir a la sociedad del mañana.

Corremos el riesgo de que este grupo social se convierta, en palabras del historiador Yuval Noah Harari, en una “clase inútil” [28]. ¿Cómo

tratar el problema? Algunas soluciones propuestas son la creación de un salario básico universal o la provisión universal y gratuita de las necesidades básicas por parte de los gobiernos.

Más allá del mercado laboral, el desarrollo e implantación de una Inteligencia Artificial centrada en las personas debería resultar en un empoderamiento de la ciudadanía, de cada uno de nosotros. Una condición necesaria para este empoderamiento es el conocimiento.

La asignatura pendiente: el 'Pensamiento Computacional'

Necesitamos invertir en educación formal e informal. De lo contrario será muy difícil, si no imposible, que como sociedad seamos capaces de tomar decisiones sobre tecnologías que no entendemos, y que en consecuencia frecuentemente tememos. Coincido plenamente con las palabras de Marie Curie, “nada en la vida debería temerse, sino entenderse. Ahora es momento de entender más para así temer menos”. Pero ¿dónde estamos? ¿Qué nivel de conocimiento tecnológico tenemos, tanto niños y jóvenes como adultos?

En el libro *Los nativos digitales no existen* [30] exploro estas preguntas en el capítulo *Erudit@s digitales*, que enfatiza la necesidad de enseñar Pensamiento Computacional en la educación obligatoria, así como de desarrollar el pensamiento crítico, la creatividad y la inteligencia social y emocional. Son habilidades que hoy descuidamos, y sin embargo cada vez van a resultar más importantes para nuestra

salud mental y nuestra coexistencia pacífica y armoniosa con la tecnología, con otros humanos y con nuestro planeta.

El término *alfabeto digital* se refiere a quienes saben utilizar una amplia gama de dispositivos digitales, como teléfonos móviles, ordenadores, tabletas, etc. Desde este punto de vista podríamos considerar que la gran mayoría de nuestros jóvenes, adolescentes y niños son hoy día alfabetos digitales. Según un estudio reciente de Pew Research Center sobre el uso de la tecnología en los adolescentes estadounidenses, un 88% de ellos tienen un teléfono móvil, un 94% tienen acceso a un ordenador y un 24% reconocen estar conectados casi constantemente.

Este nivel de adopción de la tecnología es sin duda una buena noticia, ya que numerosos estudios corroboran el impacto positivo del acceso a la tecnología e internet en el desarrollo de un país. Así pues, observando estas cifras podríamos pensar que las nuevas generaciones están plenamente preparadas para ser competentes en el mundo digital. Después de todo, son nativos digitales. ¿O tal vez no?

Es fundamental no confundir el saber usar una tecnología con saber cómo funciona. Y aunque nuestros hijos vivan enganchados a ella, tanto chicos como chicas, ¿cuántos de ellos saben cómo funciona esa tecnología alrededor de la cual gira su vida?

Parecería que muy pocos. Un estudio de Horizon 2014 en Europa enfatiza los deficientes niveles de competencia digital de los niños y adolescentes europeos. Otro informe reciente, de EU Kids Online⁸, subraya que dos tercios de los niños británicos de entre 9 y 10 años saben sobre internet tanto como sus progenitores. En lo que se refiere al conocimiento de la tecnología en la ciudadanía, también queda todavía mucho camino por recorrer. Una encuesta reciente de la Fundación Española de la Ciencia y la Tecnología (FECYT)⁹ arroja resultados preocupantes: sólo un 16.3% de los españoles sienten interés por la tecnología o la ciencia, un porcentaje que desciende al 13.7% cuando se considera únicamente a las mujeres.

En vista de estos resultados, la Comisión Europea publicó en 2018 un *Plan de acción de la educación digital*, que incluye once acciones para fomentar el uso de la tecnología y el desarrollo de competencias tecnológicas a través de la educación. El Plan señala tres prioridades: hacer un mejor uso de la tecnología digital para la enseñanza y el aprendizaje; desarrollar competencias y habilidades digitales de relevancia para la transformación digital; y mejorar los sistemas educativos a través del análisis de datos y procesos de previsión.

⁸ <http://globalkidsonline.net/eu-kids-online/>

⁹ <https://icono.fecyt.es/principales-indicadores/interes-por-la-ciencia>

Desde un punto de vista formal, la educación obligatoria de muchos países del mundo –entre los que desgraciadamente no se encuentra España– ya incorpora en todas o algunas de sus etapas una asignatura troncal de Pensamiento Computacional. El concepto Pensamiento Computacional [29] hace referencia a los procesos mentales –humanos– que ayudan a formular los problemas de manera que un ordenador pueda operar con ellos y resolverlos. Algo así como aprender a pensar como una máquina para poder utilizarla en la resolución de problemas, y de este modo conseguir que todos podamos beneficiarnos de la capacidad de los ordenadores para buscar soluciones óptimas.

No es una idea nueva. El término fue empleado por primera vez por Seymour Papert en su libro de 1980 titulado *Desafío a la mente: computadoras y educación*. Seymour era en aquel momento co-director con Marvin Minsky del laboratorio de Inteligencia Artificial de MIT y fue pionero del uso de los ordenadores en el aprendizaje de los niños. Creó, entre otros, el lenguaje de programación Logo con fines educativos.

Más recientemente, en 2006, Jeanette Wing –en ese momento en la Universidad de Carnegie Mellon– escribió un artículo [64] para la publicación de la Asociación de Maquinaria de Computación, *Communications of the ACM*, que dio visibilidad a la importancia del Pensamiento Computacional como una habilidad y una actitud valiosa para todas las personas, no solo para los expertos/as en informática. Para Wing, “el Pensamiento Computacional describe la

actividad mental orientada a formular un problema de manera que admita una solución computacional. Esta solución puede ser llevada a cabo por humanos o por máquinas, o por una combinación de humanos y máquinas”.

Como asignatura, el Pensamiento Computacional abarca cinco áreas de conocimiento básicas en un contexto tecnológico, adaptadas a cada nivel educativo: los algoritmos, los datos, las redes, la programación y el hardware. Existen ejemplos de currículos de Pensamiento Computacional de distintos países del mundo, que pueden tomarse como referencia [29,33,34]. Uno de los esfuerzos de mayor envergadura es probablemente el de Reino Unido, que incorpora un currículum de Pensamiento Computacional en todos los colegios a partir de los cinco años. Además, hay programas específicos para atraer a las niñas a las ciencias y la tecnología. Es destacable igualmente la iniciativa lanzada en 2016 en EE. UU. por el entonces presidente Barack Obama, dotada con de 4.000 millones de dólares, para universalizar el estudio de la informática y las competencias digitales en los colegios del país.

Uno de los mayores retos con respecto al éxito en la incorporación del Pensamiento Computacional en la educación obligatoria es la inversión ambiciosa en la necesaria formación al profesorado. En España existe un proyecto de ley educativa que incorpora el Pensamiento Computacional. Ojalá cuando estén leyendo este libro ya esté aprobada dicha ley. Porque los sistemas educativos de otros países ya enseñan, entre otras cosas, a programar y a diseñar algoritmos, y a representar y analizar datos en los ordenadores.

Los niños y niñas de esos países serán conscientes del valor de sus datos personales, entenderán cómo se comunican las máquinas entre sí y cómo funcionan el *World Wide Web*, los buscadores y las redes sociales. A esos niños y a esas niñas se les están brindando oportunidades para desarrollar sus competencias digitales. ¿Y a los nuestros?

Corremos el riesgo de que haya una élite minoritaria –y homogénea– de expertos que saben cómo funciona la tecnología y son capaces de crearla, erigiéndose así en constructores exclusivos de un futuro a su medida. Mientras tanto, una gran masa de gente usará esa tecnología que otros han creado y quedará excluida –excepto como consumidores– de ese futuro tecnológico.

Es importante destacar que la inacción no va a resolver la situación. Deberíamos poder ofrecer a nuestros niños la educación que les permita llegar a ser *erudit@s digitales* (vease el recuadro “*erudit@s digitales*” en la página 90). Y deberíamos también poner en marcha acciones para fomentar vocaciones científico–tecnológicas entre nuestros jóvenes –dada la inmensa demanda anticipada en profesiones tecnológicas–, especialmente entre las chicas, ya que en el ámbito tecnológico hay una preocupante falta de diversidad de género. Como veremos a continuación, la diversidad enriquece, tanto literal como metafóricamente.

El mundo necesita más *erudit@s digitales*, formarlos está en nuestras manos, como explico a continuación.



ERUDIT@S DIGITALES

Una persona *erudita digital* entiende la diferencia entre, por ejemplo, llamar vía *Skype* y por teléfono tradicional; sabe qué son y cómo se usan sus datos personales capturados *online*; conoce el término *Big Data*, y el valor de las cantidades ingentes de datos; está familiarizada con el funcionamiento de internet, una red social o un móvil/ordenador; y sabe programar, entre otras habilidades.

Ser erudito digital comporta poder apoyarse en la tecnología para desarrollar el propio potencial, y contribuir a desarrollar el potencial de la tecnología como herramienta para fomentar la creatividad, resolver problemas, crear oportunidades y, en general, mejorar la calidad de vida.

Ser *erudito digital* implica saber cómo distinguir entre el contenido veraz y el no veraz, poder contrastar contenidos digitales y crear nuevos contenidos propios.

Sin embargo, además de las capacidades técnicas será fundamental desarrollar la creatividad y los aspectos emocionales y sociales de nuestra inteligencia. Serán estas habilidades las que nos ayudarán a sacar el máximo partido de una tecnología que cada vez va a ser más potente –incluso superará nuestras habilidades–, y al mismo tiempo más adictiva.

Por ello, ser erudito digital requiere también desarrollar capacidad de autocontrol y sentido crítico. Son estas habilidades las que sirven de guía a la hora de discernir entre el uso apropiado y el no apropiado de la tecnología, entre el uso productivo, constructivo, y el que no es ni productivo ni constructivo.

Los nativos digitales no nacen, se hacen: hacia una sociedad de erudit@s digitales

Las habilidades y los conocimientos necesarios para los jóvenes de hoy deben ser enseñados, no se aprenden simplemente usando la tecnología. Es una de las conclusiones del Estudio Internacional en *Alfabetización sobre la Información y la Informática*, publicado en 2013 y que analiza el grado de competencia con los ordenadores y la capacidad para gestionar información de 60.000 alumnos de 2ª de la ESO en 21 sistemas educativos en todo el mundo.

Los estudiantes, revela el trabajo, no adquieren las capacidades digitales necesarias si estas no son enseñadas formalmente. En otras palabras, para contribuir realmente a la sociedad del futuro no basta con ser usuario de la tecnología. Ese es obviamente un primer paso para poder orientarse –¿sobrevivir? – en un mundo altamente

automatizado; pero si de verdad queremos que las próximas generaciones contribuyan a este futuro tecnológico tenemos que asegurarnos de que adquieren las capacidades para ser erudit@s digitales.

Para conseguir que nuestros jóvenes participen en el diseño del mundo que viene deberíamos enseñarles cómo funciona la tecnología, y además ayudarles a desarrollar un sentido crítico en su uso. Una cosa es usar y consumir, y otra muy distinta conocer.

Por ello propongo que dediquemos esfuerzos para fomentar una cultura de erudit@s digitales. El concepto de *erudición digital* conlleva dimensiones tanto de conocimiento técnico de la tecnología, como de desarrollo de la creatividad, el pensamiento crítico y de herramientas emocionales y sociales para tomar decisiones, colaborar y contribuir en la sociedad del futuro.

Desde un punto de vista de los conocimientos técnicos, ser erudito digital implica conocer con detalle cómo funciona la tecnología que usamos en nuestro día a día, para poder crear a su vez nuevas herramientas que contribuyan al progreso y nos ayude a afrontar los retos globales. La solución a problemas tan poco triviales como el calentamiento global, la crisis energética, el envejecimiento de la población o la brecha entre ricos y pobres tendrá en muchos casos un fuerte componente tecnológico, y de tecnología que aún no hemos inventado.

Ser erudit@ digital implica dominar el Pensamiento Computacional –como hemos descrito anteriormente, pero no únicamente–. La empatía, la paciencia, la perseverancia, la concentración mantenida en una tarea compleja, la tolerancia, la flexibilidad, la habilidad de gestionar el aburrimiento o de aceptar una gratificación a largo plazo son igualmente cualidades muy valiosas en el contexto actual.

Son, también, cualidades que difícilmente podemos desarrollar y cultivar con experiencias exclusivamente tecnológicas, diseñadas para gratificarnos inmediatamente y con frecuentes interrupciones. Veamos algunas de ellas en más detalle.

Golosinas para el cerebro

La gratificación a largo plazo se asocia con la capacidad de rechazar un premio inmediato pero pequeño, a cambio de conseguir otro mayor, más tarde. En la literatura científica se han encontrado conexiones entre la capacidad para la gratificación a largo plazo y el éxito académico, la salud física y psicológica, y las habilidades sociales. La gratificación a largo plazo está asociada con la paciencia, el control de los impulsos, la fuerza de voluntad y el autocontrol, habilidades que forman parte de la función de autorregulación de las personas.

Según numerosos trabajos [31], la gratificación a largo plazo es una habilidad con impacto positivo en nuestra vida, uno de los elementos que nos permite perseverar en tareas que no dan sus

frutos de inmediato. ¿Pero qué relación tiene la tecnología con la gratificación a largo plazo?

La tecnología de hoy en día, con una cantidad ilimitada de estímulos altamente atractivos para nuestras neuronas, es como una golosina para nuestro cerebro. Además, suele enfocarnos en el momento presente, el ahora, lo que aumenta nuestra susceptibilidad a la gratificación a corto plazo y nos hace más difícil pensar a largo plazo.

Según el profesor emérito de psicología de la Universidad de Stanford, Philip Zimbardo, autor del libro *La paradoja del tiempo: La nueva psicología del tiempo*, la tecnología está impactando nuestra percepción del tiempo y nuestra manera de pensar. “La tecnología crea una obsesión con el tiempo, pero está muy enfocada en el corto plazo, en el momento presente”, ha afirmado Zimbardo.

Desde un punto de vista psicológico lo ideal es encontrar un equilibrio entre tres horizontes temporales. Es importante mirar al futuro, porque eso nos motiva a perseverar en la obtención de objetivos a medio-largo plazo; también lo es mantener una perspectiva positiva del pasado, de manera que cuando reflexionemos sobre nuestra vida y hechos pasados tengamos una sensación placentera; y, finalmente, hay que incorporar una vivencia hedonista del presente, para poder disfrutar del momento, de los amigos y familiares.

Cada vez tenemos una relación más íntima e intensa con la tecnología, mirando constantemente nuestros dispositivos móviles, enviando y recibiendo mensajes, conectándonos a través de redes sociales y siendo interrumpidos por notificaciones cada vez más abundantes.

Según la consultora Nielsen, el 43% del tiempo que dedicamos a interaccionar con tecnología lo dedicamos a actividades de entretenimiento y auto-estimulación. En la misma línea, en 2017 un estudio de Activate cuantificaba en más de 12 horas diarias el tiempo que los adultos estadounidenses dedicaban al consumo de tecnología para el entretenimiento y las relaciones sociales, incluyendo los periodos de multitarea, es decir mientras hacen otras cosas [65].

Esta omnipresencia de la tecnología nos somete a un estado de presente hedonista. Con nuestra atención secuestrada, permanecemos concentrados mayoritariamente –y a veces exclusivamente- en el ahora, lo que dificulta nuestra capacidad para encontrar el necesario equilibrio con las otras dos perspectivas temporales fundamentales en nuestra vida: el medio-largo plazo, y el pasado.

Un estudio del Instituto Pew en 2012 encontró que entre la generación conocida como *millennials* la hiperconectividad podría contribuir a una necesidad de gratificación inmediata, y a una falta de paciencia. El estado de conexión permanente nos proporciona

estímulos de manera casi inmediata, acelerando nuestro sentido del tiempo y fomentando la impaciencia cuando algo tarda más de unos segundos en suceder.

Otro elemento que fomenta nuestro foco en el ahora y en la gratificación inmediata es la incertidumbre sobre si recibiremos o no contenido relevante –un mensaje, un post en nuestro muro de Facebook, un *Like* en la última foto que hemos subido a Instagram, etcétera–. Es decir, el premio –o gratificación– no está asegurado, y ese factor de aleatoriedad, en tanto en cuanto suceda al menos en un 25% de las ocasiones, es más efectivo, a la hora de hacernos volver repetidamente a la tecnología, que si el premio fuese consistente.

En una interesante charla TED, Tom Chatfield, diseñador de videojuegos, explica cómo, después de analizar datos sobre millones de jugadores, los diseñadores de videojuegos descubrieron que la manera más efectiva de mantener a las personas ante la pantalla es concederles un premio –por ejemplo, pasar un nivel, recolectar monedas, etcétera– al menos el 25% de las veces que intentan conseguirlo.

Otro ejemplo de la efectividad de estas gratificaciones aleatorias son las máquinas de jugar en los casinos. Estudios con animales muestran que los premios aleatorios no solo motivan a los animales a realizar una cierta tarea, sino también a hacerla mejor que si reciben premios consistentemente cada vez que la hacen bien.

Por un lado, la tecnología nos hace enfocarnos en el presente y en las gratificaciones inmediatas; por otro, los estudios corroboran el valor y la importancia de la habilidad para aceptar gratificaciones a largo plazo. Es importante ser conscientes de esta tensión, y fomentar actividades que refuercen la gratificación a largo plazo en todos nosotros, pero sobre todo en niños y adolescentes.

El mito de la multitarea

La atención humana es un bien escaso. En ese principio se basa el concepto de la economía de la atención, acuñado por Herbert Simon en 1971 y de nuevo de actualidad, teniendo en cuenta la capacidad de las nuevas tecnologías para confinar nuestra atención en el *ahora*.

Según los analistas Thomas Davenport y John Beck, autores del libro *La economía de la atención*, publicado en 2001, la atención humana “es un estado mental de concentración en un elemento en particular”. Una vez percibimos un objeto o contenido concreto le prestamos atención y, en función de lo percibido, decidimos cómo actuar. Conforme tenemos acceso a más contenidos y aplicaciones, la atención humana se convierte en el factor limitante en el consumo de información.

Además, esta lucha constante por reclamar nuestra atención nos hace caer en un estado de multitarea permanente, en el que nos dedicamos superficialmente a varias cosas a la vez, con cambios

muy rápidos de atención de una cosa a otra. De hecho, casi nos parece imposible hoy en día hacer una única cosa, sobre todo en el contexto tecnológico: vemos la televisión con el móvil en la mano, manteniendo varias conversaciones de texto a la vez o mirando contenidos en internet. La tecnología nos hace sentir como superhéroes, capaces de hacer varias cosas a la vez.

Pero ¿lo somos realmente? Numerosos estudios recientes han investigado el efecto de este estado constante de prestar atención a varios estímulos simultáneos. Y los resultados no son alentadores.

Un estudio de la Universidad de California en Irvine (EE. UU.) con trabajadores en una oficina analizó el impacto de las interrupciones, y encontró que se tardaba al menos 25 minutos en recuperar el grado de productividad previo a la llamada, el email o la notificación correspondiente. Otro estudio llevado a cabo por el Instituto de Psiquiatría de la Universidad de Londres observó que las personas distraídas por la tecnología experimentaban una disminución de su coeficiente intelectual superior al que tendrían si hubiesen consumido marihuana.

Todos estos trabajos apuntan a que la aparente capacidad de hacer varias cosas a la vez, apoyada por la tecnología, es una ilusión: el cerebro procesa la información de manera lineal. Además, el estado de multitarea casi permanente podría tener un impacto negativo en nuestra capacidad para enfocarnos en una única tarea durante largos periodos de tiempo, así como para el control emocional.

Vivimos rodeados de tecnología que compite por atrapar nuestra atención. Todos estos estímulos son muy atractivos para nuestro cerebro y por ello a menudo nos resulta difícil resistirnos. Sin embargo, es fundamental fomentar y cultivar la habilidad de concentrarse y sostener la atención en una única tarea, para conseguir una generación de erudit@s digitales.

No obstante, no siempre encontramos estímulos lo bastante interesantes como para capturar nuestra atención. Y es entonces cuando sentimos aburrimiento. ¿Qué relación tiene la tecnología con el aburrimiento? ¿Cuándo fue la última vez que estuvimos aburridos? Con un teléfono siempre en nuestras manos, conectado y cargado de opciones para el entretenimiento y la comunicación, ¿estamos perdiendo la habilidad de estar aburridos? ¿Qué consecuencias puede tener esta pérdida del aburrimiento en nuestras vidas? Exploremos estas preguntas en la siguiente sección.

El valor del aburrimiento

El aburrimiento es un estado emocional caracterizado por la falta de estímulos y por el deseo de satisfacer esta carencia. Una persona aburrida siente a menudo desinterés por los estímulos a su alcance, y puede resultarle difícil concentrarse en la actividad del momento. El psicólogo de la Universidad de York (Toronto, Canadá) John D. Eastwood enfatiza que una persona aburrida no es simplemente alguien sin nada que hacer, sino alguien que busca activamente estimulación sin encontrarla [37]. De hecho, el ansia de escapar

al aburrimiento puede llegar a ser muy intensa: en un estudio liderado por Timothy Wilson, de la Universidad de Virginia (EE. UU.), los participantes prefirieron administrarse dolorosos electroshocks antes que estar solos sin hacer nada durante unos minutos [38].

Pero el aburrimiento no tiene por qué ser necesariamente negativo. Como dijo Dorothy Parker, poeta y escritora americana, “la cura contra el aburrimiento es la curiosidad”. Andreas Elpidorou, de la Universidad de Louisville (EE. UU.), escribe en un artículo científico que “sin el aburrimiento, estaríamos atrapados en situaciones poco gratificantes y perderíamos la oportunidad de vivir experiencias emocional, cognitiva y socialmente estimulantes. El aburrimiento nos indica que no estamos haciendo algo que nos satisface y nos empuja a cambiar de actividad, buscando una actividad más estimulante” [73].

Otros beneficios del aburrimiento incluyen la oportunidad de iniciar procesos creativos y de autorreflexión [39]. Son varios los estudios que hallan conexiones entre el aburrimiento y la creatividad. Si perdemos la capacidad de estar aburridos, de frenar el procesamiento continuo de estímulos y de dejar que nuestra mente genere ideas, quizás también estaremos perdiendo la creatividad, una de las capacidades que nos identifican como humanos.

Por ello, el tercer elemento que considero importante en una cultura de erudit@s digitales es el aburrimiento.

Al proporcionarnos estimulación constante la tecnología cambia nuestra tolerancia al aburrimiento: con el tiempo nos habituamos a un cierto nivel de exposición a estímulos, y cuando ésta disminuye nos sentimos aburridos [40].

Un estudio del Centro Internacional para los Medios y la Agenda Pública (ICMPA) solicitó a un millar de estudiantes de diez países en cinco continentes que pasaran 24 horas sin ningún acceso a contenidos multimedia, a medios de comunicación ni al móvil –redes sociales, WhatsApp, etcétera–¹⁰. Una vez concluido el periodo de abstinencia digital los estudiantes compartieron sus experiencias por escrito. Sus más de medio millón de palabras revelaron tendencias y elementos comunes en su experiencia.

Así, los estudiantes utilizaron reiteradamente el término *adicción* para describir su relación con las tecnologías de comunicación (“me moría de ganas por usar un teléfono, me sentía como un drogadicto sin droga”). Además, en todos los países una mayoría de estudiantes reconocieron haber fracasado en su intento por estar desconectados durante 24 horas. También admitieron que sus teléfonos móviles ya forman parte de sus cuerpos, y por ello les resulta imposible no tenerlos cerca –de hecho, existe la *nomofobia*, el miedo irracional a no tener el móvil cerca o no saber dónde está–.

¹⁰ <https://theworldunplugged.wordpress.com/>

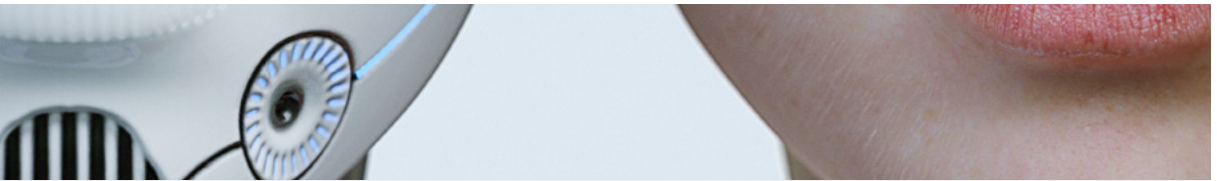
Enfatizaron que estar conectados no es simplemente un hábito, sino un elemento esencial en su capacidad de relacionarse con los demás. La soledad emergió en sus relatos como el sentimiento que aflora al estar desconectados.

Los estudiantes ganaron tiempo al estar desconectados, pero les resultó difícil imaginar cómo ocupar esas horas. El teléfono fue reconocido como elemento que proporciona seguridad y confort. En cuanto al móvil como herramienta para recibir información, en la mayoría de los casos los jóvenes reconocieron leer solo los titulares –140 caracteres de Twitter– de una noticia. Las noticias completas les parece demasiado largas. La televisión fue considerada un medio para relajarse, mientras que la música emergió como elemento de escape primordial y para influir en el estado de ánimo.

Estos resultados coinciden con los de otros trabajos que ven en los teléfonos móviles una herramienta para pasar el tiempo y combatir el aburrimiento [41]. Los móviles se han convertido en nuestros más fieles compañeros; los mantenemos próximos, a nuestro lado, en situaciones de aburrimiento como en los viajes de metro y autobús, en los momentos de espera, etcétera. Recurrimos al teléfono para pasar el tiempo, para autoestimularnos, sin ninguna tarea concreta en mente.

Desde mi punto de vista, esta realidad podría representar una oportunidad: si los móviles fuesen capaces de detectar cuándo estamos aburridos, también podrían sugerir un mejor uso de

esos momentos. Podrían recomendarnos contenidos, servicios o actividades relevantes; sugerirnos prestar atención a tareas pendientes; o ayudarnos a hacer un uso positivo de ese momento de aburrimiento, quizás fomentando la introspección y la creatividad. Exploramos estas ideas en un proyecto realizado en mi equipo de investigación en 2015 con resultados muy prometedores [42] [43]. ¿Por qué no diseñamos tecnología que nos sugiera que la apaguemos?



UNA 'APP' PARA ABURRIDOS

El aburrimiento podría abrir una inesperada ventana de oportunidad para desconectarnos de la tecnología. Junto con mi grupo de investigación desarrollamos en 2015 una aplicación para el móvil, Borapp [43], capaz de determinar si el usuario del teléfono está aburrido o no. El sistema está lejos de ser perfecto, pero representa un primer paso en el diseño de tecnología que nos entiende mejor y tiene el potencial de ayudarnos en la gestión de nuestro tiempo y nuestras emociones.

Borapp podría ser una herramienta que nos ayudase a recuperar un estado emocional que estamos empezando a perder, el aburrimiento, y a aprovechar sus aspectos positivos. Aunque un gran porcentaje de nosotros –y aún mayor de adolescentes– no podamos vivir sin nuestros móviles y estemos constantemente conectados, ¿tiene valor el tiempo que pasamos desconectados? ¿Qué pasaría si nuestro móvil nos sugiriese que lo apagásemos? ¿O será el aburrimiento una reliquia del pasado? ¿Qué pasaría con nuestra creatividad en ese caso?

La importancia del tiempo *off*

Conforme desarrollamos una relación más sinérgica e íntima con la tecnología –una tecnología que a su vez es cada vez más potente, inteligente y conectada–, se convierte en crítico el que seamos capaces de mantener un número mínimo de horas al día de descanso tecnológico, de tiempo *off*.

Jennifer Falbe, del departamento de Ecología Humana de la Universidad de California, abordó en un estudio publicado en 2015 la relación entre uso de pantallas y sueño¹¹, en más de 2.000 niños y niñas de cuarto y primero de la ESO en Massachusetts, EE. UU. Los resultados apuntan a la conveniencia de poner restricciones a

¹¹ Falbe J, Davison KK, Franckle RL, Gehre C, Gortmaker SL, Smith L, Land T, Taveras EM. Screens in children's sleep environments, sleep duration, and perceived insufficient rest. *Pediatrics*. 2015; 135(2):e367-75.

la tecnología –tabletas, teléfonos, televisión– en los dormitorios de los niños y adolescentes. Los que dormían cerca de una pantalla pequeña declararon dormir 20 minutos menos al día que los que no tenían pantallas; y en el grupo con pantallas también eran más los que sentían no haber dormido o descansado lo bastante.

Las actividades que se realizan con estos dispositivos suelen ser muy estimulantes, lo que dificulta la conciliación de un sueño que, por añadidura, puede verse interrumpido por notificaciones audibles durante la noche. Además, la luz de las pantallas brillantes envía al cuerpo la señal de que todavía es de día, lo que inhibe la producción de una hormona implicada en la regulación del sueño, la melatonina. Por ello se recomienda tener unos minutos de descanso con la luz apagada antes de ir a dormir.

Más allá de la necesidad de tener suficientes horas de descanso de calidad, varios estudios corroboran la importancia de saber desconectar y realizar actividades no tecnológicas.

Una actividad que es importante mantener son las relaciones humanas cara a cara, sin tecnología. A fin de cuentas, el *Homo sapiens* es una especie social. La profesora del MIT Sherry Turkle lleva más de treinta años investigando la relación subjetiva entre las personas y la tecnología. En su libro, “En Defensa de la conversación” [44], analiza el impacto que está teniendo la tecnología en nuestra capacidad para

conversar cara a cara, y postula que la disminución –incluso en algunos casos desaparición– de las conversaciones en nuestras vidas representa una grave amenaza para nuestras relaciones, nuestra creatividad y también nuestra productividad.

Mantener un equilibrio entre la comunicación cara y cara y la comunicación mediada por la tecnología va a ser crítico, sobre todo la tecnología que utilizamos para mantenernos conectados tenga las limitaciones actuales, forzándonos en muchos casos a comunicarnos usando únicamente texto y por tanto perdiendo la riqueza de lenguaje no verbal que caracteriza la comunicación humana.

También será necesario equilibrar nuestras interacciones con sistemas conversacionales inteligentes –asistentes personales como *Alexa* o *Google Home*– y nuestras interacciones con humanos. Estos sistemas carecen hoy en día de habilidades de las inteligencias social y emocional, y en muchos casos son incapaces de interpretar correctamente el lenguaje no verbal, tan importante y característico de la comunicación humana. Y ya sabemos que, en la naturaleza, lo que no se usa, se pierde...

Más allá de la comunicación y las relaciones, mantener una presencia y conexión físicas con el mundo que nos rodea es fundamental para nuestra salud mental, nuestro bienestar emocional, nuestra creatividad y, en último grado, nuestra felicidad.



[CAPÍTULO 5]

Hacia una IA por y para todos

La gobernanza del planeta digital

Tres de los países más poblados del mundo hoy en día tienen menos de dos décadas de historia y no aparecen en ningún atlas geográfico. Pero Facebook, WhatsApp e Instagram son especiales no solo por ser digitales y globales, sino porque su dirigente –a quien nadie ha elegido democráticamente- pueden acceder y analizar grandes cantidades de datos procedentes de miles de millones de ciudadanos y ciudadanas del mundo. Explotar estos datos le permite inferir información sobre nuestros hábitos, necesidades, intereses, relaciones, orientación sexual y política, grado de felicidad, de educación, de salud... Esta es una de las actividades que más riqueza generan en la actualidad, pero muy pocos pueden llevarla a cabo. Solo si acordamos principios de gobernanza y actuación empresarial centrados en los derechos y el bienestar de las personas lograremos un modelo de convivencia basado en una Inteligencia Artificial creada por y para todos.

Ya hemos hablado de la Cuarta Revolución Industrial y del papel central de la Inteligencia Artificial en nuestras vidas. En este

capítulo ponemos el foco en las implicaciones sociales, con el objetivo de identificar las oportunidades y retos a abordar para conseguir el empoderamiento de la sociedad gracias a la Inteligencia Artificial.

Ya hemos abordado en el capítulo anterior la brecha de conocimientos digitales y sabemos que sin conocimiento no hay empoderamiento. En este capítulo hablaremos sobre la falta de diversidad –no podemos excluir a la mitad de la población–, de la importancia de involucrar a la ciudadanía, de la discriminación algorítmica, de la transparencia y de la protección de la privacidad.

Además de la educación a niños y adolescentes, deberíamos invertir y trabajar para que la Inteligencia Artificial sea inclusiva y no deje a nadie atrás. Desgraciadamente, esta no es la situación actual, donde además tenemos un porcentaje bajísimo de mujeres: tan solo entre un 10% y un 20% de mujeres trabajan en Inteligencia Artificial y contribuyen a su investigación e incorporación en la sociedad. Tampoco estamos invirtiendo lo suficiente en formación a los profesionales –sobre todo aquellos cuyos trabajos van a verse afectados por la IA— y a la sociedad en su conjunto, para que sea participe de esta Cuarta Revolución Industrial. Necesitamos invertir en que la Inteligencia Artificial sea por y para todos, no solo para unos pocos.

La diversidad enriquece (literalmente)

El valor de la diversidad en la sociedad ha sido corroborado por numerosos estudios. La diversidad enriquece, metafórica y literalmente. Cuanto más diversos son los equipos y las instituciones –públicas y privadas–, mejores son sus resultados económicos y más innovadoras e inclusivas son sus ideas y soluciones.

Un estudio del Centro Nacional para las Mujeres y la Tecnología de la Información [68], realizado con 2.360 comunidades en diversas industrias, revela que las empresas con mujeres en sus comités de dirección obtienen más beneficios que aquellas con dirección solo masculina. Cuando los equipos de gestión son diversos aumentan los retornos de la inversión y el crecimiento. La presencia de mujeres mejora la productividad y la dinámica de los equipos, además de elevar la estabilidad financiera cuando los puestos que ellas ocupan son de responsabilidad [69] [70].

Se puede argumentar que gestionar la diversidad es mucho más complejo que la homogeneidad. Sin duda. Los seres humanos tendemos a la *homofilia*, es decir, a sentirnos mucho más cómodos rodeados de personas similares a nosotros mismos. En los equipos homogéneos no suele haber voces disonantes, porque sus miembros hablan el mismo idioma y comparten una visión similar.

En los equipos diversos, por el contrario, se cuestionan a menudo las decisiones y hay diferentes puntos de vista, voces divergentes

y personas con experiencias vitales distintas. Pero sabemos que, si conseguimos engarzar todos estos elementos, los resultados de los equipos diversos serán superiores.

Fomentar la diversidad, por tanto, no es solo un imperativo de justicia social, sino de crecimiento económico.

Sin embargo, la diversidad –incluyendo la de género– brilla por su ausencia en el sector tecnológico y en las carreras de informática o en las ingenierías aledañas. Estamos hablando de entre un 10%–20% de chicas en las facultades de estas carreras, porcentajes muy inferiores a los que alcanzamos en los años 80. ¿A qué se debe este retroceso?

Se estima que la falta de diversidad de género en el sector tecnológico le cuesta nada menos que 16.200 millones de euros anuales al PIB europeo, según un estudio de la Comisión Europea [71].

Dada la escasa cultura científico-tecnológica de la ciudadanía –reflejada en las encuestas–, considero que deberíamos hacer más divulgación para desmitificar y romper estereotipos relativos a quién trabaja en tecnología y en qué consisten estos trabajos, así como para dar visibilidad y crear referentes femeninos en el sector tecnológico. Podría ser efectivo lanzar campañas en medios de comunicación y redes sociales, con contenidos que realcen el impacto de la tecnología en nuestras vidas; muestren el ejemplo de ingenieras, investigadoras, programadoras, emprendedoras o

inventoras de tecnología que muchas veces están en la sombra; eduquen en conceptos básicos; e inspiren a los jóvenes –sobre todo a las chicas– a estudiar carreras tecnológicas.

Para fomentar la presencia de mujeres en tecnologías de la información y la comunicación han surgido decenas de iniciativas de apoyo, capacitación y sensibilización a escala nacional e internacional, aunque quizás sea necesario aunar esfuerzos para multiplicar su impacto y visibilidad. Otro paso crucial sería introducir la asignatura de Pensamiento Computacional (ver capítulo anterior) como troncal, ya que todos los estudiantes –chicos y chicas por igual– adquirirían con naturalidad competencias tecnológicas, derribando así el estereotipo de género asociado a la tecnología.

Más allá de la diversidad de género, la diversidad de área de conocimiento –la multidisciplinariedad– es clave en el contexto de la Inteligencia Artificial. La razón está en la transversalidad de la tecnología. La IA puede aplicarse a un sinnúmero de áreas de conocimiento y problemas, y con frecuencia los sistemas de IA interactúan o tienen consecuencias para la vida de miles o millones de personas. Por ello es necesario crear equipos multidisciplinarios con expertos y expertas tanto en disciplinas técnicas –como ingeniería e informática– como en las ciencias sociales –antropología, psicología, ética o sociología –, y en las áreas específicas de aplicación de la IA –medicina, educación, transporte, políticas públicas, etcétera–.

Los laboratorios urbanos

Otra dimensión para el empoderamiento social a través de la Inteligencia Artificial conlleva la generación de espacios de colaboración. Se trata de crear laboratorios urbanos –*living labs*, en la jerga– basados en la compartición de datos, tecnología y talento, para contribuir al progreso de la Inteligencia Artificial y democratizar su acceso, equilibrando las desigualdades y la asimetría.

Algunos ejemplos son el Mobile Territorial Lab [46], en Trento (Italia), con quienes colaboré cuando era Directora Científica en Telefónica, y el Laboratorio Urbano de Bogotá (Colombia), impulsado por la ONG Data–Pop Alliance, donde soy investigadora jefa de datos. En el Mobile Territorial Lab lideré un proyecto para entender, desde un punto de vista centrado en las personas, el valor monetario que los usuarios asignan a los datos captados por sus móviles. Este proyecto recibió el premio al mejor artículo científico en el congreso internacional ACM Ubicomp 2014 [47], y fue acogido con gran interés por parte de medios de comunicación nacionales e internacionales al ser la primera investigación que exploraba esta cuestión.

Big Data e IA para el bien social

A pesar de las dificultades, en la última década han surgido iniciativas en todo el mundo para promover el Big Data y la Inteligencia Artificial para el bien común, una de mis áreas

de investigación desde hace precisamente diez años, y así contribuir a la democratización en el acceso y uso de los datos. Estas son algunas de las iniciativas más establecidas a este respecto, aunque desde los últimos dos o tres años surgen nuevas con frecuencia:

- *New Deal on Data*, liderado por el profesor Sandy Pentland – mi director de tesis– desde el Foro Económico Mundial, enfocado a consensuar políticas e iniciativas para que la ciudadanía tenga control sobre la posesión, el uso y la distribución de sus datos personales.
- *Flowminder*, una ONG basada en Suecia, con quienes colaboro desde hace años, tiene una década de experiencia en proyectos de análisis de Big Data para el bien social, con casos de éxito en el uso de datos para, por ejemplo, cuantificar los desplazamientos y asentamientos humanos tras los terremotos en Haití y Nepal.
- *Naciones Unidas: el World Data Forum y el Global Partnership for Sustainable Development Data*. Esta alianza liderada por Naciones Unidas busca la consecución de los 17 Objetivos de Desarrollo Sostenible (ODS) a través del análisis de datos. Entre sus más de 150 colaboradores hay un amplio espectro de productores y usuarios de datos, incluidos gobiernos, empresas, universidades, ONGs, grupos de la sociedad civil, fundaciones, etcétera.

Los datos y la Inteligencia Artificial pueden utilizarse tanto para contribuir a conseguir los ODS como para analizar si en efecto los objetivos se están cumpliendo, así como para mejorar la toma de decisiones en las políticas públicas. La colaboración entre los organismos implicados es clave, pero exige el compromiso de todos ellos a cooperar y a involucrar a los ciudadanos.

Precisamente para abordar las oportunidades y retos del uso de los datos para la consecución de los 17 ODS, Naciones Unidas organiza el *World Data Forum* desde 2017. El Foro de 2018 concluyó con el lanzamiento de la Declaración de Dubai [72] para fomentar la inversión en el uso de los datos para el crecimiento sostenible.

En Naciones Unidas también existe, desde 2009, una unidad dedicada al análisis de datos usando técnicas de Inteligencia Artificial para el bien social, llamada *United Nations Global Pulse*, con quienes he colaborado y a cuyo comité asesor pertenezco. En 2014 demostramos que los datos agregados y anonimizados de la red de telefonía móvil, combinados con observaciones vía satélite, pueden contribuir a detectar zonas afectadas por inundaciones en México [45].

- **OPAL** [59] es un proyecto liderado por Data–Pop Alliance que aprovecha el Big Data y la Inteligencia Artificial para el bien social, preservando la privacidad de las personas de manera sostenible, escalable, estable y comercialmente viable.

OPAL, que se realiza en colaboración con entidades públicas y privadas, propone democratizar el acceso a los datos, y al conocimiento que proporcionan, haciendo que los algoritmos sean abiertos y se ejecuten donde estén los datos, y no al revés.

- *Partnership on AI* es una organización sin ánimo de lucro, creada en 2016 por Microsoft, Facebook, Amazon, IBM, Google y Apple, que agrupa a más de cincuenta organizaciones públicas y privadas. Su fin es estudiar y establecer mejores prácticas respecto a la Inteligencia Artificial, y promover el conocimiento y el debate acerca de la IA por parte de la ciudadanía¹².

- *GSMA Big Mobile Data for Social Good* es una iniciativa liderada por la asociación de operadores de telefonía móvil GSMA y Naciones Unidas, en la que participan 20 operadoras de móvil. Su fin es aplicar el análisis de datos agregados y anonimizados de la red de telefonía móvil a problemas en áreas de salud pública, cambio climático y desastres naturales¹³.

- *AI for Good Global Summit of the ITU*, la cumbre internacional de Naciones Unidas para el diálogo sobre la Inteligencia Artificial, busca aplicar la IA a la mejora de la sostenibilidad del planeta. Está gestionada por la ITU (Unión Internacional de Telecomunicaciones), el organismo de Naciones Unidas para las Tecnologías de la Información y la Comunicación¹⁴.

- El centro multidisciplinar *Center for Humane Technology*¹⁵, creado recientemente en California, defiende el desarrollo de tecnología que tiene en cuenta los valores, las necesidades y los intereses de las personas por encima de todo.
- El *Grupo de Alto Nivel de Compartición de Datos de las Empresas a los Gobiernos*, constituido en 2018 dentro de la Comisión Europea [60] y del que soy miembro, aspira a identificar buenas prácticas en la compartición de datos. En febrero de 2020, este grupo de alto nivel publicó su informe que incluye un conjunto de recomendaciones a la Comisión Europea para fomentar la compartición de datos privados con el sector público.

¹² <https://www.partnershiponai.org/>

¹³ <https://www.gsma.com/betterfuture/partnering-for-a-better-future>.

¹⁴ <https://aiforgood.itu.int/>

¹⁵ <https://humanetech.com/>

Además, numerosas empresas privadas han fomentado y desarrollado proyectos de Big Data e Inteligencia Artificial con fines de bien social. Algunos son muy cercanos a mí, como *LUCA–Big Data for Social Good*, de Telefónica y Vodafone Big Data e Inteligencia Artificial para el Bien Social, que desarrolla proyectos de salud pública, inclusión financiera, transporte y estadísticas oficiales en África y Europa. Otros ejemplos incluyen la iniciativa *BBVA Data & Analytics para el Bien Social*, por parte de BBVA; *Telenor Big Data for Social Good*, con proyectos de salud pública en Bangladesh y Pakistán; *Orange Data for Development*, dos retos pioneros donde Orange compartió datos agregados y anonimizados de Senegal y Costa de Marfil con cientos de equipos internacionales en casos de uso para el bien social; y *Turkcell Data for Refugees Challenge*, en el que la operadora Turkcell compartió datos agregados y anonimizados para contribuir a resolver la crisis de los refugiados. Las empresas tecnológicas también se han sumado a este movimiento y han creado recientemente iniciativas del uso de sus datos y la *Inteligencia Artificial para el Bien Social*, como las iniciativas de Facebook [56], Google [57] y Microsoft [58].

Cuando quien decide es un algoritmo

El desarrollo de tecnologías disruptivas ha tenido un profundo impacto en la vida y en las relaciones humanas. La agricultura, la imprenta, la máquina de vapor, la electricidad o internet han transformado nuestra manera vivir, trabajar y relacionarnos. La Inteligencia Artificial también lo está haciendo.

Hoy podemos usar cantidades masivas de datos para entrenar algoritmos de Inteligencia Artificial, y hacer que decisiones que antes eran tomadas por humanos –con frecuencia expertos– recaigan sobre estos algoritmos. Pueden ser decisiones que afectan a una o muchas personas, y sobre cuestiones nada triviales, como la contratación laboral, la concesión de créditos y préstamos, sentencias judiciales, tratamientos y diagnósticos médicos o la compraventa de acciones en bolsa.

Las decisiones algorítmicas basadas en datos pueden ser un gran avance. La historia ha demostrado que las decisiones humanas no son perfectas, pues pueden estar sujetas a conflictos de interés, a la corrupción, al egoísmo y a sesgos cognitivos, lo que resulta en procesos y resultados injustos y/o ineficientes. Por tanto, el interés hacia el uso de algoritmos puede interpretarse como el resultado de una demanda de mayor objetividad en la toma de decisiones.

Sin embargo, decidir basándose en algoritmos no es tampoco un proceso perfecto. Las palabras de Platón hace 2.400 años están sorprendentemente vigentes hoy en día: “Una buena decisión está basada en conocimiento, no en números (datos)”.

Por ejemplo, cuando las decisiones afectan a miles o millones de personas surgen dilemas éticos importantes: ¿escaparán las decisiones automáticas a nuestro control? ¿Hasta qué punto están estos sistemas protegidos contra ciberataques y usos maliciosos? ¿Podemos garantizar que sus decisiones y actuaciones no tienen

consecuencias negativas para las personas? ¿Quién es responsable de dichas decisiones?

Puede resultar especialmente escalofriante una cuestión en particular: ¿Qué sucederá cuando un algoritmo nos conozca a cada uno de nosotros mejor que nosotros mismos, y aproveche esa ventaja para manipular de manera subliminal nuestro comportamiento?

En los últimos cinco años han surgido numerosos movimientos y organizaciones, nacionales o supranacionales –por ejemplo, en la Unión Europea–, para ayudar a establecer estándares globales y a regular este ámbito. En el congreso organizado por el *Future of Life Institute* en 2017, con la participación de más de 1.200 figuras internacionales relacionadas con la innovación tecnológica y científica, se definieron los *Principios de Asilomar* para el desarrollo de la Inteligencia Artificial, con un total de 23 recomendaciones.

En 2018 la Comisión Europea nombró un Comité de Expertos de Alto Nivel en Inteligencia Artificial, que pone el foco en las implicaciones éticas, legales y sociales de la IA, y del que soy miembro reserva.

Este grupo publicó en 2019 unas guías éticas para el desarrollo de una Inteligencia Artificial confiable, lo cual implica una Inteligencia Artificial que respete tanto las leyes como los principios éticos de la sociedad donde sea implantada y que ofrezca garantías de robustez [55].

En el plano nacional destaca la declaración de Barcelona, impulsada por el Centro de Investigación en Inteligencia Artificial del CSIC, en la que se definen seis principios básicos para un desarrollo ético de la IA.

Más allá del respeto a los derechos humanos fundamentales, en la literatura técnica se han propuesto principios éticos y dimensiones de trabajo que considero necesario abordar. El lector interesado puede encontrar una versión extendida de algunos de estos principios en [48]. Resumo aquí los principios más relevantes, agrupados en cinco pilares [49] que en inglés quedan agrupados por el acrónimo *FATEN* [54].

Justicia algorítmica

La F del acrónimo *FATEN* se refiere a *Fairness*, justicia en inglés. O más bien *justicia y solidaridad; no discriminación*. La justicia debería ser un elemento central en el desarrollo de sistemas de decisión –y actuación– automáticos fruto de la Inteligencia Artificial. Las decisiones basadas en estos sistemas pueden discriminar porque los datos utilizados para entrenar los algoritmos tengan sesgos; por la aplicación de un determinado algoritmo; o por el mal uso de ciertos modelos en diferentes contextos.

En los últimos cuatro años han sido publicados ejemplos de discriminación algorítmica en el contexto de las decisiones judiciales, la medicina, las tarjetas de crédito, los sistemas de reconocimiento

facial o los sistemas de contratación. Los colectivos perjudicados han sido las personas afroamericanas o de piel oscura, las mujeres y en general las minorías.

Además, los procesos de decisión algorítmicos basados en datos pueden implicar que se denieguen oportunidades a personas no por sus propias acciones, sino por las de otros con quienes comparten ciertas características. Por ejemplo, algunas compañías de tarjetas de crédito han reducido el crédito de sus clientes no por su comportamiento, sino tras analizar el de otras personas con un historial de pagos deficiente que habían comprado en los mismos establecimientos. En la literatura ya se han propuesto diferentes soluciones para afrontar la discriminación algorítmica y maximizar la justicia.

Sin embargo, me gustaría subrayar la urgencia de que expertos y expertas de distintos campos –incluyendo el derecho, la economía, la ética, la informática, la filosofía y las ciencias políticas– inventen, evalúen y validen en el mundo real diferentes métricas de justicia algorítmica para diferentes tareas. Además de esta investigación empírica, es necesario proponer un marco de modelado teórico –avalado por la evidencia empírica— que ayude a los usuarios de los algoritmos a asegurarse de que las decisiones tomadas son lo más justas posible.

En este concepto también me gustaría incluir el de *cooperación*. Debido a la transversalidad de la Inteligencia Artificial deberíamos

fomentar y desarrollar un intercambio constructivo de recursos y conocimientos entre los sectores privado, público y la sociedad en general, para conseguir el máximo potencial de aplicación y competitividad. Esta necesidad de cooperación entre diferentes sectores, y también entre naciones –dada la globalización–, ha sido enfatizada por Yuval Noah Harari [28].

Autonomía, responsabilidad e inteligencia aumentada

La A de FATEN se desdobra en tres: *Autonomía*, *Atribución de responsabilidad* y *Aumento de inteligencia*. La *autonomía* es un valor central en la ética occidental según la cual cada persona debería poder decidir sobre sus propios pensamientos y acciones.

Sin embargo, hoy en día podemos construir –como he hecho en mis propios proyectos de investigación– modelos computacionales de nuestros deseos, necesidades, personalidad y comportamiento, que tienen la capacidad de influir subliminalmente en nuestras decisiones y acciones.

Por ello deberíamos garantizar que los sistemas inteligentes autónomos modulan su toma de decisiones, modulan su toma de decisiones teniendo en cuenta la autonomía y la dignidad humanas. Para esto debemos disponer de reglas que definan el comportamiento de estos sistemas de acuerdo con los principios éticos aceptados en la sociedad. Hay numerosos ejemplos de

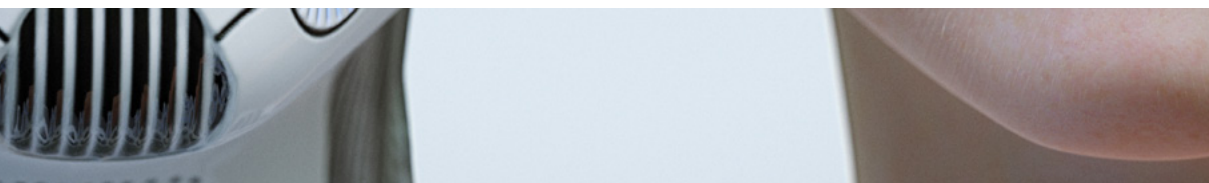
principios éticos propuestos en la literatura para este propósito, así como institutos y centros de investigación creados con este fin, como el AI Now Institute, en la Universidad de Nueva York (EE. UU.); el Digital Ethics Lab, en la Universidad de Oxford (Reino Unido); y la Oficina de Ética en la Inteligencia Artificial, creada recientemente en el Reino Unido como órgano consultivo del Gobierno británico.

Pero esta es un área activa de investigación, y no hay un método único para incorporar principios éticos a los procesos de decisión basados en datos. Es importante que los desarrolladores de sistemas de Inteligencia Artificial que afecten o interaccionen con personas –algoritmos de toma de decisiones, de recomendación y personalización, *chatbots*...– se comporten de acuerdo con un Código de Conducta y de Ética definido por las organizaciones en que trabajan. Como sabiamente dijo Disney, “no es difícil tomar decisiones cuando tienes claro cuáles son tus valores”.

También es crucial que haya claridad en la atribución de responsabilidad de las consecuencias de acciones o decisiones de sistemas autónomos. La transparencia suele considerarse un factor fundamental en este punto; sin embargo, la transparencia y las auditorías son necesarias, pero no son suficientes.

Igualmente, creo constructivo buscar la sinergia entre la Inteligencia Artificial y el ser humano. Esta visión suele llamarse *aumento de la inteligencia* –*intelligence augmentation*–, porque los sistemas de Inteligencia Artificial aumentan o

complementan la inteligencia humana. Por ejemplo, un buscador de internet puede considerarse un sistema de aumento de nuestra inteligencia, ya que expande nuestro conocimiento con la capacidad de procesar miles de millones de documentos y encontrar los más relevantes; o un sistema de traducción simultánea automática, ya que permite comunicarse a personas que no hablan el mismo idioma.



UNA INTELIGENCIA ARTIFICIAL FIABLE

La Estrategia para la Inteligencia Artificial de la Comisión Europea aspira a aumentar las inversiones públicas y privadas en Europa hasta un mínimo de 20.000 millones de euros anuales en los próximos diez años, así como “facilitar el acceso a más datos, fomentar el talento y garantizar la confianza”. Para lograr esto último, la Comisión ha identificado siete requisitos, relacionados con los principios FATEN descritos en el texto:

- **Intervención y supervisión humanas.** Los sistemas de Inteligencia Artificial deben ayudar a construir

sociedades equitativas, apoyando la intervención humana y los derechos fundamentales, en lugar de limitar la autonomía humana.

- **Robustez y seguridad.** Los algoritmos deben ser suficientemente seguros, fiables y sólidos como para resolver errores o incoherencias durante todas las fases del ciclo de vida útil de los sistemas de Inteligencia Artificial.
- **Privacidad y gestión de datos.** Los ciudadanos deben tener pleno control sobre sus datos. Los datos que les conciernen no deben utilizarse para perjudicarles o discriminarles.
- **Transparencia.** Debe garantizarse la trazabilidad de los sistemas de Inteligencia Artificial.
- **Diversidad, no discriminación y equidad.** Los sistemas de IA deben tener en cuenta el conjunto de capacidades, competencias y necesidades humanas, y garantizar la accesibilidad.
- **Bienestar social y medioambiental.** Los sistemas de IA deben impulsar el cambio social positivo y aumentar la sostenibilidad y la responsabilidad ecológicas.

- **Rendición de cuentas.** Deben implantarse mecanismos que garanticen la responsabilidad y la rendición de cuentas de los sistemas de IA y de sus resultados.

Confianza y transparencia, por favor

La *T* de nuestro acrónimo FATEN es doble, por *Confianza* –*Trust*, en inglés– y *Transparencia*.

La confianza es un pilar básico en las relaciones entre humanos e instituciones. La tecnología necesita de un entorno de confianza con sus usuarios, que cada vez más delegan –delegamos– aspectos de sus vidas en servicios digitales. Sin embargo, el sector tecnológico está experimentando una pérdida de confianza por parte de la sociedad, un fenómeno al que han contribuido escándalos recientes como el de Facebook con Cambridge Analytica.

Para que exista confianza han de cumplirse tres condiciones:

(1) la **competencia**, es decir, la habilidad para realizar con solvencia la tarea comprometida; (2) la **fiabilidad**, es decir, la competencia sostenida en el tiempo; y (3) la **honestidad y transparencia**. Por ello, la *T* también es de *Transparencia*.

La transparencia en este contexto hace referencia a la calidad de poder entender un modelo computacional. Un modelo es transparente si una persona no experta puede entenderlo con

facilidad. La transparencia, por tanto, podría contribuir a la atribución de responsabilidad de las consecuencias del uso de dicho modelo.

La socióloga Jenna Burrell, co-directora del Grupo de Trabajo sobre Justicia Algorítmica y Opacidad de la Universidad de California en Berkeley (EE. UU.) [50], propone tres tipos distintos de opacidad – falta de transparencia– en las decisiones algorítmicas. La primera es la opacidad intencionada, cuyo objetivo es la protección de la propiedad intelectual de los inventores de los algoritmos. Esta opacidad podría mitigarse con legislación que obligara al uso de software abierto, como la nueva Regulación General Europea de Protección de Datos (RGPD). Sin embargo, intereses comerciales y gubernamentales poderosos pueden dificultar la eliminación de este tipo de opacidad.

Una segunda forma de opacidad es la de conocimiento. Se refiere al hecho de que solo una minoría de personas está capacitada técnicamente para entender los algoritmos y modelos computacionales construidos a partir de los datos. Esta falta de transparencia se vería atenuada con programas educativos en competencias digitales –como he explicado anteriormente–, y también permitiendo que expertos independientes aconsejen a quienes se hayan visto afectados por procesos de decisión algorítmicos basados en datos.

La tercera opacidad es intrínseca, y surge por la naturaleza misma de ciertos métodos de aprendizaje por ordenador, como los

modelos de *deep learning* o aprendizaje profundo que hemos descrito en el capítulo 3. La comunidad científica de aprendizaje computacional está muy familiarizada con esta opacidad, a la que denomina problema de la *interpretabilidad*, siendo un área activa de investigación. Consiste en desarrollar modelos que sean *explicables*, es decir, que los humanos podamos entender cómo funcionan.

Por último, es imprescindible que los sistemas de Inteligencia Artificial sean transparentes respecto a qué datos sobre comportamiento humano captan, y con qué finalidad –esto queda contemplado en la RGPD europea–. También, por supuesto, deben ser claramente identificables las situaciones en que el interlocutor es un sistema artificial, como un *chatbot*.

Educación, bulos, burbujas y equidad

Educación es la *E* de FATEN. Educación, *Efecto beneficioso* y *Equidad*. Ya he destacado la importancia de la inversión en la educación a todos los niveles. Además, los sistemas de IA deberían ser diseñados para maximizar su efecto beneficioso sobre la sociedad. En particular, destacaría *los* conceptos de *sostenibilidad*, *veracidad*, *diversidad* y *prudencia*.

El desarrollo tecnológico en general, y de sistemas de Inteligencia Artificial, en particular, conlleva un consumo energético significativo, con impacto negativo en el medio ambiente. Las técnicas de *deep learning* requieren elevadas capacidades de computación

con costes energéticos prohibitivos, sobre todo si consideramos el despliegue de este tipo de sistemas a gran escala.

Es cada vez más importante que el desarrollo tecnológico esté alineado con la responsabilidad humana de garantizar las condiciones para la vida en nuestro planeta, y de preservar el medio ambiente para las generaciones futuras. La Inteligencia Artificial será clave a la hora de abordar algunos de los principales retos medioambientales –desde el cambio climático a la fragmentación del hábitat–, así como para desarrollar medios de transporte –coches autónomos eléctricos– y modelos energéticos más eficientes y sostenibles –las *smart grids* o redes eléctricas inteligentes, que optimizan la producción y la distribución de electricidad con herramientas computacionales–.

La capacidad de generar contenido sintético indistinguible del real utilizando técnicas de IA –*deep fakes*– contribuye a la fabricación de *fake news* o bulos, lo que aumenta el riesgo de manipulación de la opinión pública en cuestiones tan importantes como –por ejemplo– las elecciones presidenciales o la pertenencia a la Unión Europea. Por ello es de suma importancia el principio de *veracidad*, tanto en los datos usados para entrenar algoritmos de IA como en los contenidos que consumimos. De hecho, cada vez más tendremos que usar algoritmos de Inteligencia Artificial para determinar si el audio, vídeo o la noticia que nos llega es real o ha sido inventada... por otro algoritmo de Inteligencia Artificial.

Además, los algoritmos de personalización y recomendación adolecen con frecuencia de falta de diversidad en sus resultados, y tienden a encasillar a sus usuarios en ciertos patrones de gustos. Esto da lugar a lo que Eli Pariser, co-fundador de la plataforma Avaaz.org, ha denominado el *filtro burbuja* [51]. La diversidad en la personalización o en la recomendación de contenidos es muy deseable para ayudarnos a descubrir películas, libros, música, noticias o amigos diferentes a nuestros propios gustos; nos exponemos así a puntos de vista distintos, y nutrimos una mente abierta.

Un concepto relacionado es el de las *cámaras de eco* –*echo chambers*, en inglés–: las personas, al ser expuestas a contenidos que ratifican y amplifican sus propios puntos de vista, los refuerzan, mientras que, por el contrario, desarrollan desconfianza hacia formas de pensar distintas. Si bien estas cámaras de eco han existido desde hace cientos de años, las redes sociales y la ubicuidad de la tecnología amplifican enormemente su impacto. Hay autores que atribuyen a este fenómeno un peso importante en la victoria del Brexit, en el creciente éxito de Donald Trump o en el auge del movimiento antivacunas..

La *E* es también de *Equidad*. El espíritu de solidaridad y la igualdad quizás se están diluyendo con el desarrollo tecnológico.

El crecimiento de internet y del *World Wide Web* durante las Tercera y Cuarta Revoluciones Industriales ha sido sin duda clave para

generalizar el acceso al conocimiento. Sin embargo, los principios de universalización del conocimiento y democratización del acceso a la tecnología están siendo cuestionados hoy en día, en gran parte por la situación de dominancia extrema de las grandes empresas tecnológicas estadounidenses –Alphabet/Google, Amazon, Apple, Facebook, Microsoft– y chinas –Tencent, Alibaba, Baidu–. Es el fenómeno conocido como *winner takes all*, o *el ganador se lo lleva todo*.

Juntos, estos gigantes tecnológicos tienen un valor de mercado de más de 5 billones –millones de millones– de euros, y unas cuotas de mercado en EE. UU. de más de un 90% en las búsquedas de internet (Google), de un 70% de las redes sociales (Facebook) o de un 50% del comercio electrónico (Amazon).

De hecho, el siglo XXI se caracteriza por una polarización en la acumulación de la riqueza. Según un estudio reciente de Oxfam¹⁶, el 1% más rico del planeta posee la mitad de la riqueza mundial y los 2153 billonarios del mundo tienen más que los 4.600 millones de personas más pobres del mundo, es decir, que el 60% de la población mundial. Esta acumulación de riqueza en manos de unos pocos ha sido atribuida al menos parcialmente al desarrollo tecnológico y a la Cuarta Revolución Industrial.

¹⁶ <https://www.oxfam.org/en/press-releases/worlds-billionaires-have-more-wealth-46-billion-people>

Con la Revolución Agraria en el neolítico, y durante miles de años, la propiedad de la tierra conllevaba riqueza. En la Revolución Industrial la riqueza pasó a estar ligada a la propiedad de las fábricas y las máquinas. Hoy en día podríamos argumentar que los datos, y la capacidad para sacarles partido, son el activo que más riqueza genera, dando lugar a lo que se conoce como la economía de los datos.

No podemos olvidar que de los cinco países más poblados del mundo –Facebook, WhatsApp, China, India e Instagram–, tres son de Facebook. Países digitales, globales, con menos de 15 años de existencia de media, con miles de millones de ciudadanos y ciudadanas y que son gobernados por un presidente no elegido democráticamente. En consecuencia, un elevado porcentaje de los datos sobre el comportamiento humano disponibles hoy en día son datos privados, captados, analizados y explotados por estas grandes empresas tecnológicas que conocen no solo nuestros hábitos, necesidades, intereses o relaciones sociales, sino también nuestra orientación sexual o política, y nuestros niveles de felicidad, de educación o de salud mental.

Si queremos maximizar el impacto positivo de la IA en la sociedad, y dado que dicha inteligencia necesita datos para aprender, deberíamos plantearnos nuevos modelos de propiedad, gestión y regulación de los datos.

La Regulación General Europea para la Protección de Datos (RGDP) es un ejemplo en esta dirección. Sin embargo, la complejidad a la hora de aplicar esta norma pone de manifiesto cuán difícil es definir e implementar el concepto de propiedad cuando hablamos de un bien intangible, distribuido, variado, creciente, dinámico y replicable infinitas veces a coste prácticamente cero.

Finalmente, la aplicación de la Inteligencia Artificial exige cumplir requisitos estrictos para su desarrollo, tales como garantizar la disponibilidad de datos suficientes –e idealmente de calidad–, el análisis de las hipótesis de trabajo desde diversas perspectivas, y la disponibilidad de recursos –incluyendo personas con formación– para analizar e interpretar los modelos y sus resultados. El principio de la *prudencia* destaca la importancia de considerar en las fases iniciales de diseño de cualquier sistema las diversas alternativas y opciones existentes, para maximizar su impacto positivo y minimizar los potenciales riesgos y consecuencias negativas derivadas de su aplicación.

N de ‘No maleficencia’

El principio de *No maleficencia* prevé que los sistemas de IA deberían minimizar el impacto negativo que puedan tener en la sociedad. Algunas dimensiones clave en este punto incluyen la *fiabilidad*; la *seguridad*; la *reproducibilidad* de los sistemas de IA; y la *protección de datos* y el respeto hacia la *privacidad*.

La gran mayoría –si no todos– los sistemas, productos y bienes que utilizamos están sujetos a estrictos controles de calidad, seguridad y fiabilidad, para reducir su potencial impacto negativo en la sociedad. Es de esperar que los sistemas de Inteligencia Artificial estén también sujetos a procesos similares. Más allá de los procesos teóricos de seguridad, verificación y fiabilidad, quizás tendría sentido crear una autoridad que a escala europea certificase la calidad, seguridad y fiabilidad de los sistemas de Inteligencia Artificial de manera previa a su comercialización o implementación en la sociedad. En febrero de 2020, la Comisión Europea publicó un libro blanco sobre Inteligencia Artificial¹⁷ que establece las bases para el desarrollo de una Inteligencia Artificial confiable, incluyendo una propuesta de regulación de sistemas de Inteligencia Artificial utilizados en escenarios de alto riesgo, como son la salud, el transporte o la policía.

Asimismo, los sistemas autónomos deberían velar por la integridad de las personas que los utilizan o se ven afectadas por su acción, además de por su propia seguridad frente a la manipulación y los ciberataques.

Para generar confianza los sistemas deberían ser consistentes en su modo de operar, de forma que su comportamiento fuera comprensible y también *reproducible*, es decir, replicable, al ser sometido a los mismos datos de entrada o a la misma situación o contexto.

¹⁷ https://ec.europa.eu/info/publications/white-paper-artificial-intelligence-european-approach-excellence-and-trust_en

En tercer lugar, en un mundo en el que *generamos y consumimos* datos de forma ubicua y masiva los derechos a la protección de la información personal y a la *privacidad* son constantemente cuestionados, e incluso llevados al límite. Numerosos estudios han puesto el foco en el mal uso de datos personales proporcionados por los usuarios de servicios, y han alertado de la agregación de datos de diferentes fuentes por parte de entidades como los *data brokers*, con implicaciones directas en la privacidad de las personas.

A menudo se pasa por alto el hecho de que los avances en los algoritmos de aprendizaje automático, combinados con la disponibilidad de nuevas fuentes de datos, como las redes sociales, permiten inferir información privada que nunca ha sido revelada explícitamente por las personas implicadas. Información relativa, por ejemplo, a la orientación sexual, las tendencias políticas, el nivel educativo o el grado de estabilidad emocional.

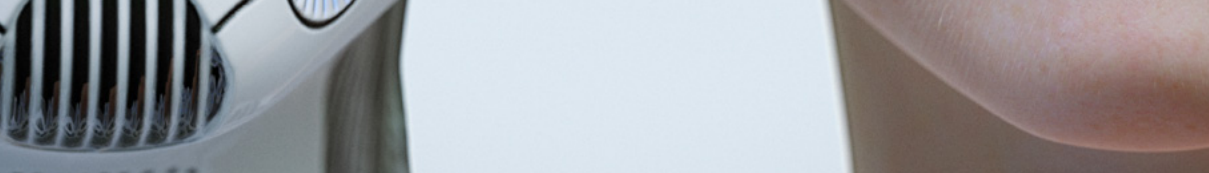
En un proyecto de investigación reciente demostramos que éramos capaces de inferir, a partir de datos no personales, atributos tan personales como algunas dimensiones de la personalidad, el nivel de educación o los intereses [52]. Comprender la existencia de esta nueva capacidad es esencial para entender las implicaciones del uso de algoritmos para modelar e incluso influenciar el comportamiento humano a escala individual, como parece ser que sucedió en el escándalo de Facebook/Cambridge Analytica.

Considero que ciertos atributos y características personales deberían permanecer en la esfera privada mientras la persona no decida expresamente lo contrario, y por tanto no deberían utilizarse o inferirse en los sistemas de IA. Como hemos dicho, Europa ha asumido cierto liderazgo mundial con la reciente entrada en vigor del Reglamento General de Protección de Datos, que se suma a derechos fundamentales como el de establecer y desarrollar relaciones con otros humanos, el de la desconexión tecnológica y el derecho a no ser vigilado.

En este contexto, otros derechos que podríamos o deberíamos agregar incluyen el derecho a un contacto humano significativo – por ejemplo, en servicios de atención operados exclusivamente por *chatbots*–, y el derecho a no ser medido, analizado, orientado o influenciado subliminalmente mediante algoritmos.

En suma, deberíamos siempre centrar el desarrollo de los sistemas de IA en las personas y fomentar la creación de entornos colaborativos para experimentar y co-crear políticas y soluciones basadas en la IA, consensuadas por los humanos.

Solamente cuando respetemos estos principios seremos capaces de avanzar y conseguir un modelo de gobernanza democrática basado en los datos y en la Inteligencia Artificial, por y para las personas.



‘BROGRAMMERS’ Y ESTEREOTIPOS

Una imagen errónea del trabajo en puestos tecnológicos y de quienes lo desempeñan es una de las causas por las que cada vez menos mujeres estudian informática. Las películas, series y medios de comunicación refuerzan el estereotipo de un sector plagado de chicos con gafas, sin inteligencia socioemocional ni higiene personal, tecleando rodeados de comida basura en un sótano sin ventanas. Esta imagen es, sin duda, muy poco atractiva para las chicas y además lejana de la realidad. Entre otros factores que expulsan a las mujeres del ámbito tecnológico destacaría:

- **Una fuerte estereotipación de género** en juguetes, libros, ropa y contenidos audiovisuales que consumen nuestros niños y adolescentes.
- **Los sesgos de género** –conscientes y subconscientes– que tanto hombres como mujeres tenemos, y que conllevan una infravaloración de las mujeres frente a homólogos masculinos con cualificaciones idénticas [53].
- **Una falta sistemática de reconocimiento hacia las mujeres en todos los contextos, y en particular en las carreras tecnológicas.** Ilustran este hecho la

brecha salarial y la escasa presencia de mujeres en distinciones y puestos de poder. El premio Turing, el equivalente al Nobel en Informática, ha sido otorgado desde su creación en 1966 únicamente a tres mujeres, frente a 62 hombres. Además, las mujeres fundadoras de empresas solo recibieron un 2% de las inversiones de capital riesgo de EE. UU. en 2017. Una compañía emergente creada por un equipo femenino atrae de media 82 dólares de inversión por cada 100 dólares obtenidos por equipos masculinos, y esto a pesar de que los resultados de estas *startups* fundadas por mujeres son por lo general mejores.

- La **inexistencia de referentes femeninos** que puedan animar a niñas y adolescentes a estudiar estas carreras.
- La **marcadamente sexista y misógina cultura *brogrammer***, en alusión al término *bro* –hermano en inglés– con que se llaman entre sí coloquialmente los programadores. En un estudio publicado por la revista *Fortune*, la camaradería masculina de la poco diversa cultura *brogrammer* fue el segundo motivo más frecuente –citado por un 68% de las mujeres– del abandono del trabajo en el sector tecnológico, después de la maternidad.



PARA ESTRECHAR LA BRECHA DE GÉNERO

Numerosas iniciativas y asociaciones buscan atraer más mujeres al ámbito tecnológico. El proyecto Mujer e Ingeniería, de la Real Academia Española de Ingeniería, apoya a chicas estudiantes de ingeniería a través de redes de tutorización. El concurso *Wisibizalas*, de la Universidad Pompeu Fabra (UPF), promueve en colegios e institutos la visibilidad de las mujeres en el ámbito tecnológico y la reflexión en torno a la brecha de género. La Asociación de Mujeres Investigadoras y Tecnólogas (AMIT) tiene entre sus fines la defensa de la equidad de género. *MujeresTech* es una asociación sin ánimo de lucro que ofrece recursos y conocimiento para aumentar la presencia femenina en el sector digital. Las Top100 es una iniciativa para identificar las diez mujeres más influyentes en España en diez categorías; incluye una categoría de emprendedoras y otra de académicas e investigadoras. El Premio Ada Byron a la Mujer Tecnóloga, de la Universidad de Deusto, que tuvo el honor de recibir en 2016, premia y visibiliza trayectorias excelentes de mujeres en diversos campos tecnológicos.

Conclusiones

Desconozco qué nos deparará el futuro y dónde estaremos dentro de veinte años. Pero sí puedo describir cómo me gustaría que fuese.

En primer lugar, deseo que sea un futuro donde la tecnología en general –y la Inteligencia Artificial en particular– forme parte integral de nuestras vidas, donde coexistamos sinérgica y armónicamente con tecnología que nos ayude a vivir más, y sobre todo mejor, a todos. El potencial de la IA es inmenso: no deberíamos desaprovechar esta oportunidad para mejorar la calidad de vida de las personas, del resto de seres vivos y de nuestro planeta. Sin embargo, este futuro del que aspiro a formar parte, y al que quiero contribuir con mi trabajo, no está garantizado. Si queremos convertirlo en realidad debemos afrontar seriamente tanto las limitaciones de los actuales sistemas de Inteligencia Artificial, como los retos que plantean.

En segundo lugar, considero que España debería invertir mucho más de lo que invierte hoy día en Inteligencia Artificial, para convertirse en líder en Europa y puente con Latinoamérica y África. Sería muy deseable una apuesta ambiciosa por la adopción de la IA en nuestras empresas y Administraciones públicas, por nuestra excelencia científica en esta área, por la formación, atracción y retención del talento.

Deberíamos actualizar nuestro sistema educativo, y estimular la creatividad y la innovación. Es la ocasión de elevar no solo el crecimiento empresarial y económico, sino sobre todo nuestra calidad de vida. Espero y deseo que no dejemos escapar esta oportunidad.

Finalmente, espero, deseo –y sueño– que cada vez haya más mujeres ingenieras, investigadoras, inventoras, innovadoras en tecnología, que sean excepcionales no en sentido literal –no porque su condición de mujeres las convierte en la excepción–, sino por la brillantez de sus ideas y el impacto de su trabajo.

La Inteligencia Artificial es una herramienta sumamente poderosa y necesaria para ayudarnos a abordar los inmensos retos que debemos superar como especie. Pero si a la hora de desarrollarla pasamos por alto los factores sociales, laborales y éticos, la Inteligencia Artificial puede convertirse en arma a favor de la desigualdad, el control y la destrucción.

Comparto las sabias palabras del genial astrofísico Stephen Hawking: “La Inteligencia Artificial puede ser lo mejor o lo peor que nos ha sucedido a la humanidad”. Durante un cuarto de siglo he dedicado mi actividad profesional a decantar la balanza hacia el lado bueno. Trabajo para que la IA sea lo mejor que nos ha pasado. Les invito, llena de esperanza, a que se unan a esta causa.

Referencias

1. N. Oliver, A. Pentland y F. Bérard, «LAFTER: a real-time face and lips tracker with facial expression recognition», *Pattern Recognition*, vol. 33, n° 8, pp. 1369-1382, 2000.
2. N. Oliver, E. Horvitz y A. Garg, «Layered representations for human activity recognition», *Proceedings of the Fourth IEEE International Conference on Multimodal Interfaces*, Pittsburgh, PA, USA, 2002.
3. N. Oliver, B. Rosario y S. Pentland, «A Bayesian computer vision system for modeling human interactions,» *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, n° 8, pp. 831-843, 2000.
4. N. Oliver y A. Pentland, «Graphical models for driver behavior recognition in a SmartCar», *Proceedings of the IEEE Intelligent Vehicles Symposium 2000*, Dearborn, MI, USA. 2000.
5. N. Oliver y F. Flores-Mangas, «HealthGear: a real-time wearable system for monitoring and analyzing physiological signals», *Proceedings of the International Workshop on Wearable and Implantable Body Sensor Networks*, Cambridge, MA, USA, 2006.
6. J. San Pedro, D. Proserpio y N. Oliver, «MobiScore: towards universal credit scoring from mobile phone data,» de *International Conference on User Modeling, Adaptation, and Personalization*, 2015.

7. Bogomolov, B. Lepri, J. Staiano, N. Oliver, Pianesi y A. F. and Pentland, «Once upon a crime: towards crime prediction from demographics and mobile data» Proceedings of the ACM Int Conf on Multimodal Interaction (ICMI), 2014.
8. M. Pielot, T. Dingler, J. San Pedro y N. Oliver, «When attention is not scarce-detecting boredom from mobile phone usage», Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing, Osaka, Japan, 2015.
9. R. de Oliveira, M. Cherubini y N. Oliver, «MoviPill: improving medication compliance for elders using a mobile persuasive social game», Proceedings of the 12th ACM international conference on Ubiquitous computing, Copenhagen, Denmark, 2010.
10. N. Oliver y F. Flores-Mangas, «MPTrain: a mobile, music and physiology-based personal trainer», Proceedings of the 8th conference on Human-computer interaction with mobile devices and services, Helsinki, Finland, 2006.
11. D. Rumelhart, G. E. E. Hinton y R. J. Williams, «Learning representations by back-propagating errors», Nature, vol. 323, pp. 533--536, 1986.
12. N. Oliver, A. Pentland y F. Berard, «LAFTER: lips and face real time tracker», Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Juan, Puerto Rico, USA, 1997.

13. M. Brand, N. Oliver y S. Pentland, «Coupled hidden Markov models for complex action recognition», Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Juan, Puerto Rico, USA, 1997.
14. Y. LeCun, Y. Bengio y G. Hinton, «Deep Learning» Nature, vol. 521, pp. 436--444, 2015.
15. A. Karatzoglou, X. Amatriain, L. Baltrunas y N. Oliver, «Multiverse recommendation: n-dimensional tensor factorization for context-aware collaborative filtering», Proceedings of the fourth ACM conference on Recommender systems, Barcelona, Spain, 2010.
16. Y. Shi, A. Karatzoglou, L. Baltrunas, M. Larson, N. Oliver y A. Hanjalic, «CLiMF: learning to maximize reciprocal rank with collaborative less-is-more filtering», Proceedings of the ACM conference on Recommender systems, New York, 2012.
17. X. Amatriain, N. Lathia, J. Pujol, H. Kwak y N. Oliver, «The wisdom of the few: a collaborative filtering approach based on expert opinions from the web», Proceedings of the 32nd international ACM SIGIR conference on Research and Development in Information Retrieval, 2009.
18. M. Willson, «Algorithms (and the) everyday», Information, Communication & Society., 2016.

19. S. Centellegher, G. Miritello, D. Villatoro, D. Parameshwar, B. Lepri and N. Oliver, «Mobile Money: Understanding and Predicting its Adoption and Use in a Developing Economy», Proceedings of the ACM Int. Conference on Interactive, Mobile, Wearable and Ubiquitous Computing (Ubicomp'19), London, 2019.
20. T. Wang., C. Rubin, D. Wagner y R. Sevieri, «Learning to detect patterns of crime», Machine Learning and Knowledge Discovery in Databases, pp. 515-530, 2013.
21. J. Floyd, P. Rente, N. Ruktanonchai, A. Tatem y N. Oliver, «Malaria parasite mobility in Mozambique estimated using mobile phone records» in NetMob 2019, Oxford, UK, 2019.
22. P. Jawandhiya, «Hardware design for machine learning», International Journal of Artificial Intelligence and Applications (IJAIA), vol. 9, n° 1, pp. 63 -- 84, 2018.
23. N. Bostrom, «SuperIntelligence: Paths, dangers, strategies, Oxford, UK: Oxford University Press, 2014. J. Hawkins, «What intelligent machines need to learn from the neocortex» IEEE Spectrum, 2017.
24. J. Hawkins, «What intelligent machines need to learn from the neocortex» IEEE Spectrum, 2017.

25. S. Ahmad y J. Hawkins, «Properties of Sparse Distributed Representations and their Application to Hierarchical Temporary Memory,» Arxiv, 2015.
26. A. Andrae, «Total Consumer Power Consumption Forecast,» de Nordic Digital Business Summit, 2017.
27. D. Lazer, M. Baum, Y. Benkler, A. J. Berinsky, K. Greenhill, F. Menczer y e. al, «The science of fake news», Science, pp. 1094--1096, 2018.
28. Y. N. Harari, «21 lessons for the 21st century», London: Penguin Random House, UK, 2018.
29. S. Bocconi, S. Chiocciariello, G. Dettori, A. Ferrari y K. Engelhardt, «Developing computational thinking in compulsory education», JCR Science for Policy Report, 2016.
30. S. Lluna y J. Pedreira, «Los nativos digitales no existen», Deusto Editorial, 2017.
31. E. Kross, W. Mischel y Y. Shoda, «Enabling self-control», Social Psychological Foundations of Clinical Psychology, nº June, pp. 375-394, 2011.
32. S. Bocconi, A. Chiocciariello, G. Dettori, A. Ferrari y K. Engelhardt (2016). «Developing computational thinking in compulsory education – Implications for policy and practice», EUR 28295 EN; doi:10.2791/792158. <http://publications.jrc.ec>.

33. S. Bocconi, A. Chiocciariello y J. Earp, «The Nordic approach to introducing Computational Thinking and programming in compulsory education», Report prepared for the Nordic@BETT2018 Steering Group. <https://doi.org/10.17471/54007>
34. M. Berry, «Computing in the national curriculum A guide for primary teachers», ISBN: 978-1-78339-143-1. <https://www.computingatschool.org.uk/data/uploads/CASPrimaryComputing.pdf>
35. J. Decety y P.L. Jackson, «The functional architecture of human empathy», Behavioral and Cognitive Neuroscience Review, vol. 3, pp. 71-100, 2004.
36. B. E. M. A. a. D. J. Jackson P.L., «Empathy examined through the neural mechanisms involved in imagining how I feel versus how you feel pain: An event-related fMRI study», Neuropsychologia, vol. 44, pp. 752-61, 2006.
37. T. Goetz, N. Hall, A. Frenzel y U. Nett «Types of boredom: An experience sampling approach», Motivation and Emotion, vol. 38, nº 3, pp. 401-19, 2014.
38. T. Wilson, D. Reinhard, E. Westgate, D. Gilbert, N. Ellerbeck, C. Hahn, C. Brown y A. Shaked, «Just think: The challenges of the disengaged mind», Science, vol. 345, pp. 75-77, 2014.
39. S. J. Vodanovich, «On the possible benefits of boredom: A neglected area in personality research», Psychology and Education: An Interdisciplinary Journal, 2003.

40. J. Eastwood, A. Frischen y M. Fenske, «The unengaged mind: Defining boredom in terms of attention», *Perspectives on Psychological Science*, vol. 7, n° 5, pp. 482-95, 2012.
41. B. Brown, M. McGregor y D. McMillan, «100 days of iPhone use: Understanding the details of mobile device use», *Proceedings of the ACM MobileHCI*, 2014.
42. A. Matic, M. Pielot y N. Oliver, «Boredom-Computer Interaction: Boredom Proneness and SmartPhone Use», *Proceedings of the ACM Int Conf on Ubiquitous Computing (Ubicomp)*, 2015.
43. M. Pielot, T. Dingler, J. San Pedro y N. Oliver, «When Attention is not Scarce: Detecting Boredom from Mobile Phone Usage», *Proceedings of the ACM Int Conf on Ubiquitous Computing (Ubicomp)*, 2015.
44. S. Turkle, «En defensa de la conversación», Ed. Atico de los libros, ISBN-10 8416222274, 2015.
45. Y. Torres Fernández, D. Pastor Escuredo, A. Morales Guzmán, J. Baue, A. Wadhwa, C. Castro Correa, L. Romanoff, J. Lee, A. Rutherford, V. Frias Martínez, N. Oliver, E. Frias-Martinez y M. Luengo-Oroz, «Flooding through the lens of mobile phone activity», *Proceedings of IEEE Global Humanitarian echnology Conference*, San Jose, CA, USA, 2014.

46. S. Centellegher, M. De Nadai, M. Caraviello, C. Leonardi, M. Vescovi, Y. Ramadian, N. Oliver, F. Pianesi, A. Pentland, F. Antonelli y B. Lepri, The Mobile Territorial Lab: a multilayered and dynamic view on parents' daily lives, EPJ Data Science, 2016.
47. J. Staiano, N. Oliver, B. Lepri, R. de Oliveira, M. Caraviello y N. Sebe, «Money walks: a human-centric study on the economics of personal mobile data», Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing, Seattle, WA, USA, 2014.
48. B. Lepri, N. Oliver, E. Letouzé, A. Pentland y P. Vinck, «Fair, Transparent, and Accountable Algorithmic Decision-making Processes», Philosophy & Technology, pp. 1-17, 2017.
49. L. Floridi, J. Cows, M. Beltrametti, R. Chatila, P. Chazerand, V. Dignum, C. Luetge, R. Madelin, U. Pagallo, F. Rossi, B. Shafer, P. Valcke y E. Vayena, «An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles and Recommendations», Minds and Machines, nº December, 2018.
50. J. Burrell, «How the machine 'thinks': Understanding opacity in machine learning algorithms», Big Data and Society, 2016.
51. E. Pariser, «The filter bubble: how the personalized web is changing what we read and how we think», New York: Penguin Books, 2012.

52. S. Park, A. Matic, K. Garg y N. Oliver, «When Simpler Data Does Not Imply Less Information: A Study of User Profiling Scenarios With Constrained View of Mobile HTTP (S) Traffic», ACM Transactions on the Web (TWEB), vol. 12, nº 9, 2018.
53. R. Steinpreis, K. Anders y D. Ritzke, «The Impact of Gender on the Review of the Curricula Vitae of Job Applicants and Tenure Candidates: A National Empirical Study», Sex Roles, vol. 41, nº 7--8, pp. 509-528, 1999.
54. N. Oliver, «Governance in the Era of Data-driven Decision-making Algorithms», Women Shaping Global Economic Governance, CEPR, July 2019
55. Comisión Europea, «European Commission. Ethics guidelines for trustworthy AI», Digital Single Market, 2019 <https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai>
56. Facebook Data for Good, «We use data to address some of the world's greatest humanitarian issues». <https://dataforgood.fb.com/>
57. «Google AI». <https://ai.google/social-good/impact-challenge/>
58. «Microsoft AI». <https://www.microsoft.com/en-us/ai/ai-for-good>
59. «Web Opal Project». <https://www.opalproject.org/>

60. «Meetings of the Expert Group on Business-to-Government Data Sharing», Digital Single Market, 2019. <https://ec.europa.eu/digital-single-market/en/news/meetings-expert-group-business-government-data-sharing>
61. OECD. «Recommendation of the Council on Artificial Intelligence». , Pub. L. No. OECD/LEGAL/0449 (2019). <https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449>
62. World Economic Forum. (2018). «The Future of Jobs Report 2018». http://www3.weforum.org/docs/WEF_Future_of_Jobs_2018.pdf
63. Kiss, M., & Parlamento Europeo. Dirección General de Servicios de Estudios Parlamentarios. (2016). «Digital skills in the EU labour market [er]: In-depth analysis». <http://dx.publications.europa.eu/10.2861/451320>
64. Wing, J. M. (2006, marzo). «Computational Thinking». COMMUNICATIONS OF THE ACM, 49 No 3. <https://www.cs.cmu.edu/~15110-s13/Wing06-ct.pdf>
65. Activate. (2018). «Activate Tech & Media Outlook 2019». <https://www.slideshare.net/ActivateInc/activate-tech-media-outlook-2019>

66. Purdy, Mark y Daugherty, Paul. Accenture (2017). «How AI Boosts Industry Profits And Innovation». https://www.accenture.com/fr-fr/_acnmedia/36dc7f76eab444cab6a7f44017cc3997.pdf
67. Human Age Institute. «Posiciones y competencias más demandadas. Informe EPyCE 2018». <https://humanageinstitute.org/wp-content/uploads/2019/03/Informe-EPYCE-2018.pdf>
68. NCWIT. «What is the impact of gender diversity on technology business performance? ». https://www.ncwit.org/sites/default/files/resources/impactgenderdiversitytechbusinessperformance_print.pdf
69. Catalyst (2014). «Infographic: Diversity Matters». <https://www.catalyst.org/research/infographic-diversity-matters/>
70. Christiansen, L., Lin, H., & Pereira, J. (2016). «Gender Diversity in Senior Positions and Firm Performance: Evidence from Europe». IMF Working Paper. <https://www.imf.org/external/pubs/ft/wp/2016/wp1650.pdf>
71. European Commission. «Increase in gender gap in the digital sector. Study on Women in Digital Age». <https://ec.europa.eu/digital-single-market/en/news/increase-gender-gap-digital-sector-study-women-digital-age>

72. Sustainable Development. «UN World Data Forum 2018 wraps up with launch of Dubai Declaration». <https://www.un.org/sustainabledevelopment/blog/2018/10/un-world-data-forum-2018-wraps-up-with-launch-of-dubai-declaration/>
73. Andreas Elpidorou, "The bright side of boredom", Front. Psychol., 03 November 2014 | <https://doi.org/10.3389/fpsyg.2014.01245>

