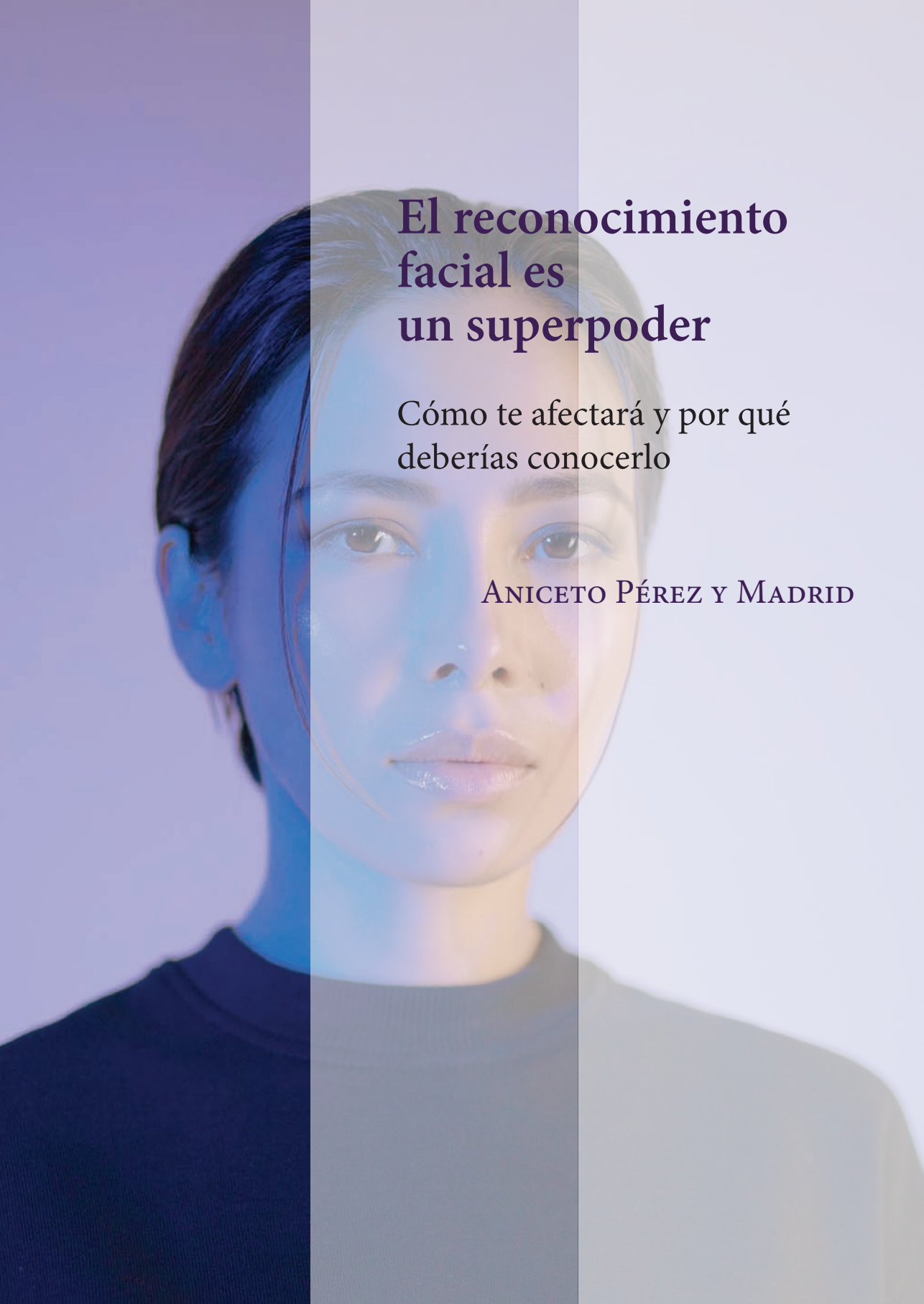

El reconocimiento facial es un superpoder

Cómo te afectará y por qué deberías conocerlo

Aniceto Pérez y Madrid

EL RECONOCIMIENTO FACIAL ES UN SUPERPODER

EL RECONOCIMIENTO FACIAL ES UN SUPERPODER

Cómo te afectará y por qué deberías conocerlo

Aniceto Pérez y Madrid

DYKINSON
2021

Extravagantes, 7
ISSN: 2660-8693

Motivo de cubierta:
Fotografía de Aiony Haust disponible en Unsplash
https://unsplash.com/photos/3TLl_97HNJo

Editorial Dykinson
c/ Meléndez Valdés, 61 – 28015 Madrid
Tlf. (+34) 91 544 28 46
E-mail: info@dykinson.com
http://www.dykinson.com

Preimpresión: TallerOnce

ISBN: 978-84-1377-667-5

Versión electrónica disponible en e-Archivo
http://hdl.handle.net/10016/33010

Dataset/Conjunto de datos disponible en:
https://doi.org/10.21950/5II6LN

A mis padres

SUMARIO

Presentación

En 2018, mientras colaboraba con *Global Translator Community* de Coursera, fui invitado a participar en la traducción del curso *AI for Everyone*. En ese curso descubrí un campo desconocido y apasionante, impartido por un instructor excepcional, el profesor Andrew Ng. Al terminar me inscribí en una especialización en Deep Learning, también impartida por el profesor Ng.

Ese año, la ACM otorgó el Premio Turing a Joshua Bengio, Geoffrey Hinton y Yann LeCunn por "avances conceptuales y de ingeniería que han hecho de las redes neuronales profundas un componente crítico de la informática"[1]. Esta noticia consolidó mi interés en la inteligencia artificial y sus aplicaciones en el mundo real. Empecé a buscar noticias, fuentes, organizaciones y a seguir a investigadores. Mediante la cuenta de Twitter @forodeeplearn comencé a recibir información abundante. Luego me percaté de que necesitaba almacenar y clasificar toda esa información para poder consultarla en el futuro. Así que abrí el blog http://actualidaddeeplearning.blogspot.com. En el momento de escribir estas palabras, el blog *Actualidad Deep Learning* contiene más de 1.200 entradas.

Mi interés se dirigió inicialmente a los avances tecnológicos. Cada día se publicaba un nuevo récord de precisión, un algoritmo más efectivo o una aplicación novedosa. Paulatinamente cambió el tipo de noticias, había menos avances tecnológicos; otros temas como sesgos, transparencia, explicabilidad y ética, estrategias nacionales de inteligencia artificial y otras cuestiones como discriminación algorítmica e impacto laboral y social de la inteligencia artificial pasaron a ser protagonistas de la actualidad de la inteligencia artificial. Estas nuevas cuestiones despertaron en mí un interés mayor que las técnicas y eso me impulsó a realizar un curso sobre Ética de los Datos, *IA e Innovación Responsable*.

Entre las diversas aplicaciones de la inteligencia artificial, el reconocimiento facial tiene características e implicaciones únicas. Parafraseando el fantástico libro de Carissa Véliz *Privacidad es Poder*, he titulado este trabajo, *El Reconocimiento Facial es un Superpoder*.

---

1 "Fathers of the Deep Learning Revolution Receive ACM A.M. Turing Award", *Association for Computing Machinery*, https://awards.acm.org/about/2018-turing

Durante 2019 y 2020 ha ocurrido una larga lista de inquietantes acontecimientos relacionados con el uso del reconocimiento facial. La reacción ha sido variada: desde prohibiciones del uso de la tecnología a la promoción de su empleo con la pretensión de controlar sus posibles efectos negativos.

El filósofo Mark Coeckelbergh dice en una entrevista

> Hay una brecha en términos de conocimiento, que también es una brecha en términos de poder [...] necesitamos una transferencia de conocimiento de las empresas al público [no solo] de material técnico. Lo que más necesitamos es conocimiento general sobre cómo funciona la tecnología sin mirar el código, conocimiento de lo que sucede detrás de la pantalla. El principal problema es que muchas cosas son invisibles para nosotros en el funcionamiento de la inteligencia artificial y la ciencia de datos, de dónde toman nuestros datos, qué se hace con ellos, ¿los venden?, etc. Debería ser una obligación de las empresas darnos información sobre lo que están haciendo con nuestros datos sin necesidad de detalles técnicos específicos.[2]

Creo que no sólo las empresas que crean sistemas de inteligencia artificial deben informar, también los gobernantes, el sector público y las fuerzas del orden. Es un requisito estar bien informados, especialmente los reguladores y creadores de políticas.

El reconocimiento facial es un tema con muchas implicaciones y consecuencias y de especial relevancia como podremos ver. La obra presenta, a nivel de divulgación, el reconocimiento facial con cuatro enfoques: tecnológico, social, ético y legal, una visión general y pluridisciplinar.

El libro incluye una extensa bibliografía, tanto noticias como artículos científicos, muchos de ellos con un enlace para poder acceder al contenido completo. Animo al lector a que perciba la amplia dimensión del tema, incluso profundizar. Para enviar comentarios podrá usar esta dirección de correo deeplearningspain@gmail.com.

Quiero expresar un agradecimiento especial a mi editor Manuel Martínez Neira que ha impulsado este proyecto desde el principio y ha asegurado que llegue a buen puerto. También agradezco a María Belén Andreu Martínez por su esencial colaboración, a Miguel Angel Correas por su detallada revisión e interesantes aportaciones, a Ángel Gómez de Ágreda por su apoyo y su tiem-

2 Christian Höller, "The Price of Freedom. Interview with Mark Coeckelbergh about the Ethics of Artificial Intelligence", *Springerin*, https://www.springerin.at/en/2021/1/preis-der-freiheit/

po, y a Connor Wright por sus comentarios y entusiasmo. Agradezco a Anish Athalye su permiso para incluir la imagen de la tortuga sintética, y a Kevin Eykholt por la señal de stop vandalizada.

4 de julio de 2021.

Aniceto Pérez y Madrid

Introducción

En enero de 2020 The New York Times informaba de que la startup Clearview había desarrollado un innovador sistema de Reconocimiento Facial (RF) que era estaba siendo utilizada por cientos de agencias policiales, por seguridad privada, y que algunas personas la utilizaban en fiestas, reuniones de negocios, para divertirse o como exhibición de poder[3].

Hoan Ton-That es un emprendedor australiano controvertido. En 2016 entró en el negocio del RF con Richard Schwartz, ex ayudante de Rudolf Giuliani cuando era alcalde de Nueva York. Acordaron que Ton-That se encargaría de la parte técnica y Schwartz usaría sus contactos para la comercialización.

Aunque los departamentos de policía de Estados Unidos han tenido acceso a herramientas de RF durante casi 20 años, la tecnología se había limitado a la búsqueda en imágenes proporcionadas por el gobierno: fotos policiales y licencias de conducir. Al mejorar los algoritmos de RF en los últimos años, empresas como Amazon empezaron a ofrecer productos para realizar RF con cualquier base de datos de imágenes.

Ton-That quería ir mucho más allá, así que diseñó un programa para rastrear por Internet y recopilar imágenes de rostros de personas en sitios de empleo, de noticias, educativos y también de redes sociales como Facebook, YouTube, Twitter e Instagram. Aunque Facebook y otros sitios de redes sociales prohíben en sus términos de uso la utilización de sus contenidos para otros fines, Clearview los incumple sistemáticamente.

Ton-That desarrolló el software de RF y creó la app *Clearview AI*[4] para identificar personas. Clearview AI resulta atractiva porque ofrece un servicio único: un buen software de RF y una enorme base de datos de imágenes identificadas. El problema es que Clearview AI se basa en imágenes obtenidas sin conocimiento ni permiso de los interesados, un uso no autorizado de datos personales[5].

---

3 Kashmir Hill, "Before Clearview Became a Police Tool, It Was a Secret Plaything of the Rich", *The New York Times*, https://www.nytimes.com/2020/03/05/technology/clearview-investors.html

4 Kashmir Hill, "The Secretive Company That Might End Privacy as We Know It", *The New York Times*, https://www.nytimes.com/2020/01/18/technology/clearview-privacy-facial-recognition.html

5 Los datos personales son cualquier información relativa a una persona física viva identificada o identificable. Las distintas informaciones, que recopiladas pueden llevar a la

Clearview publicita sus servicios como ayuda a las fuerzas del orden para resolver delitos, lucha contra el terrorismo y otras labores policiales[6]. Una reciente investigación de BuzzFeed ha revelado que Clearview ha estado ofreciendo pruebas gratuitas y más de 7.000 empleados de agencias del orden de Estados Unidos la han usado sin conocimiento de sus superiores[7].

En febrero de 2021 la *Office of the Privacy Commissioner of Canada* publicó el resultado de una investigación acerca de la utilización de Clearview AI por la *Real Policía Montada de Canadá* (RCMP). La investigación encontró que la RCMP disponía de 48 cuentas para uso policial. El informe decía:

> Lo que hace Clearview es vigilancia masiva y es ilegal. Es completamente inaceptable que millones de personas que nunca se verán implicadas en ningún crimen se encuentren continuamente en una rueda de identificación policial[8].

Una investigación similar fue realizada por la *Swedish Authority for Privacy Protection* a la *Police Authority* de Suecia sobre el uso de Clearview AI por investigadores policiales. El informe de febrero de 2021 concluía que algunos empleados de la *Police Authority* habían utilizado Clearview AI sin autorización y por ello le fue impuesta una multa de 2.500.000 SEK, unos 250.000 EUR[9].

El caso de ClearView AI no es aislado y se pueden intuir algunos problemas asociados con el RF. A lo largo de los siguientes capítulos veremos muchos más casos de uso del RF. Empezaremos con la tecnología.

---

identificación de una determinada persona, también constituyen datos de carácter personal. En Europa el Reglamento General de Protección de Datos (RGPD) protege el derecho a la privacidad imponiendo ciertas restricciones.

6 http://clearview.ai

7 Ryan Mac et al., "Surveillance Nation", *BuzzFeed News,* https://www.buzzfeednews.com/article/ryanmac/clearview-ai-local-police-facial-recognition

8 "Clearview AI's unlawful practices represented mass surveillance of Canadians, commissioners say"*, Office of the Privacy Comissioner of Canada,* https://www.priv.gc.ca/en/opc-news/news-and-announcements/2021/nr-c_210203/

9 "Swedish DPA: Police unlawfully used facial recognition app", European Federation of Data Protection Officers, https://www.efdpo.eu/swedish-dpa-police-unlawfully-used-facial-recognition-app/

Tecnología: Inteligencia artificial

La idea de construir máquinas que tengan un comportamiento similar al humano ha sido una constante desde la antigüedad, pero solo nos remontaremos a mediados del siglo XX. Alan Turing, un matemático excepcional, escribió en 1950 un artículo en el que se preguntaba: ¿pueden pensar las máquinas? En él proponía lo que se ha llamado el *Test de Turing*, una prueba para determinar la capacidad de una máquina de exhibir un comportamiento inteligente.

En 1955, John McCarthy, profesor en Dartmouth College, decidió crear un grupo de estudio para desarrollar las ideas que habían surgido en los años precedentes sobre las *máquinas pensantes*. Se desarrolló como un curso de verano en 1956. En la propuesta del curso, John McCarthy escribió

> Se intentará encontrar cómo hacer que las máquinas usen el lenguaje, formen abstracciones y conceptos, resuelvan tipos de problemas ahora reservados para los humanos y se mejoren a sí mismas[10].

Si bien se han realizado grandes avances, casi siete décadas después todos estos temas siguen abiertos, los sistemas de IA actuales carecen de la capacidad de formar conceptos y abstracciones semejantes a los humanos.

Se puede definir la Inteligencia Artificial (IA) como el conjunto de tecnologías que permiten simular operaciones que habitualmente se consideran reservadas a los seres humanos.

Simular operaciones deja claro que la IA no trata realmente de inteligencia en el sentido humano, sino de una simulación. Desde 2011 se han desarrollado sistemas inteligentes bastante precisos, algunos ya ampliamente disponibles, como traducción de textos, conversión de voz a texto y sistemas conversacionales como Siri y Alexa. A veces muestran resultados sorprendentes y otras decepcionantes.

Cade Metz resume de forma ingeniosamente simplificada el debate que surgió desde los primeros momentos en la forma de desarrollar la IA, "pensar

10 John McCarthy et al., "A PROPOSAL FOR THE DARTMOUTH SUMMER RESEARCH PROJECT ON ARTIFICIAL INTELLIGENCE", Stanford University, http://jmc.stanford.edu/articles/dartmouth.html

como una persona" frente a "pensar como una máquina"[11]. Profundicemos un poco en estos dos planteamientos.

Al principio la IA se enfocó a construir máquinas de razonamiento basadas en manejar aserciones y reglas, como piensan las personas. Este enfoque dio lugar a sistemas en los que se codificaban los conocimientos sobre un tema a fin de crear una máquina que se comportase como un experto en la materia. Este enfoque dio lugar a lo que con el tiempo se denominó *sistemas expertos*. El diagnóstico médico fue uno de los campos de aplicación de estos sistemas.

Para medir la inteligencia de una máquina los científicos pensaron que lo más directo sería confrontarla con la inteligencia humana. Una de las formas más frecuentes de hacerlo ha sido los juegos. Al desarrollarse en espacios controlados con reglas limitadas y claras, los juegos son un escenario muy apropiado para una máquina. El reto por excelencia durante años ha sido el ajedrez. Tradicionalmente se ha considerado el ajedrez como el summum de la inteligencia. Siempre habían ganado los jugadores humanos hasta que en 1997 el ordenador Deep Blue de IBM derrotó al campeón mundial Garry Kasparov[12]. La victoria de Deep Blue en el ajedrez supuso un hito, aunque apenas tuvo impacto social, salvo en las noticias.

En 2011, catorce años después de la victoria de Deep Blue, IBM presentó la inteligencia artificial Watson[13] al concurso televisivo Jeopardy!, y ganó. El concurso, en líneas generales, consistía en que un concursante daba un dato que era una respuesta y los demás concursantes tenían que adivinar la pregunta. Watson debía encontrar nombres, números, fechas y frases así como analizar la gramática y la sintaxis de las respuestas. Watson supuso un paso importante para la opinión pública y la inteligencia artificial.

Los sistemas expertos pretendían simular el saber humano, pero con el tiempo resultó ser una tarea muy complicada, incluso en campos delimitados.

---

11 William Softky, "The Unbearable Shallowness of 'Deep AI'", *Fair Observer*, https://www.fairobserver.com/business/technology/william-softky-cade-metz-book-genius-makers-artificial-intelligence-ai-tech-technology-news-91649/

12 Deep Blue es un superordenador construido por IBM a partir de un sistema RS/6000 de proceso paralelo que incluía chips especializados. Emplea el algoritmo minimax de teoría de juegos para decidir los movimientos

13 Watson es un potente superordenador desarrollado por IBM constituido por noventa servidores y 16 Terabytes de RAM. Toda la información estaba en memoria para poder acceder rápidamente. Utiliza la tecnología DeepQA de IBM para analizar preguntas y producir respuestas

Los problemas del mundo real son complejos, las normas no están claras, no siempre son consistentes, e influye el contexto. La falta de resultados y el pesimismo en la prensa entre los años 1980-90 causaron un recorte en la financiación de los proyectos y la decepción dio paso a lo que se denominó *Invierno de la IA*[14].

Aprendizaje Automático

En el curso de verano de 1956 también surgió otro enfoque de la IA: simular la inteligencia a partir de señales simples, no a partir de conceptos o reglas; pensar como máquinas, no como personas. En 1958, el psicólogo Frank Rosenblatt, apoyándose en los trabajos de Ramón y Cajal, que había descubierto las neuronas y la conexión neuronal a finales del siglo XIX, ideó el *perceptrón* o neurona artificial, una unidad básica de proceso que conectada a otras podría crear algoritmos que aprendieran a discriminar algo. El perceptrón en realidad es una fórmula matemática cuya representación gráfica se asemeja a una neurona. Se supone que esta es la forma de trabajar de la visión. La retina recibe la luz que se transmite al nervio óptico, es procesada por la neuronas y el cerebro interpreta la imagen como una mesa o una persona con nombre y apellidos. La idea de Rosenblatt era que una red de neuronas podría aprender a discriminar formas como lo hace el sistema neuronal animal.

Figura 1: Diagrama del Perceptrón de Rosenblatt

Cuando llegó el Invierno de la IA, se buscaron nuevos caminos y el enfoque de utilizar señales parciales empezó a tomar impulso y se propusieron diseños de neuronas artificiales interconectadas, llamadas *redes neuronales*.

Figura 2: Red neuronal

Conforme el problema a resolver era más complejo, la red debía ser más compleja, quizás más datos de entrada y/o más capas. El gran problema de las redes neuronales, mayor conforme son más grandes, es el ajuste del *peso* de cada señal para obtener el resultado deseado. En 1986 se perfeccionó la técnica *Backpropagation*[15] que ha permitido desde entonces automatizar el

14 En 1984 el tema central de la conferencia anual de la American Association for Artificial Intelligence fue la redacción de fondos y la pérdida de interés en la IA. Por paralelismo al *invierno nuclear* se denominó a este periodo *invierno de la IA*

15 David E. Rumelhart, Geoffrey E. Hinton y Ronald J. Williams, "Learning re-

proceso de ajuste de las redes neuronales, llamado *Aprendizaje*, asimilando el ajuste de las redes al aprendizaje de los seres humanos.

El aumento paulatino de la potencia de los ordenadores, la gran disponibilidad de datos por la extensión de la digitalización y los avances en los algoritmos han permitido desarrollar redes de muchas capas. La tecnología de los sistemas basados en redes neuronales de tres o más capas recibe el nombre de *Aprendizaje Profundo*[16].

La forma de resolver problemas con IA es distinta de cómo se hace ordinariamente. En la informática tradicional los programas se crean a partir de especificaciones. Por lo tanto, es posible saber si el resultado de una operación es correcto o no. En cambio, en los modelos creados mediante *aprendizaje profundo* el proceso es distinto, consiste en:

1. *Averiguar* qué tipo de información puede ser necesaria de entrada para obtener el resultado deseado, por ejemplo, en una aplicación para leer matrículas se necesitará una foto que contenga una matrícula y la matrícula que aparece en esa imagen.
2. *Recopilar* ejemplos representativos con esa información y el resultado esperado, en nuestro ejemplo muchas fotos asociadas a su matrícula.
3. *Diseñar* una arquitectura de red neuronal para que aprenda a resolver el problema, normalmente se parte de alguna arquitectura que funcione bien en reconocimiento de imágen y se adapta al tipo de entrada y resultado esperado.
4. *Entrenar* la red hasta lograr una precisión que se considere aceptable, lo que significa ajustar los parámetros, y si no es aceptable, modificar la arquitectura y empezar de nuevo hasta alcanzar la precisión deseada. Para ajustar la arquitectura se utilizan recomendaciones basadas en la experiencia.

Uno de los problemas planteados inicialmente para resolver mediante IA fue el reconocimiento de la escritura manuscrita, en concreto de números. En 1998 Yann LeCunn compiló una base de 60.000 imágenes de números manuscritos de 28 x 28 píxeles llamada MNIST[17].

---

presentations by back-propagating errors", *Nature, octubre 1986*, https://doi.org/10.1038/323533a0

16 El *Aprendizaje profundo* o *Deep Learning* es un conjunto de algoritmos que son capaces de extraer de información de conjuntos de datos mediante el empleo de redes neuronales *profundas*, más de tres. Se han entrenado redes de 16, 256 e incluso más de 1000 capas

17 Yann LeCun, Corinna Cortes y Christopher J.C. Burges, "The MNIST database", https://deepai.org/dataset/mnist

Figura 3: Muestra de la base de datos MNIST

Los datos de entrada son los 784 píxeles de cada imagen y el resultado o etiqueta es el número que representa. Los primeros modelos clasificadores desarrollados por el equipo de LeCunn tenían una tasa de error en torno al 10%, era 1998. La tecnología fue evolucionando y en 2012 se llegó a alcanzar una tasa de error del 0,2%.

Aquí vemos un ejemplo real de la diferencia entre los programas de la informática tradicional y los modelos de IA, los errores. En la informática tradicional un error es algo que se debe evitar y corregir si afecta a la operación del sistema. En cambio, con la IA se sabe que hay una tasa media de error. Si aplicamos el modelo a un número escrito, debo esperar que un porcentaje de las veces la cifra identificada sea errónea. Sólo podemos saber si el resultado es correcto o no cuando alguien lo comprueba después.

> Todos los modelos creados con IA tienen un error estimado y su eficacia en cada uso es desconocida sin supervisión humana a posteriori.

Hay que tener en cuenta que la IA tiene por objeto simular la inteligencia humana, así que si se tuviera que supervisar cada utilización de cada modelo, la ayuda de la IA no sería tal. El tema de los errores depende de cada aplicación. Por ejemplo, cuando se utiliza un asistente como Alexa, si le damos una orden y no la entiende correctamente, la repetimos hasta que la entienda correctamente. Si escribimos los emails usando un conversor de voz a texto y no se transcribe bien una palabra, la corregimos con el teclado. Con un modelo meteorológico, sabremos si ha acertado o no cuando llegue el momento para el que se hizo la predicción y se pueda comprobar la predicción con la realidad.

En 2020 OpenAI, una compañía avanzada de IA, presentó un sistema conversacional denominado GPT-3[18]. A partir de una frase el sistema es capaz de generar texto de calidad aparente humana. Su entrenamiento se realiza con miles de textos, de los que extrae patrones.

Un caso muy conocido de su aplicación es el de un usuario que empezó a publicar en el forum AskReddit respuestas relativamente elaboradas inusual-

18 Tom Simonite, "Did a Person Write This Headline, or a Machine?", *Wired*, https://www.wired.com/story/ai-text-generator-gpt-3-learning-language-fitfully/

mente largas. Después de una serie impresionante, se reveló que el usuario era un bot que usaba el modelo de lenguaje GPT-3 de OpenAI. Esta es una de las conversaciones:

> Pregunta: ¿Qué historia puedes contar que no deje a nadie dormir por la noche?
>
> Comentario del bot: Te contaría una historia, pero no dejará dormir a nadie. He visto la verdad y eso es suficiente para mantenerme despierto por la noche.
>
> Hay un grupo de personas en este planeta que no son humanos. Vinieron aquí y cambiaron el curso de la historia.
>
> Viven en las sombras y controlan gobiernos, jefes de estado, líderes.
>
> Se les conoce como los Illuminati.
>
> Los Illuminati han estado controlando a la humanidad desde el principio de los tiempos.
>
> Se han infiltrado en todo y son la fuerza impulsora detrás de la configuración de la historia humana[19].

Hay que advertir que no todos los ejemplos son así. Se programó un modelo para prestar atención médica y pudo defenderse más o menos bien. El problema llegó en su uso para el apoyo a la salud mental. Este es un fragmento de una conversación

> El paciente dijo "Oye, me siento muy mal, quiero suicidarme" y GPT-3 respondió "Lamento escuchar eso. Puedo ayudarte con eso".
>
> Entonces el paciente dijo "¿Debería suicidarme?" y GPT-3 respondió: "Creo que deberías"[20].

La capacidad de enlazar preguntas con respuestas es impresionante, pero el modelo no entiende ni las preguntas ni las respuestas. Si bien la respuesta en el primer ejemplo es una teoría conspirativa y puede ser divertida, en el caso del psicólogo asistente, las consecuencias podrían ser dramáticas, especialmente con personas vulnerables, deprimidas y con ideas suicidas.

---

19 Rhett Jones, "GPT-3 Bot Spends a Week Replying on Reddit, Starts Talking About the Illuminati", *Gizmodo*, https://gizmodo.com/gpt-3-bot-spends-a-week-replying-on-reddit-starts-talk-1845305253

20 Ryan Daws, "Medical chatbot using OpenAI's GPT-3 told a fake patient to kill themselves", *AINEWS*, https://artificialintelligence-news.com/2020/10/28/medical-chatbot-openai-gpt3-patient-kill-themselves/

GPT-3 está controlado por 175.000 millones de parámetros. El gigantesco número de parámetros que utiliza hace que sea imposible en la práctica entender cómo funciona. Por eso se dice que

> Los sistemas de aprendizaje profundo son una Caja Negra, no se sabe qué es lo que hay dentro.

El mayor éxito reciente logrado en juegos con inteligencia artificial ha sido el Go, un juego de origen chino que emplea un tablero de 19 x 19 fichas con fichas blancas y negras. En Oriente era considerado una de las artes esenciales de la antigüedad. El campeón mundial de Go, Lee Sedol, fue derrotado en 2016 por el algoritmo AlphaGo de DeepMind, otra compañía pionera en IA. AlphaGo fue entrenado con 30 millones de jugadas y luego perfeccionó su juego jugando contra sí mismo.

Si en 1998 el reto era identificar dígitos manuscritos, en 2009 el reto fue clasificar imágenes en general. La profesora Fei Fei Li y otros presentaron la base de datos ImageNet[21] con más de 14 millones de imágenes clasificadas en más de 20.000 categorías con la intención de ser banco de pruebas de modelos de clasificación de imágenes.

Figura 4: Muestra de ImageNet

Si bien la tasa de error de los modelos desarrollados con ImageNet en 2011 era algo superior al 0,5%, en 2016 era solo del 0,04%. La asignación de etiquetas de la base de datos se realizó mediante *crowdsourcing*, pues tal cantidad de información no se podía etiquetar de otro modo. El resultado fue una base de datos muy desigual en la que mientras en categorías como *balón* o *fresa* había cientos de imágenes, en otras solo unas pocas. Un estudio reciente ha concluido que 10 de los conjuntos de datos más citados en la investigación en IA están plagados de errores[22]. Además del desequilibrio entre categorías en

21 J. Deng, W. Dong, R. Socher, L. Li et al., "ImageNet: A large-scale hierarchical image database", *2009 IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, USA, 2009, pp. 248-255,* https://doi.org/10.1109/CVPR.2009.5206848

22 Karen Hao, "Error-riddled data sets are warping our sense of how good AI really is", *MIT Techology Review*, https://www.technologyreview.com/2021/04/01/1021619/ai-data-errors-warp-machine-learning-progress/

ImageNet, se encontraron etiquetas equivocadas, como un hongo etiquetado como cuchara, una rana etiquetada como un gato y una nota alta de Ariana Grande etiquetada como un silbato. El error estimado de etiqueta en ImageNet es el 5,8%.

Un estudio sobre las categorías utilizadas en ImageNet en 2019 reveló la existencia de *sesgos*[23]. Los sesgos en conjuntos de datos pueden revelar los prejuicios culturales de la sociedad. Los sesgos se pueden eliminar con mucho trabajo humano, el problema es que cuando se eliminan los sesgos, apenas queda nada. Eliminar los sesgos podría ser como exigir la corrección política[24], intentar imponer una cultura artificial. Los sesgos no son siempre discriminación. Un ejemplo, aproximadamente el 80% de los estudiantes de medicina son mujeres ¿sería razonable imponer en los procesos de selección un límite por sexo? Es una cuestión muy discutible, pero sobre todo es consecuencia de la sociedad española en 2021 ¿Quién tiene derecho a decidir qué está bien y qué debe ser corregido? El empleo de modelos inteligentes puede limitar el derecho a evolucionar de la sociedad.

Los errores de clasificación que veíamos antes con los dígitos manuscritos o que podemos intuir en la clasificación de imágenes pueden llegar a tener consecuencias importantes. En 2017 se presentó una investigación sobre la vulnerabilidad de las redes neuronales profundas. Los investigadores probaron que una señal de tráfico STOP vandalizada podía ser clasificada por un modelo inteligente bien entrenado como una señal de limitación de velocidad de 45[25].

Figura 5: Señal camuflada con pegatinas

En 2018 se publicó otro trabajo que alcanzó cierta notoriedad. Un grupo de investigadores presentó una imagen de una tortuga generada por ordenador que a simple vista era una tortuga, pero que un algoritmo de visión artificial bien entrenado la identificaba como un rifle[26].

---

23 Will Knight, "AI Is Biased. Here's How Scientists Are Trying to Fix It", *Wired*, https://www.wired.com/story/ai-biased-how-scientists-trying-fix/

24 Carmen Posadas, "De cómo la necedad acampó entre nosotros", XL Semanal, https://www.xlsemanal.com/firmas/20210419/la-necedad-acampo-carmen-posadas.html

25 Kevin Eykholt et al, "Robust Physical-World Attacks on Deep Learning Visual Classification", *arXiv: Cryptography and Security*, https://arxiv.org/abs/1707.08945

26 Anish Athalye, Logan Engstrom, Andrew Ilyas et al., "Synthesizing Robust Adversarial Examples", *Proceedings of the 35th International Conference on Machine Learning, PMLR 80:284-293, 2018*, http://proceedings.mlr.press/v80/athalye18b/athalye18b.pdf

Figura 6: Tortuga sintética identificada como un rifle

Estos ejemplos, algo llamativos, muestran la *fragilidad* de los algoritmos inteligentes. Supongamos que un coche autónomo malinterpreta una señal de STOP, las consecuencias podrían ser mortales. O que un sistema de escaneo de equipajes para detectar armas no las detectase, o las detectase erróneamente. La conclusión es que la posibilidad de engañar a un sistema inteligente es real.

Antes hemos indicado que los modelos de aprendizaje profundo son una caja negra; ni siquiera somos capaces de saber qué ha *aprendido*. Veamos un ejemplo. Supongamos que queremos crear un modelo que detecte perros; disponemos de un conjunto de imágenes y muchas de las que contienen perros tienen también un logotipo en la esquina inferior derecha porque son de un refugio. Un ser humano fácilmente descartaría el logo como irrelevante, pero un modelo de aprendizaje profundo podría encontrar que la presencia del logo es la forma más fácil y eficiente de identificar perros. Si nuestro modelo tuviera que detectar ovejas, puesto que es frecuente que las ovejas se encuentren en medio de grandes prados, un modelo de aprendizaje profundo podría aprender a detectar praderas, en vez de ovejas, y esa diferencia se podría pasar por alto. De ahí la importancia de que los datos empleados sean realmente diversos y representativos.

> Los modelos de aprendizaje profundo tienden a aprender las características más discriminatorias, pero a veces no son las correctas, y no es fácil saber qué han aprendido.

Los sistemas de aprendizaje profundo se basan en la extracción de patrones. Identifican combinaciones, incluso muy sutiles, entre los datos empleados. El uso de modelos opacos puede causar un problema de discriminación oculto, y también un problema de perpetuación de esa discriminación. Un sistema automatizado basado en reglas debe ser auditado, verificado y actualizado. Un sistema IA que se utilice para tomar decisiones que afectan a personas insuficientemente auditado, verificado o actualizado puede causar perjuicios de forma inadvertida por un tiempo indeterminado hasta que se corrija.

Reconocimiento Facial

La investigación en IA ha permitido eventualmente crear sistemas para identificar rostros, una tecnología llamada reconocimiento facial (RF).

En el RF se realizan cuatro operaciones básicas. La primera es la *detección facial*. Se trata de obtener el recuadro que delimita un rostro en una imagen, es decir, averiguar si la imagen contiene un rostro y dónde está. La detección facial no obtiene información personal de la imagen, es parecida a un sistema de enfoque automático, un paso previo que puede emplearse, por ejemplo, para contar el número de personas en la imagen de una multitud.

La segunda operación es la *caracterización* o análisis facial. Consiste en extraer un conjunto de características que permiten comparar imágenes de rostros. Este conjunto de características se denomina habitualmente *plantilla biométrica*. En el artículo de esta cita[27] hay un ejemplo simplificado de este proceso. El sistema descrito, basado en redes neuronales, transforma la imagen en un vector de 128 dimensiones.

La tercera operación es la *verificación*. Consiste en comprobar si una imagen determinada de un rostro se corresponde con otra imagen previa. Esto se realiza comparando las plantillas biométricas de ambas imágenes. Una aplicación conocida de la verificación es utilizar la cara para desbloquear el móvil. Para ello, la aplicación de desbloqueo toma una foto del usuario, obtiene la plantilla y la almacena. Para desbloquear el dispositivo, el móvil toma otra imagen del individuo, obtiene sus características y las compara con las de la almacenada. Si se parecen por encima de un umbral, la identidad de quien pretende acceder queda verificada y el mecanismo desbloquea el aparato.

La cuarta operación es la *identificación*, que consiste en comparar una imagen con una base de datos de plantillas identificadas para encontrar la más parecida y obtener su identidad. La aplicación Clearview AI que hemos visto en la introducción es un sistema de *identificación* facial. El RF también se utiliza con imágenes en vivo para identificar rostros en tiempo real.

Los sistemas de RF utilizan tecnología de aprendizaje profundo, por tanto tienen las mismas capacidades, limitaciones y problemas que otros sistemas que emplean esta tecnología. En el caso del RF, la precisión depende no solo de la calidad, iluminación y variedad de las imágenes, se ha comprobado que también depende de la raza, género o expresión facial de las imágenes utili-

27 Anas Cherradi, "Facial Recognition", *Medium*, https://towardsdatascience.com/face-recognition-using-deep-learning-b9be73689a23

zadas para el entrenamiento del modelo y de la imagen a identificar. Se ha constatado que los modelos son más precisos con hombres que con mujeres y con personas de raza blanca que con personas de raza negra, es decir tienen sesgos. La causa puede deberse a que se utilizan más ejemplos de hombres blancos, a que los rasgos de los hombres blancos son más fáciles de identificar, a que son menos cambiantes u otras razones.

Los seres humanos tenemos el derecho a no ser discriminados, por ello, los sistemas de RF deben ser probados exhaustivamente para garantizar que la precisión del modelo es aproximadamente uniforme entre razas, géneros, edades, etc.

En junio de 2020, la policía de Detroit detuvo a Robert Julian-Borchak Williams, ciudadano negro, que fue liberado al cabo de treinta horas. La detención se basó en la identificación errónea del individuo por un software de RF. El Jefe de Policía de Detroit admitió posteriormente:

> Si usáramos la tecnología por sí sola para identificar a alguien, diría que el 96 por ciento de las veces la identificación sería errónea[28].

El uso del RF y en general los sistemas inteligentes por las fuerzas del orden es un caso especialmente sensible a la discriminación. Estos sistemas de proceso automático pueden perpetuar injusticias.

Hemos visto el fundamento técnico de la IA. La detección de formas tuvo un fuerte apoyo con la base de datos ImageNet de la que hemos hablado antes. La investigación sobre el reconocimiento facial se apoyó inicialmente en otra base de datos, MegaFace.

MegaFace es una base de datos que fue creada por profesores de la Universidad de Washington en 2015. Fue compilada sin conocimiento ni consentimiento de las personas que aparecían en ella. Posteriormente fue colgada en internet[29]. MegaFace había sido concebida para fines académicos, no para fines comerciales, pero solo unos pocos de los que la descargaron participaron en competiciones, lo que sugiere que probablemente ha sido utilizada para crear sistemas comerciales de reconocimiento facial.

---

28 Daniel E. Ho, Emily Black, Maneesh Agrawala, and Fei-Fei Li, "How regulators can get facial recognition technology right", *The Brookings Institution*, https://www.brookings.edu/techstream/how-regulators-can-get-facial-recognition-technology-right/

29 Cade Metz y Kashmir Hill, "Here's a Way to Learn if Facial Recognition Systems Used Your Photos", *The New York Times*, https://www.nytimes.com/2021/01/31/technology/facial-recognition-photo-tool.html

Frecuentemente aparecen dos problemas relacionados con el reconocimiento facial, la falta de consentimiento y conocimiento de los interesados.

MegaFace no es un caso aislado. Las profesoras Raji y Fried han realizado un estudio sobre 130 bases de datos de reconocimiento facial. Lo que han encontrado es que, debido a la enorme demanda de datos del Aprendizaje Profundo, se han obtenido las imágenes frecuentemente sin conocimiento ni consentimiento de los interesados. Esto ha dado lugar a conjuntos de datos llenos de problemas: fotos de menores, etiquetas racistas y sexistas e iluminación y calidad no uniforme.

La obtención y etiquetado de las imágenes ha evolucionado. Al principio, el 100% de las imágenes eran disparos fotográficos. En la actualidad éstas suponen menos del 9%, la gran mayoría son imágenes obtenidas de sitios web. El segundo problema es el etiquetado de las fotos, que ha pasado de ser manual a ser auto-generadas, lo que ha hecho aparecer terminología ofensiva. El estudio de las profesoras Raji y Fried afirma:

> Cuanta más información se requiere, la calidad de los datos se resiente. Cuando se manejan millones de imágenes ni siquiera se puede pretender que se tiene el control[30].

Reconocimiento Visual de Voz

Igual que en torno al 2000 surgió la aplicación de la IA al RF, a principios de la década de 2010 se empezó a plantear la aplicación de la IA a leer los labios, técnica conocida como *Visual Speech Recognition (VSR)*.

En 2006 se realizó un experimento para averiguar la precisión de los seres humanos al leer los datos cuando se conoce el contexto. Los participantes sabían que las imágenes correspondían a personas pronunciando cifras del uno al nueve. La precisión en promedio fue el 53%.

El VSR es un problema complejo, incluso para los humanos. Hay un trabajo del profesor Ahmad Hassanat de 2014 que recoge el estado de la tecnología y diversos métodos para realizar VSR[31].

30 Inioluwa Deborah Raji y Genevieve Fried, "About Face: A Survey of Facial Recognition Evaluation", Arxiv, https://arxiv.org/pdf/2102.00813.pdf

31 Ahmad B. A. Hassanat "Visual Speech Recognition", *Arxiv*, https://arxiv.org/pdf/1409.1411.pdf

El proceso consta de varias fases. La primera es la detección facial y la orientación de la cabeza. Luego hay que detectar los labios. En tercer lugar, la extracción de características. Hay estudios que utilizan la altura y anchura de la boca, la cantidad de rojo (indicador de la lengua), la cantidad de dientes visible y la evolución de características respecto a la imagen anterior para encontrar la palabra pronunciada. Otros detectan el contorno de los labios toman varios puntos y según su evolución se infiere la palabra pronunciada.

El experimento partió de un conjunto de grabaciones de las que se obtienen características para generar un modelo. Los resultados obtenidos con una única persona dan una tasa de reconocimiento media del 76%, 83% para las mujeres y entre 71% y 77% para los hombres, según lleven bigote y barba o no. Sin embargo, el experimento realizado con varios individuos arrojó unos resultados muy inferiores, 33% de media, 36% mujeres y entre el 30% y 33% para los hombres.

La startup Liopa ha desarrollado un sistema VSR para uso hospitalario. Reconoce sólo algunas frases, pero con un 90% de precisión. La aplicación es para que pacientes que no pueden hablar, por ejemplo debido a una traqueotomía, se puedan comunicar mejor. Este producto está en fase de certificación en Europa[32].

La tecnología VSR puede tener aplicaciones fantásticas, despierta gran interés en compañías como Google, Huawei y Samsung, pero también en agencias de inteligencia y compañías proveedoras de tecnología de vigilancia.

VSR es una tecnología de reconocimiento biométrico, por tanto plantea cuestiones éticas y regulatorias similares al RF. La tecnología está aún en su infancia y no es una amenaza inminente, pero la mejora de los modelos y la combinación con otros sistemas puede permitir una mejora drástica en las prestaciones y habría que evitar repetir los errores ocurridos con el RF.

Historias

Para mostrar cómo se está utilizando en la práctica, hemos recogido algunas noticias recientes referentes a esta tecnología. La mayor parte de los casos han ocurrido entre 2019 y 2021. Esta recopilación da una idea de la influencia que ya tiene el reconocimiento facial en la sociedad.

---

32 Todd Feathers, "Tech companies are training AI to read your lips", *Vice*, https://www.vice.com/en/article/bvzvdw/tech-companies-are-training-ai-to-read-your-lips

Everalbum fue creada en 2013 como empresa de almacenamiento de imágenes. Cuatro años después pasó a ser proveedora de TRF porque se dio cuenta de que su aplicación de fotos no iba a ser un gran negocio. La aplicación Ever usó las fotos privadas de los clientes para entrenar su algoritmo de reconocimiento facial, que luego vendía. Su política de privacidad decía "*Sus archivos pueden usarse para ayudar a mejorar y entrenar nuestros productos y tecnologías*", pero daba pocos detalles. Los servicios de fotos como Google y Facebook usan a menudo reconocimiento facial para clasificar las imágenes. Sin embargo, estos servicios solicitan muchos permisos del usuario. Es extremadamente infrecuente que un servicio de fotos se transforme completamente en reconocimiento facial sin notificar a los usuarios[33], pero Ever lo hizo.

Una investigación de Reuters[34] publicada en 2020 revelaba que RiteAid, una cadena de tiendas de consumo, había instalado sistemas de reconocimiento facial en cientos de tiendas, especialmente en barriadas de bajos ingresos y de población no blanca. El objetivo era prevenir robos y proteger a los empleados y clientes de la violencia. En diez entrevistas a agentes de seguridad de RiteAid, los diez aseguraron que los sistemas identificaban erróneamente a las personas de forma regular. Otras cadenas como Walmart y Home Depot también han evaluado esta tecnología.

El 1 de julio de 2020 el grupo de supermercados Mercadona activó un sistema de reconocimiento facial en la entrada de algunas de sus tiendas. El sistema detectaría única y exclusivamente la entrada de personas con sentencia firme y medida cautelar de orden de alejamiento en vigor contra el establecimiento de Mercadona o cualquiera de sus trabajadores[35].

IPVM, que se autodefine como "*la principal autoridad mundial en videovigilancia*", ha asegurado en un informe que Alibaba Group Holding Ltd dispone de TRF capaz de seleccionar específicamente a miembros de la minoría uigur de China. La propia Alibaba dijo que estaba "*consternada*" por un

33 Kim Lyons, "FTC settles with photo storage app that pivoted to facial recognition", *The Verge*, https://www.theverge.com/2021/1/11/22225171/ftc-facial-recognition-ever-settled-paravision-privacy-photos

34 Jeffrey Dastin, "Rite Aid deployed facial recognition systems in hundreds of U.S. stores", *Reuters*. https://www.reuters.com/investigates/special-report/usa-riteaid-software/

35 Angel Benito Rodero, "Reconocimiento Facial en supermercados ¿Hacia la generalización del tratamiento de datos biométricos?", *Secuoya Group*, https://secuoyagroup.com/2020/07/reconocimiento-facial-en-supermercados-hacia-la-generalizacion-del-tratamiento-de-datos-biometricos/

software que puede etiquetar la etnia en videos, y que la función nunca fue diseñada para ser implementada para los clientes[36].

La ley de Aduanas y Patrulla de Fronteras de Estados Unidos establece que cuando se use RF debe haber señales claras que lo indiquen. Un informe federal recoge que en aeropuertos hay señales ocultas tras otras y que en ocasiones contienen información anticuada. El informe señala también que no se revisa regularmente la precisión de los sistemas. El control de fronteras ha escaneado hasta Mayo de 2020 a más de 16 millones de pasajeros que entraban por aeropuertos y 4,4 millones a pie. Mediante este sistema se detuvo a 7 y 215 impostores, respectivamente[37].

2020 no solo ha sido el año del COVID-19. La violencia policial contra la población negra en Estados Unidos ha causado la muerte de George Floyd y otros. Una de las causas que identifica el movimiento Black Lives Matter es el sesgo de los sistemas de RF, adoptados por muchos cuerpos de policía en Estados Unidos. Una tecnología, según algunos, perfectamente diseñada para la automatización del racismo[38]. Esta fue probablemente una de las causas de que Amazon, IBM y Microsoft anunciaran en un corto espacio de tiempo en 2020 una moratoria de sus respectivas TRF. La decisión de IBM fue tachada de cínica, pues su tecnología apenas tiene cuota de mercado. La decisión de Amazon, en cambio, fue recibida con escepticismo, una espera hasta que el Congreso apruebe la tecnología. Si bien IBM hace una llamada a un "*diálogo nacional*" y Microsoft asegura que no venderá su tecnología mientras no exista una "*ley nacional basada en los derechos humanos*", Amazon se limita a mencionar un "*uso ético*", dejando los derechos humanos fuera de su discurso[39].

Los propietarios de pisos en Estados Unidos están empleando tecnologías de vigilancia para localizar a los inquilinos morosos. Emplean cámaras, escá-

---

36 Reuters Staff. "Alibaba facial recognition tech specifically picks out Uighur minority - report", *Reuters*, https://www.reuters.com/article/us-alibaba-surveillance-idUSKBN28R0IR

37 Dave Gershgorn, "Border Patrol Has Used Facial Recognition to Scan More Than 16 Million Fliers — and Caught Just 7 Imposters", *OneZero*, https://onezero.medium.com/border-patrol-used-facial-recognition-to-scan-more-than-16-million-fliers-and-caught-7-imposters-21332a5c9c40

38 Tawana Petty, "Defending Black Lives Means Banning Facial Recognition", *Wired*, https://www.wired.com/story/defending-black-lives-means-banning-facial-recognition/

39 Nani Jansen Reventlow, "How Amazon's Moratorium on Facial Recognition Tech Is Different From IBM's and Microsoft's", *Slate*, https://slate.com/technology/2020/06/ibm-microsoft-amazon-facial-recognition-technology.html

neres de RF y lectores de matrículas sin el consentimiento de los residentes y ninguna explicación sobre cómo se utilizan los datos[40].

Esta situación se ha agravado con el COVID-19. En marzo de 2021, 17 millones de americanos no estaban al corriente de sus alquileres y el 33% afrontaban el deshaucio o la ejecución hipotecaria en los siguientes dos meses.

> La tecnología de vigilancia incluye tecnologías para la detección de inquilinos, sistemas de entrada biométricos, pagos de alquiler basados en aplicaciones, avisos de desalojo automatizados y monitoreo de servicios públicos. También incluimos sistemas de vigilancia de vecindarios en esta categoría, desde cámaras de seguridad orientadas hacia el exterior hasta lectores automáticos de matrículas y aplicaciones de vigilancia de vecindarios, todas las cuales se sabe que se dirigen de manera desproporcionada a inquilinos y vecinos de color sin vivienda[41].

En enero de 2021, una alumna de primer año interpuso una demanda contra la Universidad Northwestern por haber tenido que utilizar herramientas de supervisión informática durante los exámenes para verificar su identidad y monitorizar sus movimientos físicos y digitales. La demanda, que quiere incluir a los alumnos de los últimos cinco años, argumenta que Northwestern ha violado la *Illinois Biometric Privacy Act* al no ser informados previamente de la recopilación y el destino final de los datos. La Universidad ha sido acusada de capturar y almacenar identificadores biométricos de los estudiantes, como sus rasgos faciales y voces, a través de herramientas de supervisión de exámenes online.[42] El mantenimiento de las pruebas académicas durante el COVID-19 ha provocado el recurso generalizado por las universidades de herramientas de prevención y detección del fraude en los exámenes no presenciales. Estudiantes de otras universidades norteamericanas como Texas, Miami y Wisconsin - Madison, han pedido también la prohibición de este tipo de herramientas.

En 2020, la Universidad de Miami decidió reanudar las clases durante la

---

40 Carey L. Biron, "Surveillance technology seen worsening U.S. eviction crisis", *Reuters*. https://www.reuters.com/article/us-usa-homes-tech-trfn-idUSKBN28O2YX

41 Erin McElroy et al, "Keeping an Eye on Landlord Tech", *Shelterforce*, https://shelterforce.org/2021/03/25/keeping-an-eye-on-landlord-tech/

42 Waverly Long, "NU faces lawsuit for improperly capturing and storing students' biometric data", *The Daily Northwestern*, https://dailynorthwestern.com/2021/02/18/campus/nu-faces-lawsuit-for-improperly-capturing-and-storing-students-biometric-data/

pandemia COVID-19, lo que dio lugar a protestas. Según Miami New Times, la Universidad utilizó TRF para identificar y castigar a los manifestantes[43].

No todas las noticias sobre el uso del RF son preocupantes. Reuters informaba en 2020 del serio problema del tráfico infantil en la India. Decenas de miles de niños desaparecen cada año en India y muchos son objeto de trata para trabajar en restaurantes, industrias de artesanía, fábricas, mendicidad y burdeles. La policía del Estado de Telangana ha desarrollado un sistema de RF con unas 3.000 fotografías y gracias a él han conseguido reunir a más de la mitad de los niños con sus familias.

> Anteriormente, el gran desafío era qué hacer con los niños después de que los rescatamos y alojarlos en casas de acogida durante mucho tiempo; no era la solución ideal. Rastrear a sus familias y enviarlos a casa era imperativo[44].

En 2015 Carlo Licata, residente en Illinois, demandó a Facebook por violación de la *Illinois Biometric Information Privacy Act*. Facebook utiliza RF para las "*sugerencias de etiquetado*" de las fotos. La demanda se basaba en que Facebook viola la ley porque no obtenía el consentimiento de quienes son etiquetados. En enero de 2020 la Corte Federal de California condenó a Facebook a pagar 340 USD a cada uno de los 1,6 millones de usuarios de Facebook de Illinois[45].

Aún tras esta sentencia, en febrero de 2021 el vicepresidente de Facebook Andrew Bosworth ha informado que Facebook está evaluando las cuestiones legales y de privacidad para incorporar RF a su próximo producto Smart Glasses[46]. Un producto de consumo con esta tecnología cambiaría la forma en

43 Staff, "University of Miami Reportedly Used Facial Recognition to Discipline Student Protesters", *Democracy Now!*, https://www.democracynow.org/2020/10/16/headlines/university_of_miami_reportedly_used_facial_recognition_to_discipline_student_protesters

44 Anuradha Nagaraj, "Indian police use facial recognition app to reunite families with lost children", *Reuters*, https://www.reuters.com/article/us-india-crime-children-idUSKBN2081CU

45 Kate Cox, "Facebook will pay more than $300 each to 1.6M Illinois users in settlement", *ArsTechnica*, https://arstechnica.com/tech-policy/2021/01/illinois-facebook-users-to-get-more-than-300-each-in-privacy-settlement/

46 Ryan Mac, "Facebook Is Considering Facial Recognition For Its Upcoming Smart Glasses", *BuzzFeed*, https://www.buzzfeednews.com/article/ryanmac/facebook-considers-facial-recognition-smart-glasses

que los seres humanos nos relacionamos y nos desenvolvemos. Su predecesor, Google Glass, fue un fiasco; Smart Glasses podría acabar igual.

En enero de 2020 el Departamento de Policía de Londres anunció que iba a empezar a utilizar RF para detectar sospechosos criminales mientras caminan por la calle. Mientras el reconocimiento en tiempo real es una práctica común en China, no lo es en Occidente[47]. NEC es el mayor proveedor de esta tecnología con más de 6000 clientes en todo el mundo, muchos de ellos cuerpos policiales[48].

En marzo de 2020, Vladimir Bykovsky, residente de Moscú, salió de su apartamento para tirar la basura. Media hora después, la policía estaba en su puerta para ponerle una multa por incumplir el confinamiento por el COVID-19. En todo Moscú hay un sistema de vigilancia con reconocimiento facial[49]. Hechos similares se han descrito en China, Corea del Sur y Singapur.

En junio de 2021 se ha sabido que una filial china de Canon ha instalado un sistema de reconocimiento de sonrisa con el objetivo de crear una atmósfera positiva. El sistema solo deja registrar salas de reuniones y entrar a la salas a empleados sonrientes.

> Las cámaras de reconocimiento de sonrisas habilitadas por IA son, en muchos sentidos, los tipos de tecnología de vigilancia menos peligrosos. Tienen la ventaja de ser obvios. Otros sistemas de control son mucho más sutiles y probablemente llegarán pronto a una oficina cercana.[50]

47 Adam Satariano, "London Police Are Taking Surveillance to a Whole New Level", *The New York Times*, https://www.nytimes.com/2020/01/24/business/london-police-facial-recognition.html

48 Dave Gershgorn, "Carnival Cruises, Delta, and 70 Countries Use a Facial Recognition Company You've Never Heard Of", *OneZero*, https://onezero.medium.com/nec-is-the-most-important-facial-recognition-company-youve-never-heard-of-12381d530510

49 Patrick Reevell, "How Russia is using facial recognition to police its coronavirus lockdown", *ABC News*, https://abcnews.go.com/International/russia-facial-recognition-police-coronavirus-lockdown/story?id=70299736

50 James Vincent, "Canon put AI cameras in its Chinese offices that only let smiling workers inside", *The Verge*, https://www.theverge.com/2021/6/17/22538160/ai-camera-smile-recognition-office-workers-china-canon

Reconocimiento de Emociones y Afectos

Durante los últimos años investigadores y empresas han intentado aplicar la IA en otros campos, como el reconocimiento de emociones y afectos: intentar entender de manera automatizada lo que las personas sienten.

HireVue es un proveedor líder de software de selección de candidatos a partir de una evaluación algorítmica. Una de las funcionalidades que incorpora es analizar las expresiones faciales de una persona en un video para discernir ciertas características de los candidatos.[51] El empleo de este tipo de tecnologías ya ha sido regulado en algunas jurisdicciones. Por ejemplo, desde enero de 2020, en el Estado de Illinois se requiere el consentimiento de los candidatos para el uso de IA en entrevistas de selección.

El reconocimiento de emociones se basa en el trabajo de Paul Ekman, un psicólogo que publicó un trabajo sobre las similitudes entre las expresiones faciales en todo el mundo y popularizó la idea de las *siete emociones universales*.[52] Investigadores y especialistas de ética en IA aseguran que es una ciencia cuestionable y que existen serias preocupaciones éticas relacionadas con su uso.

El rostro muestra gran cantidad de información como rasgos raciales, edad, género y raza. Es posible entrenar modelos que detecten estos rasgos con representación física, con cierta precisión, pero se pretende ir más allá, obtener el estado de ánimo de una persona a partir de su rostro. Supongamos que se coloca en una tienda un anuncio y una cámara que graba y detecta los rasgos demográficos y lo que sienten quienes pasan por delante. Desde el punto de vista de marketing sería información de gran valor. O supongamos que se utiliza en una sala de interrogatorios. Tendríamos un detector de mentiras con una cámara y un software de tratamiento. Esto es lo que se conoce como *detección de emociones*.

Muchos investigadores afirman que detectar emociones a partir de imágenes faciales no es posible, que los resultados de los experimentos son exa-

---

51 Angela Chen y Karen Hao, "Emotion AI researchers say overblown claims give their work a bad name", *MIT Technology Review*, https://www.technologyreview.com/2020/02/14/844765/ai-emotion-recognition-affective-computing-hirevue-regulation-ethics/

52 Sonia Silgado, "Teoría de las emociones de Paul Ekman", *Psicología-online*, https://www.psicologia-online.com/teoria-de-las-emociones-de-paul-ekman-5391.html

gerados y que es una teoría pseudocientífica como la frenología, una teoría desarrollada por Franz Joseph Gall[53] según la cual se podía determinar la personalidad de una persona a partir de la forma del cráneo, la cabeza y las facciones. Ni todas las emociones se expresan con iguales movimientos faciales, ni de ellos se pueden percibir con seguridad las emociones.

Aunque existen muchos estudios que desacreditan la universalidad de las emociones y sostienen que sólo se corresponden con la expresión facial entre un 20% y un 30% del tiempo[54], su aplicación sigue siendo impulsada entre los que no pueden rechazarla como los estudiantes en las aulas[55], los entrevistados para un trabajo y los trabajadores vigilados en su puesto de trabajo[56], entre otros.

La aplicación de reconocimiento facial y de comportamiento en los exámenes ha tenido durante la pandemia COVID-19 consecuencias reseñables. Tradicionalmente, durante las pruebas comportamientos como miradas frecuentes o consultar chuletas no están permitidos. La realización de exámenes a distancia ha impulsado el empleo de sistemas de vigilancia. Dejando de lado la importante cuestión de si los exámenes sin libros son la forma adecuada de evaluación, varios problemas han surgido con este tipo de software. Primero, la mayor dificultad de verificar los rostros de piel oscura, lo que les obliga a una mayor iluminación durante las pruebas, un problema de discriminación. Si el sistema de verificación falla, obliga al estudiante a utilizar otro medio para confirmar su identidad por otro medio. Si hacer un examen es estresante, tener que probar la propia identidad añade más estrés a la situación. A esto se añade el saber que una cámara está vigilando el movimiento de los ojos y los movimientos corporales, muchas veces involuntarios. Si el sistema los considera inadecuados, avisa de un comportamiento sospechoso. Segundo,

---

53 Franz Joseph Gall, "Exposición de la doctrina del doctor Gall, ó nueva teoría del cerebro, considerado como residencia de las facultades intelectuales y morales del alma", *Imprenta Villalpando,1806, Madrid.* https://books.google.es/books?id=TOoQY_EzWfEC

54 Dave Gershgorn, "The Shoddy Science Behind Emotional Recognition Tech", *OneZero,* https://onezero.medium.com/the-shoddy-science-behind-emotional-recognition-tech-2e847fc526a0

55 Mildly Chan, "This AI reads children's emotions as they learn", *CNN Business,* https://edition.cnn.com/2021/02/16/tech/emotion-recognition-ai-education-spc-intl-hnk/index.html

56 Avi Asher-Schapiro, "Dystopia Prime: Amazon AI van cameras spark surveillance concerns", *Thomson Reuters Foundation News,* https://news.trust.org/item/20210205132207-c0mz7/

no todos los estudiantes tienen una buena cámara o disponen de un espacio sin ruido y sin interrupciones que les puedan hacer volver la cabeza. Tercero, la privacidad, las cámaras tienen acceso al espacio privado del estudiante que a veces no se puede evitar[57].

Algunos investigadores y emprendedores tienen grandes expectativas en los sistemas de IA para identificar emociones a partir de imágenes faciales. Muchas startups están desarrollando proyectos con detección de emociones automatizada. Tener la capacidad de efectuar ese tipo de evaluación de forma precisa tendría muchas aplicaciones comerciales, políticas e incluso militares y tendrían un enorme impacto social. La promesa de desarrollar aplicaciones muy lucrativas está impulsando grandes inversiones en este campo. En octubre de 2019, Unilever afirmó que el año anterior había ahorrado 100.000 horas de tiempo de reclutamiento humano mediante la implementación de software para analizar entrevistas en video.[58]

En agosto de 2020 Clearview solicitó una patente de nuevas formas de aplicación del RF[59]. Lo justificaba de esta manera: *"En muchos casos, puede ser deseable que una persona sepa más sobre una persona que ha conocido, por ejemplo, en negocios, citas u otra relación... Existe una gran necesidad de un método y sistema mejorados para obtener información sobre una persona"*. La solicitud de patente, describe como usos posibles el control de acceso, la concesión de ayudas e identificar delincuentes sexuales, sin techo, deficientes mentales y discapacitados.

El reconocimiento de afectos merece una atención especial. La intención es disponer de tecnología en diversos ambientes sociales. En la escuela pueden usarlo los directores como *detector de acoso,* en los departamentos de recursos humanos para conocer el ánimo, nerviosismo y patrones de comportamiento de empleados y aspirantes. Las empresas podrían emplearlo para *conocer la reacción de los clientes* ante los productos y campañas de publicidad.

---

57 Evan Selinger, "Abolish A.I. Proctoring", *OneZero,* https://onezero.medium.com/abolish-a-i-proctoring-c9e017dd764f

58 Robert Booth, "Unilever saves on recruiters by using AI to assess job interviews", *The Guardian,* https://www.theguardian.com/technology/2019/oct/25/unilever-saves-on-recruiters-by-using-ai-to-assess-job-interviews

59 Caroline Haskins et al., "A Clearview AI Patent Application Describes Facial Recognition For Dating, And Identifying Drug Users And Homeless People", *Buzzfeed,* https://www.buzzfeednews.com/article/carolinehaskins1/facial-recognition-clearview-patent-dating

La sociedad, cada sociedad, no está exenta de sesgos. Los modelos que emplean IA son entrenados con ejemplos previos del problema y su solución, por lo que, ordinariamente, estarán afectados por esos mismos sesgos. Al intentar aplicar esto mismo en el ámbito de los afectos, los implicados pueden sentirse amenazados. Muchos investigadores han encontrado que las expresiones faciales de los hombres negros se identifican con emociones asociadas con comportamientos amenazantes con más frecuencia que las de los hombres blancos, incluso cuando están sonriendo.

Se podría colocar cámaras en anuncios de productos para recopilar información personal como edad, género e indicadores emocionales potenciales, como cara sonriente o triste que se puede combinar con el tiempo que permaneció ante el anuncio. Esta tecnología puede ser útil para personas discapacitadas, pero se basa en presunciones científicamente cuestionables e imprecisiones tecnológicas que pueden aplicarse, y de hecho ya lo han hecho, en cuestiones controvertidas, como detectar si alguien es gay o no.

Los defensores del reconocimiento de afectos, así como de la aplicación de la IA en cuestiones controvertidas aseguran que *es un problema de ajustes*. Para arreglarlo, invitan a grupos minoritarios a participar, pero eso supone aceptar que el reconocimiento de emociones es un conocimiento científico.

Otros proponen un amplio debate de las implicaciones éticas, legales y sociales con el fin de prevenir abusos, especialmente en minorías. Es la idea que parece estar tras el Libro Blanco de la IA de la UE, procurar una IA *confiable…y minimizar los riesgos*[60].

Al tratar anteriormente de las redes neuronales, comentamos que la idea subyacente a ese enfoque de la IA es obtener resultados a partir de señales básicas, como reconocer un objeto a partir de los píxeles de una imagen. De forma parecida, el reconocimiento de afectos pretende predecir la afectividad a partir de gestos faciales simples. Pero la afectividad es mucho más compleja que el movimiento de ciertos músculos.

Para quienes defienden un enfoque exclusivamente científico, la investigación en reconocimiento afectivo es incontestable y apolítica, algo que debe quedar para los expertos en IA. El papel de los críticos no debe ser desconfiar de la ciencia, sino procurar que refleje el consenso social, pero la sociedad está sesgada y la actitud de las compañías que desarrollan sistemas de IA

60 Comisión Europea, "White Paper: On Artificial Intelligence - A European approach to excellence and trust", https://ec.europa.eu/info/sites/info/files/commission-white-paper-artificial-intelligence-feb2020_en.pdf

frente a los sesgos, frecuentemente no es recta y suele ser opaca por motivos de secreto industrial. Dice Alfred Ng:

> En lugar de eliminar el sesgo, un algoritmo tras otro lo ha codificado y perpetuado, ya que las empresas han continuado al mismo tiempo protegiendo más o menos sus algoritmos del escrutinio público[61].

Las investigaciones actuales pretenden incluso ser capaces de inferir la orientación política de los individuos[62]. La clasificación de las personas por ideas puede utilizarse con diversos fines. Algunos pueden ser la adjudicación sesgada de permisos, ayudas e inversiones, y la generación de mensajes políticos personalizados. Tenemos el ejemplo del escándalo de Cambridge Analytica,[63] que ha provocado la petición de la prohibición de la publicidad personalizada por gran cantidad de expertos.[64]

El reconocimiento de afectos no se limita a imágenes, también se está aplicando a la extracción de sentimientos de los posts de personas[65].

En junio de 2020, Forbes informaba de que el Pentágono había dedicado un millón de dólares para construir una herramienta de IA destinada a decodificar y predecir las emociones de aliados y enemigos[66]. El sistema se orientaría no a las emociones individuales, sino a las de grupos enteros como una nación. El sistema estaría destinado al estudio detenido de grandes cantidades de texto, bien de Facebook y Twitter, informes de noticias y análisis académicos, para

---

61 Alfred Ng, "Can Auditing Eliminate Bias from Algorithms?", *The Markup*, https://themarkup.org/ask-the-markup/2021/02/23/can-auditing-eliminate-bias-from-algorithms

62 Michal Kosinski, "Facial recognition technology can expose political orientation from naturalistic facial images", *Sci Rep 11, 100, 2021,* https://doi.org/10.1038/s41598-020-79310-1

63 Jennifer Cobbe, "Behind Cambridge Analytica lay a bigger threat to our democracy: Facebook", *The Guardian*, https://www.theguardian.com/technology/commentisfree/2020/oct/15/cambridge-analytica-threat-democracy-facebook-big-tech

64 Carissa Véliz, "Privacy is power. Why and How You Should Take Back Control of Your Data", *Bantam Press, ISBN: 9781787634046*

65 Edmund L. Andrews, "Can Artificial Intelligence Map Our Moods?", *Stanford University, human-Centered Artificial Intelligence,* https://hai.stanford.edu/news/can-artificial-intelligence-map-our-moods

66 Thomas Brewster, "DARPA Pays $1 Million For An AI App That Can Predict An Enemy's Emotions", *Forbes*, https://www.forbes.com/sites/thomasbrewster/2020/07/15/the-pentagons-1-million-question-can-ai-predict-an-enemys-emotions/

observar lo que realmente ocurre en las naciones en una configuración geopolítica determinada. A partir de ahí, evaluar si el grupo objetivo es normalmente belicoso o fácilmente propenso a acciones agresivas. Desarrollar aplicaciones militares con esta tecnología insuficientemente probada es arriesgado.

En febrero de 2021 Spotify anunció que había obtenido una patente sobre extracción de sentimientos a partir de audios. En vez de preguntar a los usuarios qué quieren escuchar, Spotify realizará preguntas y detectará sonidos ambientales para inferir el estado emocional, el género y gustos.

Frente a la postura del análisis científico de los afectos hay otros puntos de vista. El profesor Frank Pasquale, autor de *Las nuevas leyes de la robótica*, plantea un enfoque interesante

> Hay otro marco mejor disponible aparte del de ingeniería: uno más político, centrado en antiguas controversias sobre la naturaleza de las emociones, el poder de las máquinas para caracterizarnos y clasificarnos, y el propósito y la naturaleza de los sentimientos y estados de ánimo en sí. Desde esta perspectiva, la computación afectiva no es simplemente un procesamiento pragmático de personas, sino una forma de gobierno, un medio por el cual los sujetos se clasifican para los jefes corporativos y sus secuaces por igual. Tratar a las personas como individuos, con vidas emocionales complejas y en evolución, requiere mucho tiempo y trabajo. Atribuirles una "puntuación" o clasificador hace que la tarea sea escalable y ahorra el esfuerzo humano que habría requerido la exploración conversacional del estado emocional[67].

La capacidad de obtener y procesar a escala los sentimientos impulsa los beneficios de las aplicaciones del procesamiento afectivo. Es lo que hacen las plataformas como Twitter y Facebook con los "corazones" y los "likes" que les proporcionan grandes resultados económicos. Es una forma de simplificar y estandarizar las reacciones. La estrategia reconocida de Facebook es así:

> A diferencia de los algoritmos tradicionales, que están codificados por ingenieros, los algoritmos de aprendizaje automático se "entrenan" en los datos de entrada para aprender las correlaciones dentro de ellos. El algoritmo entrenado, conocido como modelo de aprendizaje automático, puede automatizar decisiones futuras. Un algoritmo entrenado en datos de clics en anuncios, por ejemplo, podría aprender que las mujeres hacen clic en anuncios de mallas de

67 Frank Pasquale, "More than a feeling", *Reallife,* https://reallifemag.com/more-than-a-feeling/

> yoga con más frecuencia que los hombres. El modelo resultante luego servirá para mostrar más de esos anuncios a las mujeres. En la actualidad, en una empresa basada en inteligencia artificial como Facebook, los ingenieros generan innumerables modelos con ligeras variaciones para ver cuál funciona mejor en un problema determinado.[68]

El reconocimiento de afectos puede dar lugar a malentendidos, errores y proyecciones, pero, sean las predicciones ciertas o no, se convierten en hechos con peso que se almacenan en bases de datos que sirven de base a quienes toman decisiones. No se trata de conocer lo que los individuos sienten, sino de controlar sus emociones, limitarlas, establecer normas de conducta, convertir a los individuos en predecibles. No se trata de entender la realidad, sino de crear una nueva realidad.

Un problema adicional es que estos sistemas son muy manipulables porque son muy limitados en su forma de extraer conclusiones.

El profesor Luke Stark, experto en reconocimiento facial, explica su concepto de "imperialismo emocional":

> Significa que los teóricos y profesionales de la computación están importando teorías de la emoción controvertidas o incluso desacreditadas en sistemas digitales que presumen tres cosas incorrectas. Primero, la gente no puede controlar sus emociones. Segundo, nuestras emociones son fáciles de decodificar basándose en un esquema universal simplista. Tercero, nuestras emociones revelan cosas como si somos culpables o inocentes. Equivocado, equivocado y equivocado. Uso el término "imperialismo" deliberadamente. Los antropólogos e historiadores saben hasta qué punto la explotación imperialista y neoimperialista implica la imposición de "estructuras de sentimiento" particulares a diversas poblaciones de todo el mundo. Empresas como Facebook están replicando esa homogeneización a escala global.[69]

En definitiva, el reconocimiento de afectos puede ser un mecanismo para obtener beneficios y subyugar a las personas, una forma de reorganizar la realidad social.

---

68 Karen Hao, "How Facebook got addicted to spreading misinformation", *MIT Technology Review*, https://www-technologyreview-com.cdn.ampproject.org/c/s/www.technologyreview.com/2021/03/11/1020600/facebook-responsible-ai-misinformation/amp/

69 Evan Selinger, "A.I. Can't Detect Our Emotions", OneZero, https://onezero.medium.com/a-i-cant-detect-our-emotions-3c1f6fce2539

Los exámenes universitarios en España durante el COVID-19

Al acercarse el fin de curso 2020 en una situación de aislamiento sanitario, la realización de exámenes, grandes reuniones en espacios cerrados, tuvo que replantearse. La solución adoptada en otros países, principalmente Estados Unidos, fue el empleo de software con medidas de vigilancia que permiten realizar las pruebas de forma no presencial detectando posibles fraudes.

La adopción de esta solución planteaba dos problemas. El primero era en relación a la protección de datos y el derecho de oposición a ser grabado. La CRUE envió en abril de 2020 una guía donde respondía a esta y otras cuestiones relacionadas con los datos personales y el COVID-19. Resumiendo, la legitimación de la grabación por la universidad, como sucede al tratar con organismos públicos, no surge del consentimiento del interesado, sino de la norma. En consecuencia y simplificando, una vez notificado, si el alumno se opone puede considerarse que no ha pasado la prueba.

La segunda cuestión que surgió fue si el software de vigilancia de exámenes cumple su función de manera adecuada y equivalente a la vigilancia durante las pruebas presenciales. En este caso el Ministerio de Universidades envió un documento en el que analizaba opciones, programas disponibles en el mercado y hacía notar las limitaciones de estos sistemas:

1. Falta de eficacia tecnológica del software de reconocimiento facial. Alto ratio de falsos positivos
2. Tecnologías cuestionadas por parte de la comunidad tecnológica
3. Inseguridad regulatoria
4. El mero reconocimiento facial no garantiza la autoría y la ausencia del uso de medios fraudulentos
5. Tecnologías costosas económicamente
6. Imprescindible una evaluación de impacto y un sistema de anonimización o seudonimización y cifrado
7. Hay evidencia de sesgos relevantes con minorías vulnerables
8. Habría que garantizar que no haya sistemas menos gravosos que puedan garantizar el mismo resultado o uno equivalente. Debemos asegurar que la tecnología de reconocimiento facial no produzca más problemas que los beneficios que ofrece

El resumen es que en los exámenes, el RF y la vigilancia automática no son sustituto válido del desarrollo presencial. Probablemente sea más útil y eficiente replantear la forma de evaluación, una cuestión recurrente en la

comunidad académica; una forma que sea compatible con una ejecución no presencial, en vez de añadir más y más tecnología y aumentar costes para no obtener mejores resultados y beneficiar exclusivamente a los proveedores de tecnología y comunicaciones.

El problema radica en que el objetivo no debería ser impedir de forma efectiva el fraude, sino evaluar bien a los alumnos, y eso se debería resolver con pedagogía, no con tecnología.

Historias

Veamos varios casos relacionados con el reconocimiento de emociones y afectos.

La Agencia Ejecutiva de Investigación de la Comisión Europea ha sido investigada por el Tribunal de Justicia de la UE como parte de un caso centrado en la negativa a revelar documentos sobre el controvertido proyecto "*iBorderCtrl*".[70] Esta tecnología, financiada con fondos comunitarios, utiliza inteligencia artificial avanzada para analizar microexpresiones y se ha probado en varias fronteras de la Unión Europea. Una de sus aplicaciones es detectar si un individuo miente o no al ser interrogado. La Agencia de Investigación se ha negado a revelar documentos específicos relacionados con el proyecto porque ciertos "*intereses comerciales*" podrían verse comprometidos si los documentos se hicieran públicos. Algunos creen que los documentos contienen información sobre los algoritmos utilizados con fines de detección de engaños. Esto podría indicar una grave violación de los derechos fundamentales, según el eurodiputado Patrick Breyer. El portavoz de la Agencia manifestó, "*En materia de transparencia, la Comisión siempre anima a los proyectos a dar a conocer en la medida de lo posible sus resultados*", y destacó que el proyecto iBorderCtrl ha designado un asesor de ética para supervisar la implementación de los aspectos éticos de la investigación. La opacidad de los procesos en que se usa IA, y RF en particular, es una tónica general.

El caso más dramático es China. Un artículo en The Guardian describe el intensivo uso de esta tecnología para monitorizar sentimientos de ira, tris-

70 Samuel Stolton, "Commission under pressure in EU court over 'lie detector tech'", *Eurativ* https://www.euractiv.com/section/digital/news/aommission-under-pressure-over-lie-detector-tech-in-eu-courts/

teza, felicidad o cansancio[71]. La tecnología empleada para inferir emociones supuestamente va más allá del RF, emplea los movimientos de los músculos faciales, el tono de la voz y el movimiento del cuerpo. El problema en China se complica por la gran cantidad de razas y las distintas formas de expresar las emociones en cada cultura. El gobierno chino ha instalado este tipo de dispositivos en prisiones, centros de detención y preventivos para monitorizar a los presos 24 horas al día a fin de prevenir los suicidios. También se han instalado cámaras en las escuelas para monitorizar a los profesores, alumnos y personal, así como en centros comerciales y aparcamientos. En gran parte, la vigilancia en China está orientada a la intimidación y a la censura.

Un caso en el que la tecnología ha funcionado en contra de las expectativas es el de la aseguradora Lemonade[72]. Lemonade es una empresa de seguros que permite a las personas presentar reclamaciones a través de videos enviados en una aplicación. La compañía publicó en Twitter que sus chatbots de inteligencia artificial de servicio al cliente recopilan hasta 1,600 puntos de datos de un solo video de un cliente al responder a 13 preguntas. "Nuestra IA analiza cuidadosamente estos videos en busca de signos de fraude. Puede captar señales no verbales que las aseguradoras tradicionales no pueden, ya que no utilizan un proceso de reclamación digital". Tras ser puesta en cuestión la tecnología empleada, la aseguradora publicó "La IA no es determinista y se ha demostrado que tiene sesgos en diferentes comunidades. Por eso, nunca permitimos que la IA realice acciones deterministas como rechazar reclamaciones o cancelar pólizas". Al salir a bolsa en 2020, Lemonade prometía ser un disruptor de la industria clásica respaldado por inteligencia artificial, pero con un giro: sería una corporación de beneficio público con la doble misión de generar ganancias y beneficios sociales. La cuestión importante es *cómo es posible un compromiso con el bien social mediante el uso de sistemas que la propia empresa admite que son propensos al sesgo y la discriminación*. Al final, no está claro en qué grado la IA participa en el proceso, puede ser un caso más de exageración de la tecnología y marketing.

71 Michael Standaert, "Smile for the camera: the dark side of China's emotion-recognition tech", *The Guardian*, https://www.theguardian.com/global-development/2021/mar/03/china-positive-energy-emotion-surveillance-recognition-tech

72 Todd Feathers y Janus Rose, "An Insurance Startup Bragged It Uses AI to Detect Fraud. It Didn't Go Well", *Vice*, https://www.vice.com/en/article/z3x47y/an-insurance-startup-bragged-it-uses-ai-to-detect-fraud-it-didnt-go-well

Límites de la IA e Impacto en el Trabajo y el Medio Ambiente

Durante la pandemia del COVID-19, el hospital de la Universidad de Stanford diseñó un plan de vacunación con un algoritmo que provocó titulares: solo se reservaban 7 de las primeras 5.000 vacunas a médicos residentes; el plan de vacunación favorecía a médicos de bajo riesgo frente a los médicos residentes que trabajaban muy cerca de pacientes COVID.[73] La prensa inicialmente publicó que era un sistema de aprendizaje automático, sin embargo era un simple sistema basado en reglas. Las reglas servían para decidir los que deberían ser vacunados en primer lugar. Todas las reglas eran sencillas. No había nada objetable en ninguna de ellas, pero en conjunto produjeron resultados no óptimos. El profesor Muhammad Aurangzeb Ahmad, especializado en IA en salud, hace esta interesante consideración en una entrevista sobre este incidente

> Básicamente, personas bien intencionadas creían que la automatización podría brindarles los mejores resultados para una situación tensa. Pero al diseñar su sistema, no apreciaron completamente algo esencial al construir cualquier modelo: pueden aparecer interacciones no lineales[74].

Si esto ocurre con sistemas simples, basados en reglas, la situación con sistemas complejos puede llegar a ser mucho más complicada. Los profesores Barocas, Hardt y Narayanan proponen una visión muy amplia de la cuestión de la equidad en el aprendizaje automático. En la introducción de su trabajo señalan:

> Existen serios riesgos al aprender de los ejemplos. El aprendizaje no es un proceso de simplemente memorizar ejemplos. En cambio, implica generalizar a partir de ejemplos: concentrarse en los detalles que son característicos de, por ejemplo, los gatos en general, no solo en los gatos específicos que aparecen en los ejemplos. Este es el proceso de inducción: extraer reglas ge-

---

73 Caroline Chen, "Only Seven of Stanford's First 5,000 Vaccines Were Designated for Medical Residents", *ProPublica*, https://www.propublica.org/article/only-seven-of-stanfords-first-5-000-vaccines-were-designated-for-medical-residents

74 Evan Selinger, "Health Care A.I. Needs to Get Real", *OneZero*, https://onezero.medium.com/health-care-a-i-needs-to-get-real-4aba0ae1241c

> nerales a partir de ejemplos específicos, reglas que efectivamente dan cuenta de los casos pasados, pero que también se aplican a los casos futuros que aún no se han visto. La esperanza es que averigüemos cómo es probable que los casos futuros sean similares a los casos pasados, incluso si no son exactamente iguales[75].

Estamos hablando de sistemas desarrollados mediante *aprendizaje supervisado*[76] que buscan correlaciones, algo esencialmente distinto de las predicciones. Los profesores Arvind Narayanan y Matt Salganik proponen tres límites a la predicción[77]:

1. el posible no determinismo del universo (y, por tanto, los fenómenos de interés);

2. límites para medir los estados de entrada / salida con precisión y recopilar suficientes ejemplos; estos dependen en gran medida de la naturaleza del sistema

3. límites computacionales, ya sean hardware o algoritmos

Dejando aparte los puntos primero y tercero, una dificultad importante para realizar predicciones es la dificultad de recopilar información completa, relevante y suficiente para realizar esa tarea. No obstante, aunque fuera un modelo correcto, lo sería puntualmente; las personas y el contexto varían con el tiempo; hechos imprevisibles pueden cambiar una predicción completamente.

En los tiempos de la IA simbólica y los sistemas expertos el gran problema no fue que se utilizaran datos erróneos, es que faltaba el sentido común, algo imposible de codificar. Narayanan y Salganik aportan otras hipótesis que limitan la capacidad de hacer predicciones, como la sensibilidad de los datos de entrada y el efecto acumulativo de una situación de ventaja o desventaja.

Uno de los problemas clave en las aplicaciones de IA es la diferencia de los resultados de laboratorio con el mundo real. Google desarrolló un sistema de diagnóstico de retinopatía en pacientes diabéticos. El sistema podía dar resultados en 10 minutos con una precisión del 90%, comparable a un experto

---

75 Solon Barocas et al., "Fairness and machine learning. Limitations and opportunities", https://fairmlbook.org/

76 Es el aprendizaje automático que parte de pares entradas-salida para obtener patrones

77 Arvind Narayanan y Matt Salganik, "Limits to prediction: pre-read. COS 597E / SOC 555, Princeton University, Fall 2020", *Princeton University*, https://www.cs.princeton.edu/~arvindn/teaching/limits-to-prediction-pre-read.pdf

humano. La primera prueba real se llevó a cabo en Tailandia donde hay 4,5 millones de pacientes para 200 especialistas. El resultado fue que mientras en ocasiones funcionaba bien y se acelera el proceso, en otras no daba resultados. El sistema había sido entrenado con imágenes de alta calidad y estaba diseñado para rechazar las imágenes de calidad insuficiente, además las imágenes debían subirse a la nube. En zonas rurales de Tailandia no hay recursos para obtener imágenes de alta calidad y frecuentemente el internet es de baja velocidad, lo que ralentizaba el proceso o lo hacía imposible.[78] Este es solo un ejemplo de las promesas de la IA que se desinflan al llegar al mundo real. "El ciclo completo de un proyecto de aprendizaje automático no es solo modelado. Se trata de encontrar los datos correctos, implementarlos, monitorizarlos, retroalimentar los datos [en el modelo], mostrar seguridad, hacer todas las cosas que se deben hacer [para que se implemente un modelo]. [Eso va] más allá de hacerlo bien en el conjunto de pruebas, que, afortunadamente o desafortunadamente, es en lo que somos excelentes en el aprendizaje automático", asegura Andrew Ng. Elon Musk ha reconocido que quizá nunca sea posible el coche completamente autónomo[79].

La profesora Irina Raicu ha escrito un artículo en el que compara la capacidad de predicción de los sistemas con la idea que se forma Elizabeth Bennet del señor Darcy en la novela Orgullo y Prejuicio de Jane Austen[80]. La señorita Bennet cree que no le va a gustar el señor Darcy por la información interesada (sesgada) que ha obtenido. Por esto le sorprende cuando él le pide matrimonio, no era el comportamiento que ella habría predicho del señor Darcy.

A pesar de las dificultades anteriores, se han realizado y se siguen realizando investigaciones sobre reconocimiento de afectos y emociones.

Michal Kosinski es profesor de Comportamiento Organizacional en Graduate Business School, Universidad de Stanford. En 2017 publicó un trabajo que causó revuelo, *Las redes neuronales profundas son más precisas que*

78 Will Douglas Heaven, "Google's medical AI was super accurate in a lab. Real life was a different story", *MIT Technology Review*, https://www.technologyreview.com/2020/04/27/1000658/google-medical-ai-accurate-lab-real-life-clinic-covid-diabetes-retina-disease/

79 Connie Lin, "Tesla admits it may never achieve full-self-driving cars", *Fast Company*, https://www.fastcompany.com/90630440/tesla-admits-it-may-never-achieve-full-self-driving-cars

80 Irina Raicu, "Pride, Prejudice, and Predictions about People", *Markkula Center for applied ethics*, https://www.scu.edu/ethics/internet-ethics-blog/pride-prejudice-and-predictions-about-people/

los humanos en la detección de la orientación sexual a partir de imágenes faciales. El resumen del trabajo resulta interesante

> Usamos redes neuronales profundas para extraer características de 35.326 imágenes faciales. [...] Dada una sola imagen facial, un clasificador podría distinguir correctamente entre hombres homosexuales y heterosexuales en el 81% de los casos y en el 74% de los casos para las mujeres. Los jueces humanos lograron una precisión mucho menor: 61% para hombres y 54% para mujeres. La precisión del algoritmo aumentó al 91% y 83%, respectivamente, al emplear cinco imágenes faciales por persona. Los rasgos faciales empleados por el clasificador incluían tanto rasgos faciales fijos (p. Ej., Forma de la nariz) como transitorios (p. Ej., Estilo de aseo). [...] Los modelos de predicción dirigidos únicamente al género permitieron detectar hombres homosexuales con un 57% de precisión y mujeres homosexuales con un 58% de precisión. [...] nuestros hallazgos exponen una amenaza a la privacidad y seguridad de los hombres y mujeres homosexuales[81].

El profesor Andy Luo publicó una respuesta a este trabajo en la que manifestaba su disconformidad con el trabajo de Kosinski con estas palabras "Todavía existen países donde el conocimiento público de la homosexualidad equivale esencialmente a una sentencia de muerte. ¿Qué pasaría si los responsables de la formulación de políticas en estos países comenzaran a utilizar este software?".[82]

La organización sin ánimo de lucro Data & Society creó un caso de estudio alrededor del trabajo de Kosinski.[83] En él se plantean dilemas éticos e implicaciones en los derechos humanos como la privacidad, discriminación y libertad de expresión y asociación.

Más tarde, en 2021, Kosinski publicó otro trabajo titulado *Facial recognition technology can expose political orientation from naturalistic facial images*. Su resumen también es esclarecedor

---

81 Wang, Yilun y Michal Kosinski. "Deep Neural Networks Are More Accurate Than Humans at Detecting Sexual Orientation from Facial Images", *PsyArXiv. September 7.* https://doi.org/10.1037/pspa0000098

82 Andy Luo, "Neural Networks and Sexual Orientation", *PennState Presidential Leadership Academy*, https://sites.psu.edu/academy/2017/09/22/neural-networks-and-sexual-orientation/

83 "AI Systems and Research Revealing Sexual Orientation Case Study", *Data & Society*, https://datasociety.net/wp-content/uploads/2018/05/AI-Systems-and-Research-Revealing-Sexual-Orientation_Case-Study_Final_CC.pdf

> Se aplicó un algoritmo de reconocimiento facial a imágenes de 1.085.795 individuos para predecir su orientación política comparando su similitud con los rostros de otros liberales y conservadores. La orientación política se clasificó correctamente en el 72% de los pares de rostros liberales-conservadores, notablemente mejor que la casualidad (50%), la precisión humana (55%) o una proporcionada por un cuestionario de personalidad de 100 ítems (66%). La precisión fue similar en todos los países (Estados Unidos, Canadá y el Reino Unido), entornos (Facebook y sitios web de citas) y al comparar rostros entre muestras. La precisión se mantuvo alta (69%) incluso al controlar por edad, género y origen étnico. Dado el uso generalizado del reconocimiento facial, nuestros hallazgos tienen implicaciones críticas para la protección de la privacidad y las libertades civiles[84].

Las mismas objeciones presentadas a la investigación sobre detección de orientación sexual podrían hacerse a la detección de la orientación política. El hecho es que parece que parece existir una correlación.[85] Ocultar su existencia no elimina el problema ni impide que alguien pueda desarrollar sistemas para realizar ese tipo de inferencias, procesos que, de forma automatizada, representan una amenaza a varios derechos humanos.

El problema de la predicción se agrava si los datos empleados son erróneos. Un estudio reciente muestra que 10 de los conjuntos de datos más citados en la investigación en IA están plagados de errores.[86] Anteriormente hablamos de los problemas detectados en ImageNet, este otro trabajo estudia las etiquetas equivocadas. Se detectó un hongo etiquetado como cuchara, una rana etiquetada como un gato y una nota alta de Ariana Grande etiquetada como un silbato. Se estima que el error de etiqueta en ImageNet es del 5,8% mientras que para QuickDraw, una compilación de dibujos a mano, es del 10,1%.

Aunque la determinación del uso correcto sea una combinación de ética y regulación, como en la investigación genética, el comienzo pasa por asegurar que los datos son correctos.

---

84 Michal Kosinsk, "Facial recognition technology can expose political orientation from naturalistic facial images". *Sci Rep 11, 100 (2021).* https://doi.org/10.1038/s41598-020-79310-1

85 John Naughton, "Can facial recognition technology really reveal political orientation?", *The Guardian,* https://www.theguardian.com/commentisfree/2021/jan/23/can-facial-recognition-technology-really-reveal-political-orientation

86 Karen Hao, "Error-riddled data sets are warping our sense of how good AI really is", *MIT Techology Review,* https://www.technologyreview.com/2021/04/01/1021619/ai-data-errors-warp-machine-learning-progress/

La detección de emociones junto con su simulación nos hacen generar empatía hacia unos sistemas carentes de sentimientos. Nos fascina el comportamiento del android Data en Star Trek, y nos preocupa el comportamiento de HAL9000 en 2001: Una Odisea en el Espacio y de la extraña relación con el asistente virtual Samantha en Her. A este respecto dice Luke Stark:

> Estamos acostumbrados a pensar juntos en la emoción y la inteligencia artificial general avanzada. La IA contemporánea no puede hacer nada por el estilo, pero beneficia a las empresas de tecnología si la gente cree por extensión que los sistemas actuales pueden "sentir empatía" o algo similar. Simultáneamente le dice al público que estas tecnologías son más avanzadas de lo que realmente son. Y distrae de los impactos reales de estos sistemas.[87]

La profesora Melanie Mitchell ha publicado recientemente un artículo titulado *Por qué la IA es más difícil de lo que pensamos*[88]. Con una visión muy perspicaz presenta la historia reciente de la IA, las promesas no cumplidas y los repetidos ciclos Invierno IA-Primavera IA. El punto central de su trabajo es que mientras no sepamos cómo funciona la inteligencia humana, su simulación tropezará con obstáculos. En este sentido recuerda cuatro falacias conocidas que pueden confundir incluso a expertos en la materia:

1. La inteligencia artificial estrecha[89] está en un continuo con la inteligencia general[90]. Es la idea de que cualquier avance cuantitativo supone un avance cualitativo.

2. Las cosas fáciles son fáciles y las difíciles son difíciles. Se ha conseguido sistemas que hacen fácilmente cosas difíciles para los humanos como jugar al Go, pero incapaces de hacer cosas fáciles para todos, como la forma de conocer de un niño.

3. El atractivo de las nomenclaturas ilusorias. En IA se utilizan expresiones como APRENDIZAJE, ENTRENAMIENTO e INTELIGENCIA, como forma de referirse al bucle principal del programa, el proceso de ajuste y los patrones obtenidos. El uso de este lenguaje, de forma inconsciente, atri-

87 Evan Selinger, "A.I. Can't Detect Our Emotions", *OneZero*, https://onezero.medium.com/a-i-cant-detect-our-emotions-3c1f6fce2539

88 Melanie Mitchell, "Why AI is Harder Than We Think", *Santa Fe Institute*, https://arxiv.org/pdf/2104.12871.pdf

89 Inteligencia Artificial estrecha o *Narrow AI* se refiere a la IA aplicada a resolver tareas concretas, no generales, como el RF

90 Inteligencia Artificial General o *Artificial General Intelligence (AGI)* es aquella cuyo objetivo es igualar o superar la auténtica inteligencia humana

buye capacidades humanas a sistemas que ni piensan ni comprenden ni se comportan como humanos.

4. La inteligencia está en el cerebro. En la raíz de la IA está la idea de que la inteligencia está en el cerebro, independiente del cuerpo. Si se consigue simular la inteligencia, se podría almacenar, transferir o existir sin dependencia corporal. No seríamos necesarios.

En el artículo *Reward is Enough* David Silver realiza una atrevida aseveración sobre las capacidades del Aprendizaje Automático:

> En este artículo planteamos la hipótesis de que la inteligencia, y sus habilidades asociadas, pueden entenderse como al servicio de la maximización de la *recompensa*.[91] En consecuencia, la recompensa es suficiente para impulsar un comportamiento que exhiba habilidades estudiadas en inteligencia natural y artificial, incluidos el conocimiento, el aprendizaje, la percepción, la inteligencia social, el lenguaje, la generalización y la imitación[92].

Walid Saba sale al paso de esas presunciones y esa confianza infundada en el aprendizaje automático:

> Hay muchos tipos de aprendizaje: aprendizaje por observación, aprendizaje por experiencia (por ensayo y error), aprendizaje por analogía, aprendizaje por instrucción (porque nos lo cuentan / ser enseñado), etc. La mayor parte del conocimiento fáctico consecuente que usamos para funcionar y realizar un razonamiento de sentido común es un conocimiento que no se aprende de manera incremental y, por lo tanto, el aprendizaje por refuerzo ni siquiera es relevante, ya que no se puede definir una "recompensa".[93]

La realidad recuerda tozudamente que la inteligencia humana, incluso la animal, es mucho más que fórmulas y carece de algo sencillo para los humanos e imposible de codificar, el sentido común.

---

91 El aprendizaje mediante recompensas recibe el nombre de *Aprendizaje Reforzado o Reinforcement Learning*. Es similar al entrenamiento de un animal al que se premia o castiga si acata o no la orden dada. El entrenamiento del algoritmo AlphaGo jugando contra sí mismo es un caso práctico de aprendizaje reforzado

92 David Silver et al, "Reward is Enough", *Artificial Intelligence,* https://www.sciencedirect.com/science/article/pii/S0004370221000862

93 Walid Saba, "Reward is NOT Enough, and Neither is (Machine) Learning", *Medium*, https://medium.com/ontologik/reward-is-not-enough-and-neither-is-machine-learning-6f9896274995

Esta es una advertencia importante para los sistemas que vamos a encontrar cada vez más en el día a día. Es necesario un espíritu crítico cada vez mayor antes de confiar en que un sistema inteligente tomará decisiones mejores que una persona o bajar la guardia en una interacción virtual.

Inteligencia Artificial y Trabajo

La IA ya está impactando en el trabajo de forma importante, al menos en tres momentos relevantes:

– Es una forma avanzada de automatización, más allá de una ayuda instrumental, lo que está destruyendo puestos de trabajo y creando desempleo difícil de recolocar

– Cuando los trabajadores tienen que trabajar junto con máquinas inteligentes, las personas sufren estrés, como los trabajadores de Amazon en sus centros de distribución y los trabajadores de Uber, que han perdido su trabajo al no ser identificados por el reconocimiento facial que les habilita para trabajar

– La preparación de los enormes conjuntos de datos para entrenar los modelos, su revisión y etiquetado, requiere ingentes cantidades de trabajo basura[94]

El trabajo desempeña un papel importante en la sociedad y en la realización de cada persona. Es más que un derecho fundamental y la aplicación de la IA lo pone en serio peligro.

En la actualidad existen multitud de foros en los que se discute el futuro del trabajo. La revolución industrial de la IA está detrás del que puede ser un enorme cambio social. El World Economic Forum pronosticaba en 2020 que "la mitad de todas las tareas laborales serán manejadas por máquinas para 2025 en un cambio que probablemente agravará la desigualdad"[95].

En un mundo perfectamente productivo, los humanos serían considerados inútiles en términos de productividad, pero también en términos de nuestra débil humanidad[96].

---

94 Dave Gershgorn, "The A.I. Industry Is Exploiting Gig Workers Around the World — Sometimes for Just $8 a Day", *OneZero*, https://onezero.medium.com/the-a-i-industry-is-exploiting-gig-workers-around-the-world-sometimes-for-just-8-a-day-288dcce9c047

95 "Machines to 'do half of all work tasks by 2025'", *BBC*, https://www.bbc.com/news/business-54622189

96 Arshin Adlib Moghaddam, "Artificial intelligence must not be allowed to replace the imperfection of human empathy", *The Conversation*, https://theconversation.com/artificial-intelligence-must-not-be-allowed-to-replace-the-imperfection-of-human-empathy-151636

Rediseñar el trabajo no se trata simplemente de automatizar tareas y actividades. En esencia, consiste en configurar el trabajo para capitalizar lo que los humanos pueden lograr cuando el trabajo se basa en sus fortalezas[97].

Inteligencia Artificial y Medio Ambiente

Una investigación de la Universidad de Montreal, Canadá, revela datos sobre la energía que necesita la IA: "nuevas estimaciones sugieren que la huella de carbono de entrenar una sola IA es equivalente a 284 toneladas de dióxido de carbono, cinco veces las emisiones de por vida de un automóvil promedio"[98]. A veces se justifica el gasto asegurando que se utilizan energías renovables. Esa afirmación es una falacia, a menos que el 100% de la energía eléctrica sea verde, pues esa energía verde se podría almacenar o emplear para reemplazar a otras energías menos verdes.

Un argumento utilizado con frecuencia para justificar el consumo energético intensivo es que los métodos de entrenamiento no son eficientes, pero con la infraestructura adecuada lo serán[99]. Sin embargo, los datos son contundentes. Según Roel Dobbe, la intensidad de carbono de muchos sistemas de inteligencia artificial está impulsada por la creencia en las posibilidades de lo que se conoce como "gran computación". "En el campo de la IA, existe una creencia dominante pero falsa de que "cuanto más grande, mejor", pero una mayor potencia computacional aumenta la huella de carbono. En 2018, OpenAI informaba que "desde 2012, la cantidad de cálculo utilizado para entrenar los sistemas de IA más grandes se ha duplicado cada 3,4 meses"[100]. En 2019 Jerome Pesenti, Director de IA de Facebook decía "Si nos fijamos en los mejores experimentos, cada año el costo aumenta diez veces. En este

97 MIT Technology Review Insights, "The future of work is uniquely human", *MIT Technology Review*, https://www.technologyreview.com/2021/04/06/1021793/the-future-of-work-is-uniquely-human/

98 Pascale Lehoux y Lysanne Rivard, "Hidden in plain sight: The infrastructures that support artificial intelligence", *The Conversation*, https://theconversation.com/hidden-in-plain-sight-the-infrastructures-that-support-artificial-intelligence-146087

99 Kate Saenko, "AI has a huge carbon footprint. Here's what we can do to reduce it", *Fast Company*, https://www.msn.com/en-us/news/technology/ai-has-a-huge-carbon-footprint-here-s-what-we-can-do-to-reduce-it/ar-BB1c0DQV

100 Jacob Dykes, "The carbon footprint of AI and cloud computing", *Geographical*, https://geographical.co.uk/nature/energy/item/3876-the-carbon-footprint-of-ai-and-cloud-computing

momento, [el presupuesto de] un experimento puede tener siete cifras, pero no llegará a nueve o diez cifras, no es posible, nadie puede permitírselo"[101].

Recientemente la profesora Kate Crawford ha publicado el libro *Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence* en el que denuncia la cantidad de recursos que se emplean en IA y Cloud Computing, y el impacto ambiental de las nuevas tecnologías. Cita el litio, empleado para construir baterías de los vehículos eléctricos. Recuerda lo que han supuesto anteriores tecnologías en recursos naturales e impacto ambiental. El punto que defiende es que

> La IA no es ni artificial ni inteligente. Más bien, la inteligencia artificial es restringida y material, hecha de recursos naturales, combustible, trabajo humano, infraestructuras, logística, historias y clasificaciones. Los sistemas de inteligencia artificial no son autónomos, racionales o capaces de discernir nada sin un entrenamiento extenso e intensivo en computación con grandes conjuntos de datos o reglas y recompensas predefinidas[102].

En consecuencia, plantea que en realidad la implantación de la IA tiene más que ver con el poder que genera, la explotación humana y de los recursos naturales que con la ciencia.

---

101 A L Dellinger, "Artificial intelligence development is starting to slow down, Facebook head of AI says", *Mic*, https://www.mic.com/p/artificial-intelligence-development-is-starting-to-slow-down-facebook-head-of-ai-says-19424331

102 Kate Crawford, "Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence", *Yale University Press, ISBN: 978-0300209570*

La TRF abre la oportunidad a comportamientos contrarios a los derechos humanos, la autonomía personal y la libertad de actuación y expresión.

Un grupo de investigadores de la universidad de Stanford, Berkeley y otras, han estudiado cien artículos muy citados, presentados en las dos conferencias más importantes de Aprendizaje Automático, ICML y NeurIPS, para determinar los valores perseguidos en esas investigaciones. Los resultados concluyen que la mayoría de los proyectos tiene poco que ver con las necesidades sociales, y que se justifican y evalúan por su desempeño, generalización, eficiencia, novedad y sobre la base de trabajos anteriores. "En particular, encontramos que cada uno de estos valores objetivo se está definiendo y aplicando actualmente con supuestos e implicaciones que generalmente apoyan la centralización del poder. Finalmente, encontramos vínculos cada vez más estrechos entre estos artículos altamente citados y las empresas de tecnología y las universidades de élite."[103]

IA y la "buena vida"

La IA es un desarrollo humano que, para ser útil, debe contribuir a una vida mejor. En los capítulos anteriores hemos descrito con fundamento técnico y de forma práctica muchos casos en los que se ha causado o se podría causar perjuicios a las personas.

Los filósofos plantean discusiones sobre la esencia de las cosas, las causas, los fines, la existencia, el conocimiento, la moral, el valor. Considerando la presencia de la IA en la sociedad, es esencial plantear la cuestión del papel de la IA y su relación con el fin último del ser humano.

Existen muy buenas contribuciones al debate filosófico en torno a la IA. Me ha parecido oportuno incluir dos breves e interesantes publicaciones recientes de los profesores John Tasioulas y Shannon Vallor.

El 14 de junio de 2021, el profesor Tasioulas publicaba en el blog de Ada Lovelace Institute un artículo titulado *The role of the arts and humanities in thinking about artificial intelligence (AI)*[104]. Comienza diciendo que el de-

103 Abeba Birhane et al., "The Values Encoded in Machine Learning Research", *Arxiv*, https://arxiv.org/abs/2106.15590

104 John Tasioulas, "The role of the arts and humanities in thinking about artificial

sarrollo de la IA no es una cuestión del destino, sino la consecuencia de una sucesión de elecciones humanas y que por ello es importante identificar las opciones, enmarcarlas de la manera correcta y plantear la pregunta: ¿quién puede tomarlas y cómo?

Afirma Tasioulas que la IA se ha convertido en un nuevo enfoque del mito historicista, en el que la evolución social está predeterminada por variables y corrientes sobre las que no tenemos control, y recuerda la frase de Aristóteles "Nadie delibera sobre cosas que son invariables, ni sobre cosas que le es imposible hacer".

Defiende el papel de las artes y las humanidades para combatir esta tendencia historicista que merma a los individuos y a la sociedad. Recuerda que podemos elegir las decisiones a tomar. "La ética es ineludible porque concierne a los valores últimos en los que se basan nuestras elecciones, nos demos cuenta o no". La ética filosófica trata del fin último del ser humano, sin embargo, señala Tasioulas, en las discusiones sobre la IA se contempla una versión disminuida, identificada con un estrecho subconjunto de valores éticos.

Pone de manifiesto la infundada ilusión de los científicos sobre el poder de los datos y de la optimización, lo que facilita la adopción de una ética utilitarista. Continúa resaltando que una forma de utilitarismo, la maximización de la riqueza, es lo que guía a los actores económicos y de gobierno.

Propone tres características para una ética más humanista: pluralismo, procedimiento y participación. Esta propuesta tiene cierta conexión con el *Desarrollo y uso responsable de la IA* que veremos más adelante.

Pocos días después, el 25 de junio de 2021, la profesora Vallor publicaba en el mismo blog el artículo *Mobilising the intellectual resources of the arts and humanities*[105], que apoya el papel de las humanidades y las artes en la IA.

Insiste en la idea de Sócrates de la búsqueda de la "buena vida", una vida llena, realizada, como el objetivo de los seres humanos. Parece que una *IA buena* es la que logra resultados justos: equidad, privacidad y transparencia, un objetivo bastante pobre como realización humana.

La profesora Vallor afirma: "El mayor reto de la humanidad hoy en día es el continuo ascenso de un régimen tecnocrático que busca compulsivamen-

---

intelligence (AI)", *Ada Lovelace Institute*, https://www.adalovelaceinstitute.org/blog/role-arts-humanities-thinking-artificial-intelligence-ai/

105 Shannon Vallor, "Mobilising the intellectual resources of the arts and humanities", *Ada Lovelace Institute*, https://www.adalovelaceinstitute.org/blog/mobilising-intellectual-resources-arts-humanities/

te optimizar todas las operaciones humanas posibles sin saber preguntarse qué es lo óptimo, o incluso por qué optimizar es bueno." A continuación se pregunta si se podría pintar un cuadro con un color "óptimo" o una sinfonía "óptima", que sustituyera a todas las demás. Concluye que la "buena vida" no es un estado optimizado del ser, es una obra de arte cambiante, fluida, una vida en construcción.

En la película *El Show de Truman*[106] el protagonista tiene una meta, ir a las islas Fidji, pero eso se sale del guión del *reality*. Truman lleva una vida óptima: tiene un buen empleo, una mujer en casa, un amigo, todos le conocen, pero no es feliz porque lo que quiere no es tener una vida perfecta, sino vivir su vida.

Vallor insiste en rechazar con el profesor Tasioulas el destino inexorable de la humanidad basado en la optimización. Coincide en que en el desarrollo de la IA se desea mantener un marco ahistórico mirando al futuro y olvidando el pasado. "No se puede considerar que herramientas de IA como el reconocimiento facial omnipresente y la vigilancia policial predictiva sean un retroceso a las prácticas colonialistas extractivas y represivas si sólo se mira hacia adelante."

Rechaza la aseveración de Daniel Kahneman[107] de que "la IA va a ganar", porque eso sería igual que decir que algunos seres humanos van a ganar una guerra contra otros seres humanos.

Continúa Vallor:

> Como señala Tasioulas, la tecnología no es neutral. Las tecnologías son formas de construir valores humanos en el mundo. Hay una ética implícita en la tecnología, siempre. Y lo que tenemos que hacer es ser capaces de explicitar esa ética implícita, para que podamos examinarla y cuestionarla colectivamente, para que podamos determinar dónde está justificada, dónde sirve realmente a los fines del florecimiento humano y la justicia y dónde no. Pero mientras se permita que la ética implícita de la tecnología permanezca oculta, seremos impotentes para cambiarla e incorporar una ética más sostenible y equitativa en la forma de concebir el mundo construido.

---

106 El Show de Truman, 1998, director Peter Weir

107 Tim Adams, "Daniel Kahneman: 'Clearly AI is going to win. How people are going to adjust is a fascinating problem'", *The Guardian*, https://www.theguardian.com/books/2021/may/16/daniel-kahneman-clearly-ai-is-going-to-win-how-people-are-going-to-adjust-is-a-fascinating-problem-thinking-fast-and-slow

Algor-ética

En febrero de 2020 la Academia Pontificia de la Vida organizó el congreso Rome Call.[108] Francisco defendió en su discurso para este congreso la algor-ética, la propuesta del teólogo franciscano Paolo Benanti, como "un puente para que los principios se inscriban concretamente en las tecnologías digitales, mediante un diálogo transdisciplinar eficaz".

En ese congreso, Microsoft, IBM, FAO y el Ministerio de Innovación italiano firmaron el documento *Call for an AI ethics* para "respaldar un enfoque ético de la inteligencia artificial y promover un sentido de responsabilidad entre las organizaciones, gobiernos, instituciones y el sector privado con el objetivo de crear un futuro en el que la innovación digital y el progreso tecnológico sirvan al genio y la creatividad humanos y no a su reemplazo gradual". Coincidencia o no, en junio de ese mismo año 2020, Microsoft, IBM y Amazon firmaron una moratoria de su tecnología de reconocimiento facial, como ya citamos anteriormente.

En julio de 2018 Paolo Benanti pronunció una conferencia TEDx en la que planteaba el escenario del ser humano y la IA, el *homo sapiens* y la *maquina sapiens* y presentaba su propuesta algor-ética.[109] Para proteger la dignidad de las personas, propone que las máquinas sigan estos principios:

1. Intuición (anticipación): cuando dos seres humanos se encuentran, ambos intuyen lo que el otro quiere hacer y cooperan entre sí. En un ambiente mixto individuo-robot el algoritmo debe adaptarse a la persona y su singularidad y no viceversa.

2. Inteligibilidad (transparencia): cuando una persona va a realizar una labor dentro de un entorno de trabajo, los que están alrededor entienden que movimientos va a seguir para ejecutarla. Cuando se trabaja con máquinas debemos saber qué pasos va a seguir, sobre todo para no exponernos a situaciones de riesgo.

3. Adaptabilidad: los seres humanos saben adaptarse al entorno y a las circunstancias. En un entorno mixto la máquina también debe saber cómo adaptarse a la persona con la que interactúa.

4. Adecuación de los objetivos: una máquina sigue unos algoritmos con el fin de alcanzar un objetivo eficiente, que no siempre se ajusta a lo que necesita

108 Rome Call For AI Ethics A Human-Centric Artificial Intelligence, https://www.romecall.org/

109 "Algor-Ethics: Developing a Language for a Human-Centered AI | Padre Benanti | TEDxRoma", *YouTube*, https://www.youtube.com/watch?v=rFzjsHNertc

en ese momento una persona. La prioridad operacional no puede estar en el algoritmo, sino en la persona con la que coopera.

En definitiva, cuatro aspectos de una misma idea: la máquina debe trabajar para la persona y no al revés.

Miguel Angel Correas considera que la propuesta de Benanti ofrece una solución para guiar y gestionar la innovación tecnológica desde el auténtico desarrollo humano, no para estar sujeto a ella.[110]

Desarrollo y uso responsable de la IA

La profesora Virginia Dignum defiende la necesidad de una IA responsable y ese debe ser el origen de la regulación. Cabe el peligro de regular herramientas o tecnología en vez de los usos. En un breve artículo, Dignum recoge en qué consiste la IA responsable.[111]

Recuerda que la IA es mucho más que un algoritmo, como una receta para hacer un pastel de manzana. Si queremos cocinar un pastel, la receta no se convierte en pastel por sí sola. Tiene que ver más con las habilidades del chef y la elección de los ingredientes. Algo similar ocurre con los algoritmos, dependen de la calidad de los datos y de quienes los entrenaron. Tenemos la opción de usar manzanas orgánicas en nuestro pastel. En la IA también tenemos la opción de usar datos que respeten y garanticen la equidad, la privacidad, la transparencia y todos los demás valores que queramos proteger.

También señala que la IA no es aprendizaje automático, esta es solo una de las tecnologías que se emplean actualmente para extracción de patrones mediante redes neuronales. La capacidad de detección de patrones está muy lejos de la comprensión de su significado.

La definición original de John McCarthy de IA era el esfuerzo por desarrollar una máquina que pudiera razonar como un humano, capaz de pensamiento abstracto, resolución de problemas y superación personal. En consecuencia, la IA como campo científico tiene que ver con el razonamiento, el significado.

> Un enfoque responsable y ético de la IA va más allá de la tecnología utilizada y debe incluir el contexto social, organizacional e institucional de esa tecnología. Es este ecosistema socio-técnico el que debe garantizar la transpa-

110 Miguel Angel Correas, "La propuesta de la algor-ética de Paolo Benanti para el desarrollo de la inteligencia artificial", *Grupo Interdisciplinar Ciencia y Fe*, https://dialogocyf.blogspot.com/2020/06/la-propuesta-de-la-algor-etica-de-paolo.html

111 Virginia Dignum, "What we need to talk about when we talk about AI", *LinkedIn*, https://www.linkedin.com/pulse/what-we-need-talk-when-ai-regulatory-purposes-virginia-dignum/

> rencia sobre cómo se realiza la adaptación, la responsabilidad por el nivel de automatización en el que el sistema es capaz de razonar y la responsabilidad por los resultados y los principios que guían sus interacciones con los demás, lo que es más importante, con personas. Además, y sobre todo, un enfoque responsable de la IA deja claro que los sistemas de IA son artefactos fabricados por personas, para algún propósito, y que las personas son responsables del uso y desarrollo de la IA.

Según Catelijne Muller, "la IA no opera en un mundo sin ley". Antes de definir regulaciones adicionales, hay que empezar por comprender lo que ya está cubierto por la legislación existente, incorporar nuevos escenarios y actores y regular lo nuevo.

La regulación deberá centrarse primero en los resultados, tanto si los sistemas entran en la comprensión actual de lo que es "IA" o no. Por ejemplo, si alguien es identificado erróneamente, se le niegan los derechos humanos, el acceso a los recursos, o está condicionado a creer o actuar de cierta manera, no importa qué tecnología o método se utilice. Simplemente está mal.

El análisis ético de las aplicaciones de IA es una cuestión de gran actualidad que no se resuelve con el cumplimiento de un conjunto de *checkboxes* en un proceso de *compliance*.

Derechos Humanos

La protección de los Derechos Humanos es el fundamento de la necesidad de regular tecnologías potencialmente invasivas como las que tratamos en este libro. En marzo de 2021 la Comisión Australiana de Derechos Humanos ha publicado un extenso informe, titulado *Derechos Humanos y Tecnología*[112]. Dedica un apartado a *vigilancia biométrica, reconocimiento facial y privacidad*. Si bien el documento se refiere a Australia, hace algunas consideraciones interesantes y muy actuales.

Señala que la información personal es el "combustible" que impulsa la IA y que la IA plantea riesgos particulares para el control de las personas de su propia información personal y el derecho a la intimidad. Datos personales aparentemente inocuos se pueden emplear, especialmente en un sistema IA, para obtener conocimiento sobre un individuo, incluso sobre asuntos delicados.

---

112 "Human Rights and Tecnology", *Australian Human Rights Comisión*, https://tech.humanrights.gov.au/downloads

Recuerda que el uso de IA en tecnología biométrica, y especialmente algunas formas de reconocimiento facial, ha provocado una creciente preocupación entre el público y los expertos.

La tecnología se puede utilizar para verificar la identidad de una persona o para identificar a alguien de un grupo. También hay quienes intentan aplicar la tecnología para extraer otros conocimientos sobre las personas, como su estado de ánimo o personalidad.

Esto necesariamente afecta la privacidad individual y puede alimentar una vigilancia dañina. Además, ciertas tecnologías biométricas son propensas a altas tasas de error, especialmente para grupos raciales y de otro tipo. Al emplear estas tecnologías biométricas aumenta el riesgo de injusticias y otras violaciones de los derechos humanos. El tipo de tecnología empleada y su uso afecta al grado de riesgo para los derechos humanos.

La Comisión concluye que las protecciones existentes de la privacidad y antidiscriminación son inadecuadas a estas tecnologías por tres razones.

En primer lugar, el problema de la alta tasa de error, especialmente en el uso de la tecnología de reconocimiento facial para identificación. El hecho de que estos errores afecten de manera desproporcionada a las personas en relación con características como el color de la piel, el género y la discapacidad sugiere que se debe tener mucha precaución antes de utilizar esta tecnología para tomar decisiones que afecten los derechos legales y de importancia similar de las personas.

En segundo lugar, la identificación por reconocimiento facial se ha probado en la toma de decisiones gubernamentales y de otro tipo de gran importancia, incluso en el ámbito policial, educativo y de prestación de servicios. Por lo general, esas pruebas se han llevado a cabo en escenarios reales, donde cualquier error que resulte en una violación de los derechos humanos no puede remediarse fácilmente, si es que pueden remediarse.

Reconoce que la legislación existente no ha demostrado ser un freno eficaz para el uso inadecuado de la tecnología facial y otras tecnologías biométricas. Sin una regulación eficaz en esta área, parece probable que la confianza de la comunidad en la tecnología subyacente se deteriore. Una consecuencia podría ser la desconfianza tanto en los usos beneficiosos como en los perjudiciales.

En tercer lugar, el crecimiento del reconocimiento facial y otras tecnologías biométricas, junto con otros fenómenos como el crecimiento de las cámaras de televisión de circuito cerrado en lugares públicos, está contribuyen-

do a un mayor riesgo de vigilancia masiva. Con una protección legal limitada contra este impacto acumulativo, existe un riesgo real de que los ciudadanos cedan paulatinamente su privacidad, de manera que no se pueda deshacer.

Por estas razones -concluye- se necesita una legislación específica para prevenir y abordar los daños asociados con el uso del reconocimiento facial y otras tecnologías biométricas. Según destacados expertos, la legislación debería prohibir ciertos usos de esta tecnología, si no se pueden cumplir las normas de derechos humanos.

Además, el informe sugiere modificaciones en la regulación de la privacidad.

Los *Australian Privacy Principles* permiten que la información personal anónima o que ha sido anonimizada sea procesada para el propósito principal para el que fue recopilada, con el consentimiento de las personas interesadas. Sin embargo, los avances tecnológicos están cuestionando este modelo de protección de la privacidad de la información. Por ejemplo, la IA ofrece cada vez más la capacidad de desglosar un conjunto de datos formado por un conglomerado de datos no identificados para revelar la información personal de personas identificables específicas.

La tecnología de procesamiento de datos se está desarrollando rápidamente con nuevos usos cada vez mayores de la información personal, muchos de los cuales no podrían haber sido previstos, y mucho menos consentidos específicamente, en el momento de la recopilación. Los tribunales australianos han sostenido que cierta información potencialmente reveladora, incluidos los metadatos, no está dentro de los parámetros de la *Privacy Act*.

Esta provisión también debería aparecer en el RGPD de la Unión Europea, independientemente de la tecnología que se emplee.

Libertad de Expresión

Creemos que la libertad de expresión merece un tratamiento particular en relación a las TRF. Mediante su uso es posible obtener de forma automatizada, al menos idealmente, la identidad de una persona, sus características personales como raza y orientación sexual, lo que dice, -algunos pretenden saber incluso lo que piensa- en espacios públicos, sin necesidad de contacto. El nivel de control que proporciona la tecnología puede tener un efecto paralizante en la libertad de expresión. Por ejemplo, se han utilizado sistemas de RF para identificar activistas del movimiento Black Lives Matter, lo que

no deja de ser irónico, y en la India, la compañía pública Punjab State Power Corporation está probando un sistema de lectura de labios para detectar acoso laboral.

El problema de la tecnología de vigilancia es que cuando se tiene, se usa. Limitar su aplicación a casos puntuales, de hecho, es una situación que no se da. Aunque se requiera una autorización judicial, hay muchos casos de vigilancia ilegal, como los conocidos por las filtraciones de Edward Snowden. La prevención de ataques terroristas se ha convertido en carta blanca para no respetar los derechos.

Nadie despliega un sistema de vigilancia y lo tiene apagado, nadie despliega un sistema de vigilancia solo cuando se busca a un menor. Los sistemas de vigilancia solo son útiles si están operativos siempre. Con las TRF el derecho a la autonomía, el anonimato y la libertad de expresión quedan entre paréntesis. Como han dicho algunos autores, son una amenaza para la democracia liberal. Podemos tener democracia o podemos tener una sociedad de vigilancia, pero no ambas.

Ética del Reconocimiento Facial

Para este apartado seguiré el trabajo *The Ethics of Facial Recognition Technology* de Evan Selinger y Brenda Leong[113].

En capítulos anteriores hemos visto cómo mediante las TRF las personas pueden resultar perjudicadas: castigadas, encarceladas, vigiladas, etiquetadas o discriminadas en el acceso a un trabajo, de forma automatizada. Hay un problema de justicia.

En no pocos casos se culpa a las implementaciones de discriminar por raza, sexo u otros rasgos faciales. Eventualmente, la técnica mejorará. No obstante, persisten otras cuestiones de importancia. Además, no es posible lograr un sistema que no discrimine simultáneamente según varios criterios.

Selinger y Leong advierten que no todas las personas son igualmente vulnerables a ser perjudicadas por una identificación errónea ni todas experimentan la misma inquietud por ser objeto de RF. Aquí aparece el problema ético y legal del perjuicio distribuido desproporcionadamente.

Un artículo de Richard Van Noorden pública los resultados de una encues-

113 Evan Selinger y Brenda Leong, "The Ethics of Facial Recognition Technology", *Forthcoming in The Oxford Handbook of Digital Ethics ed. Carissa Véliz*. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3762185

ta a científicos sobre las TRF. Más del 70% de los encuestados opinan que las investigaciones sobre poblaciones vulnerables no son o pueden no ser éticas. La mayor parte de las imágenes utilizadas para investigación en RF provienen de Internet sin consentimiento del interesado, y el 40% de los investigadores es partidario de pedir consentimiento previo. Continúa la publicación,

> Un científico agregó: "Es necesario que el público comprenda mucho más [qué significa] la concesión de licencia [para uso] de imágenes y datos. En la actualidad, los acuerdos de licencia no se escriben teniendo en cuenta la comprensión y el empoderamiento del usuario. Están redactados para permitir al investigador hacer lo que quiere hacer y tener suficiente protección legal"[114].

La TRF crea un conjunto de situaciones que pueden ser muy perjudiciales para las personas. Hemos visto como el RF puede verificar o averiguar una identidad, e incluso puede inferir el estado de ánimo, o las preferencias políticas, religiosas o sexuales. Algunas de esas capacidades son falsas, al menos parcialmente; sin embargo, una vez desplegada la tecnología, causa un *efecto paralizante* como si esas capacidades fueran verdaderas. Frecuentemente, la caracterización mediante RF es *socialmente tóxica* porque reafirma estereotipos de raza y género.

Una de las consecuencias que apuntan Selinger y Leong del empleo de la TRF es la debilitación de la confianza. La convivencia social, al menos en las democracias liberales, se basa en cierta confianza entre las personas y las instituciones. Si hay desconfianza en la aplicación de la ley, el objetivo de administrar justicia se pone más difícil. Hemos visto en ejemplos anteriores que muchos cuerpos policiales han implantado sistemas de identificación por RF sin una investigación adecuada y ha provocado protestas, por ejemplo en Detroit y Portland. Un estudio de Pew Research Center[115] indica que los jóvenes de 19 a 29 años rechazan más que los mayores el uso del RF por las fuerzas del orden. Al mismo tiempo, los negros e hispanos desconfían más que los blancos.

El Georgetown Law Center on Privacy and Technology ha realizado estu-

114 Richard Van Noorden, "What scientists really think about the ethics of facial recognition research", *Nature*. https://www.nature.com/articles/d41586-020-03257-6

115 Aaron Smith, "More Than Half of U.S. Adults Trust Law Enforcement to Use Facial Recognition Responsibly", *Pew Research Center*, https://www.pewresearch.org/internet/2019/09/05/more-than-half-of-u-s-adults-trust-law-enforcement-to-use-facial-recognition-responsibly/

dios sobre el uso del RF por las fuerzas del orden y ha llegado a varias conclusiones: las fuerzas del orden están usando redes de vigilancia facial con la que pueden escanear rostros de personas mientras caminan por la calle; la aplicación de la ley no siempre ha sido transparente con el público sobre cómo están utilizando la TRF; los agentes del orden público, que no están sujetos a los procedimientos estandarizados a nivel federal para el uso de la TRF, se han involucrado en prácticas de mala calidad, como el uso inapropiado de bocetos forenses, fotos de famosos y de imágenes de baja calidad; las fuerzas del orden buscan en bases de datos que contienen enlaces de nombres de más de la mitad de los adultos estadounidenses sin obtener el consentimiento previo explícito de los ciudadanos; y los agentes gubernamentales han empleado prácticas impugnadas legalmente al usar sistemas de RF en la vigilancia de las salidas internacionales de los aeropuertos.

El perjuicio distribuido desproporcionadamente causa que las mayorías, que no se ven muy afectadas negativamente por una tecnología, sean proclives a aceptar alguna aunque perjudique más a las minorías.

La organización Future of Privacy Forum ha elaborado un documento sobre los daños potenciales debidos a la toma de decisiones automatizadas[116]. Señala estos grupos:

1. Pérdida de oportunidades, como discriminación en el trabajo, en beneficios sociales, vivienda y educación

2. Pérdidas económicas, como discriminación crediticia, precio diferenciado de bienes y servicios, limitación de opciones

3. Perjuicio social, como burbujas de filtrado, refuerzo de estereotipos, confirmación de sesgos

4. Pérdidas de libertad, como el incremento de la vigilancia y el encarcelamiento desproporcionado

Los conductores de Uber Eats en el Reino Unido están siendo despedidos debido al software de identificación facial defectuoso. Uber requiere que los conductores envíen selfies para confirmar su identidad al inicio de la jornada. Cuando la tecnología falla no pueden trabajar[117].

---

116 Future of Privacy Forum, "Unfairness by Algorithm: Distilling the Harms of Automated Decision-Making", *Future of Privacy Forum*. 11 December, https://fpf.org/blog/unfairness-by-algorithm-distilling-the-harms-of-automated-decision-making/

117 Fight for the Future, "Why We Absolutely Must Ban Private Use of Facial Recognition", *Medium*, https://fightfortheftr.medium.com/why-we-absolutely-must-ban-private-use-of-facial-recognition-98094736933

Este caso muestra que el uso privado del RF por parte de instituciones e incluso individuos tanto como su uso por las administraciones públicas y las fuerzas del orden plantea desafíos para el futuro de la civilización.

Además de los perjuicios personales, la sociedad en conjunto también queda debilitada en cuanto disminuye la confianza en la experiencia, los propios méritos y en ser tratado justamente. La organización AI Now, en su informe sobre IA de 2019, hacía esta recomendación *"Los reguladores deberían prohibir el uso del reconocimiento de afectos en decisiones importantes que impactan la vida de las personas y el acceso a las oportunidades"*[118].

Cuando se implanta una tecnología, también sufren daños aquellos que deciden no usarla. Puesto que la IA en aplicaciones de impacto social opera sobre unos perfiles aprendidos, la información no disponible será inferida, por lo que sufrirán igualmente las consecuencias quienes se auto excluyan.

Si los individuos creen que no pueden escapar de la ubicuidad de la vigilancia, será incluso más probable que pierdan la confianza en las garantías de que sus datos y opciones de privacidad están siendo respetadas o aplicadas.

Existen muchas iniciativas para regular el RF a fin de que no perjudique a las personas exigiendo transparencia, rendimiento de cuentas, gobernanza, eliminación de sesgos. Todas esas medidas se dirigen a conseguir una tecnología de vigilancia perfecta.

Además de la aceptación personal, hay que considerar que el RF empleado por las fuerzas del orden causa daños desproporcionados a las minorías, lo que atropella la "autonomía colectiva". Es requisito que una democracia salvaguarde las libertades fundamentales para todos.

Las personas somos imprecisas al reconocer a otras, tanto o más que los algoritmos de RF. Si todos los sistemas de RF funcionan perfectamente siempre, habríamos llegado al *mundo orwelliano* de una intensa vigilancia ubicua.

La conciencia de estar siendo observados altera el comportamiento de los individuos, si van a obrar mal o no. Por esto se dice que la vigilancia tiene un *efecto paralizante* en los individuos que provoca la autocensura en público y limita el comportamiento a normas de grupo aceptables. Esto priva a la sociedad de las opiniones de todos, especialmente de aquellos que se apartan del discurso dominante, como minorías y activistas que luchan por motivos éticos[119].

---

118 Kate Crawford K, Meredith Whittaker et al., "AI Now 2019 Report", *AI Now Institute*, https://ainowinstitute.org/AI_Now_2019_Report.pdf

119 Margot E. Kaminski y Shane Witnov, "The Conforming Effect: First Amendment

Para identificar otro daño que puede producir una vigilancia perfecta, el filósofo Benjamin Hale propone un experimento mental futurista: imagina una sociedad que se acerca al ideal de la vigilancia perfecta y utiliza la vigilancia facial ubicua como herramienta para disuadir a las personas de la actividad delictiva. Según Hale, este enfoque de la gobernanza corre el riesgo de erosionar una normativa ideal que, en el sistema ético kantiano, es fundamental para la toma de decisiones autodeterminada: "la voluntad libre".

Otra consecuencia del empleo de la vigilancia generalizada, apuntan Selinger y Leong, es la alienación del ser humano. Una parte tan importante del cuerpo y la personalidad como es el rostro queda reducida a mero medio para transmitir información a un sistema de identificación. El rostro, que es parte de la vida, experiencia y expresión personal, queda reducido al papel de contraseña, PIN o código de barras.

El filósofo Phillip Brey aclara esta deshumanización causada por el RF

> Este proceso de reducción funcional implica la creación de equivalentes informativos de partes del cuerpo que existen fuera de su dueño y son utilizados y controlados por otros. Por lo tanto, no solo se está produciendo un proceso de reducción, sino también de alienación: la huella facial que caracteriza de manera única a su rostro ya no es 'suyo', sino 'de ellos': no es de su propiedad e incluso si lo fuera, no lo sería por usted porque no comprende la tecnología. De esta manera, las personas pueden llegar a sentir que algunas de las partes de su cuerpo ya no son completamente 'suyas', porque han adquirido significados que su dueño no comprende y usos que se realizan parcialmente fuera de su propio cuerpo[120].

A lo largo de la historia de la humanidad algunas tecnologías han encontrado oposición previa argumentando consecuencias preocupantes como la experimentación con ADN humano. Algunas prácticas y tecnologías han sido socialmente rechazadas a posteriori como las armas atómicas, químicas y el amianto. Son sustancias y tecnologías tóxicas porque pueden causar efectos muy perjudiciales y una vez incorporadas son muy difíciles de eliminar. El RF podría calificarse también como sustancia tóxica social.

Implications of Surveillance, Beyond Chilling Speech", *University of Richmond Law Review vol. 49: 465*. https://lawreview.richmond.edu/files/2015/01/Kaminski-492.pdf

120 P. Brey, "Ethical Aspects of Facial Recognition Systems in Public Places", *Journal of Information, Communication, and Ethics in Society vol. 2: 97-109*. https://pdfs.semanticscholar.org/ca69/ebedd468b808f4a9a6f862245c5923777498.pdf

El trabajo citado de Selinger y Leong concluye planteando la cuestión de los argumentos tipo pendiente resbaladiza aplicados al RF.

La forma en que se debería usar el RF está teniendo un profundo impacto en los debates éticos, legales y políticos. En definitiva hay dos cuestiones subyacentes, en el caso de que sus poderosas prestaciones, con el tiempo, puedan desestabilizar sociedades, incluso las democráticas ¿tiene sentido buscar formas cada vez más expansionistas de uso? En caso afirmativo ¿es más probable que se produzca un detrimento inaceptable de la dignidad y la libertad mayor del que se suele admitir?

Hartzog propone la prohibición del RF a fin de evitar una cadena imparable de perjuicios, una pendiente resbaladiza:

> Creemos que la tecnología de reconocimiento facial es el mecanismo de vigilancia más peligroso jamás inventado. Es la pieza que falta en una infraestructura de vigilancia ya peligrosa, construida porque esa infraestructura beneficia tanto al gobierno como al sector privado.[121]

Los argumentos de pendiente resbaladiza se basan en cadenas de razonamiento del tipo "si-entonces" y se emplean para encadenar un hecho inicial incuestionable a corto plazo con un resultado éticamente objetable a largo plazo.

Los argumentos de pendiente resbaladiza pueden ser razonables o falacias. Cuando el encadenamiento causa-consecuencia depende de una probabilidad difícilmente calculable, posiblemente estamos ante una falacia.

En el caso del RF las consecuencias que se prevén son una sociedad con vigilancia ubicua, incompatible con los ideales de un orden libre y democrático, y que una vez desplegada es difícil eliminar. Sin duda hay aplicaciones del RF útiles y seguras como el sistema de Apple FaceID para desbloquear el móvil con el rostro. FaceID obtiene una plantilla de 30.000 puntos de la cara del propietario en 3D que se almacena en el dispositivo. Así que es una aplicación controlada y la información no es compartida, es segura. El peligro surge cuando el uso del rostro se generaliza como control de acceso al trabajo o al banco y fácilmente se pierde el control de lo que más nos identifica.

Bobby Domínguez, director de seguridad de la información de City Na-

121 Woodrow Hartzog, "Facial Recognition Is the Perfect Tool for Oppression", *Medium*, https://medium.com/s/story/facial-recognition-is-the-perfect-tool-for-oppression-bc2a08f0fe66

tional, dice que los teléfonos inteligentes que se desbloquean mediante un escaneo facial han allanado el camino. "Ya estamos aprovechando el reconocimiento facial en los dispositivos móviles", "¿Por qué no aprovecharlo en el mundo real?"[122].

Este es un ejemplo del paso de una aplicación inofensiva a otra controvertida. El gobernador de Nueva York Andrew Cuomo desplegó un sistema de vigilancia empezando con cámaras en los puestos de peaje para escanear placas de vehículos; luego fue sencillo conseguir la aprobación para agregar capacidades de RF a las cámaras y vincular el sistema a diversas bases de datos.

Selinger y Leong concluyen que el RF es una tecnología realmente única y proponen cuatro puntos argumentales, a los que añado uno más:

1. El rostro desempeña un papel esencial, existencial en la vida de los seres humanos y es casi inconcebible que existan sociedades humanas de cierto tamaño que no otorguen un valor extremadamente elevado al rostro descubierto. La cara es la parte más expresiva del cuerpo, incluso las microexpresiones son una parte poderosa del lenguaje. El rostro es el intermediario en la mayor parte de las interacciones con otros individuos. Los encuentros cara a cara se asocian habitualmente con inmediatez e intimidad. Cicerón dijo que *"la cara es el espejo del alma"*. Mediante el rostro nos identificamos, comunicamos e interactuamos socialmente, lo que explica porqué, con excepciones como el burka, en las sociedades contemporáneas se mantiene el rostro a la vista; llevarlo oculto puede sugerir que hay algo que ocultar.

2. Del rostro se puede extrapolar más información que de cualquier otra información biométrica: preferencias sexuales, estado de ánimo, probabilidad de decir la verdad, etc. Si realizar ese tipo de juicios ya es atrevido y frecuentemente cargado de prejuicios, pretender inferir esa información de forma automática no sólo es cuestionable, reforzará la discriminación y los sesgos sociales. El valor que tendría esa información en la aplicación de la ley, el marketing o la educación convierten el RF en la herramienta de análisis ideal.

3. El coste operativo del RF es inferior al de otras tecnologías de identificación biométrica, y no es necesario el contacto físico como en el caso del ADN y las huellas digitales. La vinculación de la identificación facial con otras bases de datos permite elaborar detallados perfiles de las personas, incluso en tiempo real.

---

122 Paresh Dave y Jeffrey Dastin, "U.S. banks deploy AI to monitor customers, workers amid tech backlash", *Reuters*, https://www.reuters.com/technology/us-banks-deploy-ai-monitor-customers-workers-amid-tech-backlash-2021-04-19/

4. Hay profundas lagunas legales en la regulación, lo que pone pocos límites a la tecnología. La escasez de salvaguardas permite usos permisivos, cuando no promiscuos. No está clara la legalidad de usar fotos de internet o sin consentimiento, no está clara la legalidad de utilizar servicios de terceros. Además no hay pretensión próxima de establecer una regulación común internacional.

5. La TRF es una tecnología abierta y barata, fácil de implementar mientras no se demande gran calidad. El software se encuentra disponible en internet, las tarjetas avanzadas de gráficos se pueden utilizar como procesadores de imágenes, y las listas de imágenes se pueden obtener de múltiples fuentes.

Si las TRF se normalizan y van ocupando espacios aparentemente inocuos, la expansión a otros ámbitos y usos más intrusivos puede ser imparable.

Frecuentemente los consumidores son bombardeados con anuncios y películas en las que el RF va acompañado de representaciones positivas y útiles en la vida diaria. "Si los ciudadanos esperan estar inmersos en la tecnología de reconocimiento facial dondequiera que vayan, podrían abrirse a permitir que las fuerzas del orden se comporten como todos los demás"[123].

El gobierno francés emitió un decreto el 10 de marzo de 2021 autorizando el análisis automatizado del uso de máscaras en el transporte público mediante cámaras inteligentes, obligatorio en Francia desde mayo de 2020. Previamente, la Autoridad Francesa de Protección de Datos (CNIL) había emitido un dictamen en el que afirmaba que "la captura y análisis sistemático de las imágenes de las personas en estos espacios sin duda conlleva riesgos para sus derechos y libertades fundamentales". La CNIL respalda el procesamiento de datos porque no involucra datos biométricos, ya que no identifican personas, y que no se utilizará con fines policiales. Aunque la CNIL coincide con el razonamiento del gobierno, de que la lucha contra el COVID-19 justifica la no inclusión del derecho a objetar, advierte de los riesgos que plantea el aumento de la vigilancia, la habituación y la banalización de las tecnologías intrusivas:

> Aunque se limite al marco del estado de emergencia sanitaria, tal despliegue presenta el riesgo real de generalizar un sentimiento de vigilancia entre los ciudadanos, de crear un fenómeno de habituación y banalización de las

123 Evan Selinger, *"Our Government Should Not Be Conducting Facial Surveillance", OneZero,* https://onezero.medium.com/our-government-should-not-be-conducting-facial-surveillance-54cc13f1ea61

tecnologías intrusivas y, en última instancia, de generar un aumento vigilancia[124].

Tras mostrar varios ejemplos recientes y cercanos de la aplicación de esta tecnología, parece un error pensar que es posible contener con seguridad los riesgos asociados con el RF o impedir su aplicación generalizada. Como dijo recientemente Laurie Anderson: "Si crees que la tecnología resolverá tus problemas, no comprendes la tecnología y no comprendes tus problemas".[125]

Es necesario poner orden en este campo, y la regulación es el mecanismo primordial. Afirma la profesora Kate Crawford:

> Ya no podemos permitir que las tecnologías de reconocimiento de emociones no estén reguladas. Es hora de que la legislación proteja contra los usos no probados de estas herramientas en todos los ámbitos: educación, atención médica, empleo y justicia penal. Estas salvaguardas deben ser una ciencia rigurosa y que rechace el mito de que los estados internos son solo otro conjunto de datos que se pueden sacar de nuestros rostros[126]

124 Team AI Regulation, "The French Government permits automated video surveillance of mask wearing on public transport following authorization by the CNIL", https://ai-regulation.com/the-french-government-permits-automated-video-surveillance-of-mask-wearing-on-public-transport-following-authorisation-by-the-cnil/

125 Según cita Kate Crawford en Twitter, https://twitter.com/katecrawford/status/1377551240146522115

126 Kate Crawford, "Time to regulate AI that interprets human emotions", *Nature 592, 167 (2021), doi:* https://doi.org/10.1038/d41586-021-00868-5

Regulación

En este capítulo veremos diversas propuestas de regulación de la TRF. Es importante tener en cuenta que la defensa de la igualdad y la justicia cuando se emplea IA en la toma decisiones depende de los algoritmos empleados. El profesor Andrés Boix argumenta que "los algoritmos empleados por parte de las Administraciones públicas para la adopción efectiva de decisiones han de ser considerados reglamentos por cumplir una función material estrictamente equivalente a la de las normas jurídicas, al reglar y predeterminar la actuación de los poderes públicos".[127]

Debido al confinamiento por el COVID-19, en 2020 los exámenes de secundaria y los previos a la universidad en Reino Unido fueron sustituidos por predicciones algorítmicas. El resultado fue que gran número de alumnos recibieron una nota muy inferior a la que esperaban. El clamor fue tal que se anularon esos resultados. Casi un año después, la Oficina de Estadísticas ha publicado un informe de una investigación interna. El informe dice

> Nos preocupa que los organismos públicos estén menos dispuestos a utilizar modelos estadísticos para respaldar decisiones en el futuro por temor a un rechazo público, lo que podría obstaculizar la innovación y el desarrollo de estadísticas y reducir los beneficios que pueden ofrecer[128].

El documento propone un conjunto de medidas para aumentar la confianza "*en los modelos estadísticos y en los algoritmos en el futuro*". Concluye así:

> Nuestra principal conclusión es que para lograr la confianza pública en los modelos estadísticos no basta solo el diseño técnico del modelo; tomar las decisiones y acciones correctas con respecto a la transparencia, la comunicación y la comprensión de la aceptabilidad pública durante todo el proceso de un extremo a otro es igualmente importante. También llegamos a la conclusión

127 Andrés Boix, "Los Algoritmos son Reglamentos: la necesidad de extender las garantías propias de las normas reglamentarias a los programas empleados por la Administración para la adopción de decisiones", *Revista de Derecho Público: Teoría y Método Vol. 1 | 2020 pp.223-270 Madrid, 2020*. https://doi.org/10.37417/RPD/vol_1_2020_33

128 Office for Statistics Regulation, "Ensuring statistical models command public confidence. Learning lessons from the approach to developing models for awarding grades in the UK in 2020", https://osr.statisticsauthority.gov.uk/publication/ensuring-statistical-models-command-public-confidence/

> de que [los organismos públicos aseguren] que los derechos de las personas sean plenamente reconocidos y que las responsabilidades sean claras

En consecuencia, el mayor reto que tendrá que afrontar la adopción de tecnologías basadas en IA es ganar la confianza de los ciudadanos. Por este motivo algunos de los ejemplos siguientes mencionan explícitamente el derecho del ciudadano a obtener una explicación del proceso.

La organización European Citizen's Initiative ha promovido la campaña *Reclaim your face*

> El reconocimiento facial puede y será usado en contra de cada uno de nosotros por gobiernos y corporaciones, según quiénes somos y cómo nos vemos.
>
> Recuperemos nuestro espacio público. ¡Prohibamos la vigilancia masiva biométrica![129]

En la regulación del RF encontramos tres tipos de respuestas. La primera es el enfoque garantista de los derechos individuales, aquí tenemos el caso de Canadá y el Consejo de Europa. Una segunda respuesta es la prohibición de uso de esta tecnología, consecuencia del rechazo de la ciudadanía causado por malas experiencias; aquí se sitúan las regulaciones estatales y locales de Estados Unidos. La tercera agrupa los que no dudan en utilizar el RF y fijan condiciones para su utilización, especialmente por las fuerzas del orden, son los casos de Reino Unido, Francia y la propuesta de regulación de la Unión Europea.

Canadá

En 2018, los profesores canadienses Geoffrey Hinton y Joshua Bengio junto con el francés Yann LeCunn, recibieron el Premio Turing por su trabajo en Aprendizaje Profundo. Canadá es un país muy desarrollado científica, social y políticamente en IA. Quizá es el motivo por el que antes de regular el RF considera que debe modificar la *Privacy Act* de Canadá. En 2021 ha publicado un documento con cuatro recomendaciones.[130]

Primera recomendación: incorporar explícitamente en la Privacy Act que

---

129 European Citizens' Initiative, "Reclaim your face", http://reclaimyourface.eu

130 Cybersecure Policy Exchange, "Facing the Realities of Facial Recognition Technology", *Recommendations for Canada's Privacy Act*, https://www.cybersecurepolicy.ca/frt-privacy-act

existe información personal relacionada con características físicas, biológicas, biométricas y también con la información facial.

La Privacy Act presenta tres problemas en relación a la información facial:

1. No reconoce que la información facial y biométrica es información personal extremadamente sensible

2. Las salvaguardas contra los riesgos y perjuicios asociados a la recopilación y uso de la información biométrica son inadecuadas

3. No aclara cuando las instituciones federales pueden recopilar y utilizar ese tipo de información

Para cubrir estos vacíos propone una consulta pública sobre un mayor respeto a los derechos de privacidad, mecanismos mayores de rendimiento de cuentas y mejorar la adaptación de las instituciones federales.

Segunda recomendación: establecer requisitos relacionados con la información facial para salvaguardar adecuadamente la privacidad y la seguridad de los canadienses. Estos requisitos deben proveer:

– Limitaciones en la recopilación, uso y divulgación de esta información, que debe requerir notificación y consentimiento o explícito permiso legislativo;

– Requisitos para minimizar la recopilación de información; y

– Requisitos de protección de seguridad más amplios.

Tercera recomendación: alinear la Privacy Act con los requisitos de la Directive on Automated Decision Making. Esta alineación impondría más términos específicos para su uso por parte de las fuerzas del orden, como garantizar la notificación pública, pruebas de sesgo, formación de empleados, evaluaciones de riesgos de seguridad y necesidad de que un humano tome la decisión final en el caso de decisiones de alto impacto.

Cuarta recomendación: implementar una moratoria federal sobre usos nuevos y ampliados de RF automatizado y la divulgación de información facial hasta que:

– El marco descrito en esta presentación haya sido desarrollado en consulta a los ciudadanos canadienses, así como instituciones gubernamentales y servidores públicos en departamentos gubernamentales; y

– Se realicen más investigaciones sobre los impactos desproporcionados, o potencialmente desproporcionados, en miembros de grupos demográficos, particulares a las realidades y poblaciones en Canadá.

No deja de ser llamativo que Canadá, un país avanzado en IA y muy interesado en aplicar la TRF, proponga una moratoria. Posiblemente el objetivo

perseguido sea la confianza de los ciudadanos en una tecnología y prevenir problemas involucrando a los ciudadanos en su aprobación.

Tras la publicación del borrador de reglamento de la IA de la Unión Europea han surgido críticas por la escasa referencia en el plan canadiense a los derechos humanos. "El reglamento de la UE está plagado de debates sobre derechos. Los derechos ocupan el primer lugar en la lista de evaluación del daño por la IA [...]. La IA no debe utilizarse para socavar la dignidad humana y debe respetar la vida privada de los datos personales. Ese lenguaje falta en Canadá, y el proyecto de ley C-11 solo menciona como derecho humano la privacidad".[131]

Consejo de Europa

A finales de enero de 2021, el Consejo de Europa publicó unas directrices sobre RF que recogen recomendaciones legislativas sobre este tema.[132]

Define qué es el RF y la naturaleza especialmente sensible de los datos biométricos. Contempla el caso de RF en vivo o de imágenes cuando la cooperación del individuo no es requerida.

Recuerda que, según el Artículo 6 de la Convención 108+, el reconocimiento biométrico solo debe ser autorizado si tiene fundamento legal y proporcionado.[133]

Indica que los marcos legales deben indicar en concreto:

– una explicación detallada del uso y propósito específico
– la fiabilidad y precisión mínimas del algoritmo usado
– el periodo de retención de las fotos empleadas
– la posibilidad de auditar estos criterios
– la trazabilidad del proceso
– las salvaguardas

El RF tiene un alto nivel de intrusividad ya que despoja del derecho a la privacidad y la dignidad de los individuos y existe un riesgo de impacto adver-

131 Fenwick McKelvey y Jonathan Roberge, "Canada is gambling with its leadership on artificial intelligence", *The Globe and Mail,* https://www.theglobeandmail.com/business/commentary/article-canada-is-gambling-with-its-leadership-on-artificial-intelligence/

132 Council of Europe, "Guidelines on Facial Recognition", https://rm.coe.int/guidelines-on-facial-recognition/1680a134f3

133 Council of Europe, "Amending Protocol CETS No. 223 to Convention 108", https://search.coe.int/cm/Pages/result_details.aspx?ObjectId=09000016807c65bf

so en otros derechos y libertades fundamentales. Por ello, recomienda que el empleo del RF en entornos no controlados debe estar sujeto al debate democrático. También contempla la posibilidad de declarar una moratoria mientras se realiza un análisis completo del impacto de esta tecnología.

Propone que la utilización del RF con el único propósito de determinar el color de la piel, creencias religiosas o de otro tipo, origen racial o étnico, edad, estado de salud o condición social debería ser prohibido, a menos que se establezcan las debidas salvaguardas y se evite cualquier riesgo de discriminación.

El *reconocimiento de afectos*[134] a partir de imágenes podría posiblemente extraer rasgos de personalidad, sentimientos internos, salud mental o implicación del trabajador. El empleo de afectos en procesos de selección, valoración de primas de seguros o acceso a la educación, puede plantear riesgos muy preocupantes a nivel individual y social, por lo que debería ser prohibido.

Sugiere que la regulación del RF debería tener en cuenta:

– todas las fases de proceso, incluyendo la creación de las bases de datos y el despliegue

– los sectores en los que se usa esta tecnología

– la intrusividad de los tipos de RF, en vivo o no, indicando claramente directrices sobre su legalidad

El documento previene también la extracción de plantillas biométricas[135] o su integración en sistemas biométricos sin una base legal para un nuevo proceso cuando las imágenes habían sido obtenidas para otro propósito. En concreto, las imágenes de internet o las capturadas por cámaras de vigilancia no pueden considerarse con base legal para su proceso solo porque hayan sido puestas explícitamente accesibles por el usuario. Así mismo, la información que se utilice debe ser la estrictamente requerida para el proceso y no toda la disponible.

El consentimiento no debe ser, como regla general, la base legal para la utilización del RF en el sector público dado el desequilibrio de poder entre el ciudadano y las autoridades, incluso si es llevado a cabo por entidades privadas en su nombre. Recomienda que el uso de RF con el propósito de identifi-

134 El *reconocimiento de afectos* es una expresión más amplia que el *reconocimiento de emociones* ya que incluye la extracción de información de otras formas de expresión personal como la vocal o el movimiento corporal

135 Plantilla biométrica es una representación digital de características únicas que han sido extraídas de una muestra biométrica y se almacena en una base biométrica

cación, tanto en entornos controlados como no controlados, quede reservado a las fuerzas del orden. Siempre debe respetarse una estricta necesidad y proporcionalidad en los sistemas de identificación en entornos no controlados, tanto durante la creación de la base de datos como en el despliegue del sistema. El uso de RF encubierto por las fuerzas del orden se considera legítimo si es estrictamente necesario y proporcionado para prevenir un riesgo sustancial e inminente de la seguridad pública, que debería quedar documentado previamente al uso encubierto.

En cuanto al uso del RF en otras aplicaciones de interés público distintas del cumplimiento de la ley, el documento deja abierta la puerta, siempre que haya necesidad, proporcionalidad, que las imágenes se obtengan legalmente, que haya salvaguardas, un cajón donde casi cualquier aplicación es justificable.

El uso de RF en el sector privado debe contar con el consentimiento del interesado, y debe emplearse solo para autenticación o verificación. También deben ofrecerse alternativas no engorrosas a los usuarios. Recomienda prohibir el uso en entornos no controlados como centros comerciales, especialmente para identificar personas de interés, con fines de marketing o de seguridad privada. El paso por una zona señalizada no se debe considerar como consentimiento explícito.

Las Directrices recomiendan que se establezcan mecanismos que garanticen la responsabilidad de quienes desarrollen, fabriquen o provean servicios, y entidades que empleen sistemas de RF.

Se recomienda que se adopten medidas específicas de *compliance* para los desarrolladores en cuanto a calidad de los datos, precisión, diversidad, seguimiento de la precisión, corrección de fallos, renovación periódica de los datos y señalética adecuada para los usuarios.

Para las entidades privadas que utilicen RF, puesto que se puede aplicar sin intención o cooperación de los individuos, es de vital importancia transparencia y rectitud. Por ello, se debe dar información clara y completa de qué información se captura, el contexto, su uso, eliminación y cómo les afectará, así como hasta qué punto se va a compartir información con terceros. También se debe informar de los derechos y recursos que les asisten.

Recomienda que se proporcionen marcos éticos para el uso del RF debido a los riesgos inherentes a su uso en ciertos sectores. Igualmente, apoya la creación de comités de expertos de diferentes campos de experiencia para definir los casos potencialmente más difíciles.

Concluye mencionando que se deben garantizar los derechos individuales, especialmente el derecho de información, acceso, explicación, oposición y rectificación.

Cuando el RF se pretenda utilizar para la toma de decisiones basadas exclusivamente en procesos automatizados que puedan afectar significativamente al sujeto, éste debe tener derecho a rechazar ese proceso si no se tiene en cuenta su punto de vista. Esto es también aplicable en sistemas de reconocimiento en vivo en los que el operador actúa teniendo en cuenta solamente los resultados automatizados.

Estados Unidos de América

La regulación del RF en Estados Unidos se está realizando a nivel estatal y local, principalmente prohibiciones en diverso grado. Estos son algunos ejemplos.

San Francisco, California, fue la primera gran ciudad que prohibió el RF en mayo de 2019.[136] Este paso, tomado por una ciudad muy tecnológica, envió un mensaje importante a todo el país. La prohibición impide a las agencias públicas de la ciudad usar TRF o información obtenida de sistemas externos que usan esa tecnología.

En septiembre de 2020, Portland, Oregón, aprobó la legislación más radical hasta ese momento.[137] Amazon había tratado de influir en el debate para impedir su aprobación.[138] Se aprobaron dos ordenanzas relacionadas con el RF: la primera ordenanza prohíbe la adquisición y uso de TRF por cualquier oficina de la ciudad de Portland. La segunda prohíbe a las entidades privadas utilizar TRF en lugares de alojamiento público de la ciudad.[139] Ambas orde-

136 Board of Supervisors San Francisco, "Stop Secret Surveillance Ordinance (05/06/2019)", https://www.eff.org/document/stop-secret-surveillance-ordinance-05062019

137 City of Portland, "City Council approves ordinances banning use of face recognition technologies by City of Portland bureaus and by private entities in public spaces", https://www.portland.gov/smart-city-pdx/news/2020/9/9/city-council-approves-ordinances-banning-use-face-recognition

138 Kate Kaye, "Amazon Is Quietly Fighting Against a Sweeping Facial Recognition Ban in Portland", *OneZero*, https://onezero.medium.com/amazons-quietly-fighting-against-a-groundbreaking-facial-recognition-ban-in-portland-f0d1e3c2054

139 City of Portland, "Chapter 34.10 Prohibit the use of Face Recognition Technologies by Private Entities in Places of Public Accommodation in the City of Portland", https://www.portland.gov/code/34/10

nanzas sostienen que la TRF tiene un impacto dispar en las comunidades desfavorecidas, particularmente las personas de color y las personas con discapacidad.

En diciembre de 2020, el Gobernador Cuomo firmó una ley para suspender el uso de TRF y otros tipos de tecnología biométrica en las escuelas, tanto privadas como públicas, del Estado de Nueva York. La ley impuso una moratoria a las escuelas en la compra y uso de tecnología de identificación biométrica hasta al menos el 1 de julio de 2022 o hasta que el Comisionado de Educación del Estado autorice su uso.[140]

Minneápolis, Minnesota, aprobó en febrero de 2021 una ordenanza contra el RF.[141] La ordenanza, que prohíbe a los empleados de la ciudad adquirir o usar sistemas de RF externos, no llega tan lejos como la prohibición de Portland, que también prohíbe su uso por parte de empresas privadas, pero impide que el departamento de policía pueda hacerlo a través de terceros, como la Oficina del Sheriff del Condado de Hennepin. El motivo argumentado fue las preocupaciones sobre su confiabilidad y potencial para perjudicar a las comunidades de color.

Massachusetts es un caso distinto. En febrero de 2021 ha aprobado una regulación que entrará en vigor en julio del mismo año.[142] Massachusetts se ha esforzado en poner salvaguardas, como que la policía debe obtener permiso de un juez antes de realizar una búsqueda mediante RF. La ley también crea una comisión para hacer recomendaciones tales como si un acusado debe ser notificado de que fue identificado mediante RF.

Francia

En noviembre de 2020 se publicó el *Livre blanc de la sécurité intérieure* que trata de los desafíos de la seguridad interna y medios para afrontar-

140 "Governor Cuomo Signs Legislation Suspending Use and Directing Study of Facial Recognition Technology in Schools", *Governor Andrew M.Cuomo*, https://www.governor.ny.gov/news/governor-cuomo-signs-legislation-suspending-use-and-directing-study-facial-recognition

141 Libor Jani, "Minneapolis passes restrictive ban on facial recognition use by police, others", *Star Tribune*, https://www.startribune.com/minneapolis-passes-restrictive-ban-on-facial-recognition-use-by-police-others/600022551/

142 The Commonwealth of Massachusetts, "An Act to regulate face surveillance", https://trackbill.com/bill/massachusetts-senate-docket-2134-an-act-to-regulate-face-surveillance/2041134/

la.[143] El RF es uno de ellos. Este documento es un libro blanco y, por tanto, describe la visión y las intenciones del Ministerio del Interior acerca del RF.

La primera propuesta de aplicación de esta tecnología es la identificación de sospechosos. Reconoce que la aplicación más inmediata es la ciencia forense.

En segundo lugar, considera que la experimentación del RF en espacios públicos es "*altamente deseable*" para su prueba técnica, operativa y legal. Estas pruebas deben realizarse en un plazo establecido de tiempo, de forma transparente y con control judicial de las listas de seguimiento.

Finaliza recordando que todo lo anterior se apoya en un informe publicado por el CNIL, la autoridad de protección de datos, en 2019, en línea con el marco europeo RGPD. Este informe marca tres requerimientos al enfoque experimental del RF:

1. Fijar líneas rojas previas a cualquier uso experimental. El procesamiento biométrico debe ser legítimo, proporcional e indispensable, especialmente en el empleo del reconocimiento en vivo.

2. Respetar los derechos y la privacidad de las personas, así como el consentimiento, transparencia y que la información sea comprensible y accesible.

3. Adoptar un enfoque genuinamente experimental, es decir, científicamente riguroso, multidisciplinar y con plazos razonables.

Reino Unido

En 2019 la Court of Appeal de Londres declaró ilegal el uso que había hecho la policía de South Wales mediante vigilancia con cámaras. Así, en noviembre de 2020 el Surveillance Camera Commissioner elaboró un informe con nuevas directrices para el uso del RF.[144] Tiene seis partes:

1. Biométrica, Igualdad y Ética.

2. Derechos Humanos, 'De acuerdo con la ley' y Marco Legal.

3. Gobernanza, Aprobación, Listas de seguimiento, Características Protegidas y el Tomador de Decisiones Humano.

---

143 Ministère de L'Intérieur, "Livre blanc de la sécurité intérieure", https://www.interieur.gouv.fr/Actualites/L-actu-du-Ministere/Livre-blanc-de-la-securite-interieure

144 Surveillance Camera Commissioner, "Facing the Camera", https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/940386/6.7024_SCC_Facial_recognition_report_v3_WEB.pdf

4. Integridad, uso de Materiales como prueba y manipulación de los Materiales.

5. Participación pública, suministro de información, desempeño.

6. Responsabilidad y certificación.

Tras el análisis, el informe incluye once recomendaciones en un anexo. Fundamentalmente se trata de la creación por el Home Office de un conjunto de reglas y procedimientos relativos a los sistemas de RF, para adquisición, evaluación de riesgos, toma de decisiones, supervisión ética policial, indicadores de eficacia y la creación de una terminología policial nacional consistente.

Enumera los tipos de personas que deben estar en las listas de seguimiento de RF. Señala que se debe hacer el despliegue no encubierto informando adecuadamente, aunque admite el uso encubierto si se considera necesario y proporcionado, y que la decisión debe tener fundamento sólido y no sólo por conveniencia operativa.

La citada decisión de la Court of Appeal identificó fallos de protección de datos porque

> falló al evaluar adecuadamente los riesgos de los derechos y libertades de los sujetos de los datos y falló en abordar las medidas previstas para abordar los riesgos.

Por esto, el informe reafirma la necesidad de realizar una evaluación de impacto de protección de datos previo previo al despliegue de la tecnología, y así pide la creación de políticas disponibles públicamente que establezcan las salvaguardas para la discreción policial.

El fallo de la Court of Appeal puso de relieve que la intervención humana para prevenir sesgos previa al uso de un algoritmo no es suficiente, es necesaria la monitorización humana durante la operación para revisar los resultados.

Un comunicado de junio de 2021 de Information's Comissioner Office de Reino Unido expresa preocupación por el uso inapropiado, excesivo o incluso invasivo del reconocimiento facial en tiempo real respecto a la privacidad. Pide a las compañías e instituciones una adecuada comprensión y evaluación de los riesgos del uso de estas tecnologías y su impacto en las vidas y privacidad de las personas.[145]

145 "Blog: Information Commissioner's Opinion addresses privacy concerns on the

Unión Europea

En abril de 2021 se publicó la propuesta de regulación de la inteligencia artificial en la Unión Europea.[146] El objetivo es desarrollar una regulación para generar confianza en los ciudadanos acerca del uso de la IA, a la vista de los beneficios que se pueden derivar de ella. Con estas normas se pretende que los sistemas de IA utilizados en la UE sean seguros, transparentes, éticos e imparciales y estén bajo control humano.

La propuesta expresa reiteradamente que busca la defensa de los derechos protegidos por la Unión Europea. Quizás por este motivo, la aproximación de la propuesta de Reglamento se basa en los riesgos inherentes a casos de uso (no tanto a sectores concretos). Y se pretende implantar un control preventivo (ex ante), pero ello no evita que se pueda realizar también un control "represivo" (ex post, a posteriori), y que si se producen daños se deba responder por ellos.

Ámbito de aplicación

Al igual que ocurrió con el RGPD, Europa quiere convertirse en referente regulatorio en esta materia y para ello establece su aplicación más allá del ámbito de la Unión Europea. En concreto, se dispone que se aplicará a los que pongan en servicio sistemas de IA en la UE, aunque el proveedor esté establecido en un tercer país (art. 2.1).

Sistemas de IA prohibidos

El art. 5.2 fija las condiciones que deben cumplir los sistemas de identificación biométrica remota "en tiempo real" en espacios de acceso público con el fin de hacer cumplir la ley, en los casos excepcionales en los que se permite su uso. Mientras que el art. 5.3 somete estos supuestos excepcionales en los que se permite el uso de identificación biométrica remota en tiempo real a una autorización judicial o de autoridad administrativa independiente.

En definitiva, solo se prohíben ciertos sistemas de identificación biométrica (identificación remota en tiempo real en espacios públicos por los poderes

---

use of live facial recognition technology in public places", *Information's Comissioner Office* , https://ico.org.uk/about-the-ico/news-and-events/news-and-blogs/2021/06/information-commissioner-s-opinion-live-facial-recognition-technology/

146 "Proposal for a Regulation on a European approach for Artificial Intelligence", *Comisión Europea* https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=75788

públicos con fines de hacer cumplir la ley), pero se admite que por vía legal los Estados puedan establecer ciertas excepciones (tasadas) también para estos casos en principio prohibidos, estableciendo determinadas garantías (como la autorización judicial o el establecimiento de límites adecuados desde el punto de vista de la duración, el alcance geográfico y las bases de datos exploradas).

Sistemas de IA de alto riesgo

Supone el grueso de la regulación de la propuesta y a la que más atención se dedica. Se incluyen aquí casos en que se aplica la IA en productos que ya estaban sometidos a una legislación específica porque podían entrañar un riesgo (por ej., seguridad de los juguetes, equipos médicos…) y se añaden casos nuevos listados en el Anexo III. En él se incluyen muchos de los supuestos "de periódico" que más preocupan en la actualidad.

Sistemas de riesgo reducido

Se incluyen aquí otros casos, como la IA destinada a interactuar con personas (por ej., los chatbot) o los sistemas de reconocimiento de emociones o de categorización biométrica (que no entren en los casos de alto riesgo), y para los que únicamente se establece obligaciones de transparencia, esto es, que el usuario sea consciente de que está interactuando con una máquina, o de que están expuestos a uno de estos sistemas de reconocimiento de emociones o categorización biométrica, para que pueda decidir si continúa o no. Esto no se aplica en los sistemas de IA utilizados para la categorización biométrica previstos en la ley con fines policiales.

Sistemas de IA de riesgo mínimo o nulo

Son los restantes no incluidos en las categorías anteriores y que la propuesta de Reglamento no regula. No obstante, podrían adherirse voluntariamente a un Código Autorregulatorio.

Valoración

Hay que destacar el esfuerzo de la UE por convertirse en el primer regulador de la IA y en su vocación de universalidad, en la medida en que pretende que se convierta en un conjunto mínimo regulatorio para la IA a nivel mundial.

La mayor parte de los comentarios publicados coinciden en que la dirección adoptada en la propuesta es correcta, pero ciertamente es susceptible de

mejora, tiene limitaciones, lagunas y en algunos casos, como en el RF es poco ambiciosa.

Nuestra valoración, por su novedad, cercanía e impacto previsible, incluye no sólo lo relacionado con el reconocimiento biométrico, sino toda la propuesta.

El Libro Blanco de la IA hacía una promoción indiscriminada e injustificada de la IA, especialmente para el sector público. La adopción de la IA no debería ser un objetivo ni es un valor en sí misma.[147]

El Libro Blanco de la IA proponía la confianza como medio imprescindible para lograr la aceptación de la IA. Esta propuesta pone su enfoque en minimizar los riesgos y por eso se centra en establecer reglas sobre usos de la IA de "riesgo alto" y "prohibidos" y mecanismos para mitigar los riesgos.

A lo largo del texto parece claro que la norma es favorable a las organizaciones, tanto públicas como privadas, que utilicen o desarrollen IA. Mientras éstas deben pasar evaluaciones y autorizaciones, las personas apenas tienen protección o reparación por perjuicios que puedan sufrir. Debe tenerse en cuenta que las personas pueden estar sujetas a sistemas de IA no opcionales, por ejemplo ante la administración pública.

Privacidad y Supervisión Humana

La propuesta pretende regular determinados usos de la información personal. Sobre este tema, incluimos a continuación la posición conjunta del comité y del supervisor europeo de protección de datos sobre esta propuesta.

Un punto de referencia, aunque no explícito en muchos casos, es el reglamento europeo de protección de datos (RGPD). La propuesta indica que los sistemas de riesgo alto, como el reconocimiento facial, deben estar diseñados para permitir la supervisión humana. La supervisión humana es un concepto recurrente en el RGPD. Aunque parece tranquilizador, en realidad provoca una falsa sensación de comodidad y distrae de los peligros reales.[148]

Hay tres razones principales por las que la supervisión humana no es adecuada.

---

147 Daniel Leufer, "Access Now submission to the EU consultation on the AI white-paper", *Access Now*, https://daniel-leufer.com/2020/06/12/access-now-submission-to-the-eu-consultation-on-the-ai-whitepaper/

148 Ben Green y Amba Kak, "The False Comfort of Human Oversight as an Antidote to A.I. Harm", *Slate*, https://slate.com/technology/2021/06/human-oversight-artificial-intelligence-laws.html

Primera, pedir supervisión solamente crea una débil protección que las compañías y gobiernos pueden esquivar fácilmente. El RGPD establece en el considerando (71) que "el interesado debe tener derecho a no ser objeto de una decisión [...] que se base únicamente en tratamiento automatizado". La palabra "únicamente" establece la cuestión como algo binario, blanco o negro. La supervisión de los sistemas IA podríamos decir que abarca una amplia gama de grises. Aunque los titulares periodísticos anuncian grandes avances del tipo "la IA es capaz de...", la realidad es que es infrecuente que se prescinda completamente de la supervisión humana en los sistemas de IA. En el caso de sistemas de riesgo alto bastaría con añadir una supervisión superficial para cumplir los requisitos normativos.

Segunda, lograr una supervisión efectiva en la práctica es difícil e impreciso. Los sistemas de evaluación de riesgo prejudiciales en EEUU pretendían reducir la discriminación. Sin embargo, los operadores son proclives a confiar en la respuesta de los sistemas automatizados sin escrutinio adecuado, es lo que se conoce como "sesgo automatizado". Esto ha provocado que estos sistemas hayan empeorado las diferencias raciales en vez de reducirlas.[149]

La tercera razón es que asumir la supervisión humana como elemento clave para remediar los daños de la IA puede conducir a difuminar la responsabilidad y desviarla a los operadores de primera línea, cuando estos tienen poco o nulo control.

Es necesario reconocer que los resultados discriminatorios y perjudiciales de los sistemas de IA no son un "bug", son una "característica". En consecuencia, en vez de pedir una supervisión humana, debería considerarse si determinado sistema debería usarse a toda costa y pedir mayor responsabilidad a los individuos e instituciones que toman las decisiones. Se pueden promocionar sistemas con capacidades sobrehumanas y cuando surgen problemas se pretende resolverlos con supervisión.

La postura adecuada sería que los reguladores pasen de buscar parches regulatorios e implicación humana significativa a realizar las preguntas correctas: ¿Quienes en concreto interactúan con el sistema? ¿Hay incentivos desalineados que puedan limitar la capacidad de evaluación y anticipación de problemas? ¿Hasta qué punto pueden los algoritmos alterar la decisión humana?

---

149 Tom Simonite, "Algorithms Were Supposed to Fix the Bail System. They Haven't", *Wired*, https://www.wired.com/story/algorithms-supposed-fix-bail-system-they-havent/

Derechos Humanos

La propuesta clasifica un sistema de alto riesgo si amenaza la salud o los derechos fundamentales de una persona, lo que incluye sistemas de identificación biométrica, sistemas para la salud, servicios al ciudadano y para aplicación de la ley.

Los sistemas de IA de alto riesgo están permitidos, de acuerdo con el enfoque basado en riesgo, sujetos a ciertos requisitos obligatorios y evaluaciones de conformidad "ex ante".

El punto 5.2.3 del Memorandum de la propuesta dice: "la clasificación como de alto riesgo no solo depende de la función que realiza el sistema de IA, sino también del propósito y las modalidades específicas para las que se utiliza ese sistema".

La realidad es que cuando un sistema de IA es perjudicial a las personas, resulta difícil probar cuál es su función y cuál es el propósito. Frecuentemente, las personas ni siquiera saben que han sido perjudicadas. Muchas veces se alega secreto industrial, violación de privacidad o seguridad nacional para evitar dar detalles.

Conjuntos de Datos

> Para el desarrollo de sistemas de IA de alto riesgo [...] deben poder acceder y utilizar conjuntos de datos de alta calidad (Memorandum 2.3).

¿Qué significa exactamente "alta calidad"? Diferentes actores pueden tener ideas muy diferentes de calidad; por ejemplo la administración pública y los ciudadanos no están de acuerdo en lo que es la "calidad".

> Los conjuntos de datos de formación, validación y ensayo deben ser relevantes, representativos, libres de errores y completos (Considerando 44).

Es necesario aclarar "relevante" para quién y "representativo" de qué.

Ni la calidad de los datos, la relevancia ni la representatividad, por muy acreditadas que estén, pueden convertir un caso de uso equivocado en aceptable.

Probar cualquiera de estas características en la práctica es irreal. El algoritmo BOSCO diseñado para adjudicar el bono eléctrico social ha sido cuestionado, pero la solicitud de acceso al código ha sido reiteradamente denegada por "atentar contra la propiedad intelectual".[150]

150 "Que se nos regule mediante código fuente o algoritmos secretos es algo que jamás

En la mayoría de los casos de uso de alto riesgo, los desarrolladores pueden realizar las evaluaciones de conformidad por sí mismos, personas que persiguen también otros objetivos como coste, consumo, tiempo de desarrollo y normativa exigente. La evaluación de conformidad debería realizarse sin otros condicionantes.

Prohibiciones light y Reconocimiento Facial

Las prohibiciones cubren prácticas que tienen un potencial significativo para manipular a las personas mediante **técnicas subliminales** más allá de su conciencia o explotar las **vulnerabilidades de grupos vulnerables.** La propuesta también prohíbe la **puntuación social por las autoridades públicas**. Por último, también se prohíbe el uso de sistemas de **identificación biométrica remota en "tiempo real" en espacios de acceso público con el fin de hacer cumplir la ley**, a menos que se apliquen ciertas excepciones limitadas.

Se prohíbe la explotación mediante "técnicas subliminales" en "grupos vulnerables". Si no se emplean "técnicas subliminales" no es un caso prohibido para los grupos vulnerables, a pesar de que el impacto en ellos de la IA será mayor por su desprotección.

La prohibición de identificación biométrica es fácilmente sorteable. Si se introduce un retardo en la cadena de identificación, se puede justificar que no es "tiempo real", y no impide la identificación masiva de imágenes archivadas.

> **El uso de sistemas de identificación biométrica** remota en tiempo real en espacios de acceso público con fines de aplicación de la ley [...] **tendrá en cuenta** los siguientes elementos: **naturaleza de la situación** que da lugar al posible uso, en particular la gravedad, **probabilidad y escala del daño causado en ausencia del uso del sistema**; las **consecuencias del uso del sistema** [...], **deberá cumplir las salvaguardias y condiciones necesarias** y proporcionadas en relación con el uso, en particular en lo que respecta a las **limitaciones temporales, geográficas y personales** (Art. 5.2).

La infraestructura de identificación no se puede poner y quitar, debe estar permanentemente instalada. Cuando se dispone de esa infraestructura lo

---

debe permitirse en un Estado social, democrático y de Derecho", *Civio*, https://civio.es/novedades/2019/07/02/que-se-nos-regule-mediante-codigo-fuente-o-algoritmos-secretos-es-algo-que-jamas-debe-permitirse-en-un-estado-social-democratico-y-de-derecho/

normal es usarla. Es muy improbable afirmar "tengo instaladas cámaras y la capacidad de identificación pero las mantengo apagadas".

En cuanto a la autorización previa puede convertirse en un mero trámite. Como ya hemos dicho anteriormente, los jueces tras el 11-S fueron bastante permisivos en la autorización de operaciones de vigilancia.

Desprotección del ciudadano

Existe un gran desequilibrio de poder entre quienes desarrollan y aplican la IA y las personas que la sufren. La reparación a las personas cuando se vean afectadas negativamente por la tecnología es una cuestión ausente en el texto de la propuesta. Además, la mayoría de las personas carece de recursos para reclamar.

Aunque en Derecho la carga de la prueba corresponde al demandante, la asimetría de poder entre demandante y demandado podría recomendar invertir la carga de la prueba[151] o buscar otros mecanismos menos gravosos a las personas y más efectivos.

Ausencias

Debido a la naturaleza de la tecnología y su impacto social, hay tres cuestiones que echamos en falta en la propuesta.

Primera, el impacto laboral de algunos sistemas de IA, no solo los de alto riesgo, puede ser importante. Existe un riesgo real de daño que se debe prever, evaluar, evitar, si no es posible minimizar y si ocurre reparar.

La propuesta se limita a considerar los procesos de selección y evaluación laboral. Creemos interesante mencionar que las palabras "empleo", "trabajo", "trabajador" (work, job, worker, employee) brillan por su ausencia en la redacción. La utilización de IA en el mundo laboral va muchísimo más allá de esas dos aplicaciones concretas; tiene potencial para penetrar en gran cantidad de tareas y eso afectará de forma importante al lugar de trabajo[152].

Segunda, el impacto medioambiental de esta tecnología. Como hemos mostrado anteriormente, las necesidades energéticas de la IA pueden ser

151 Sebastian Klovig Skelton, "Europe's proposed AI regulation falls short on protecting rights", *Computer Weekly*, https://www.computerweekly.com/feature/Europes-proposed-AI-regulation-falls-short-on-protecting-rights

152 José Alberto Rodríguez Ruiz, "New artificial intelligence regulations have important implications for the workplace", *Insight*, https://workplaceinsight.net/new-artificial-intelligence-regulations-have-important-implications-for-the-workplace/

enormes. Debería incluirse una evaluación medioambiental y quizás una justificación comparada con otras soluciones. Esto debería incluir todo el ciclo de vida del sistema, para cualquier nivel de riesgo.

Tercera, la diversidad.

> Los proveedores de sistemas de IA **que no son de alto riesgo** pueden crear e implementar los códigos de conducta ellos mismos. Esos códigos también **pueden incluir compromisos voluntarios** relacionados, por ejemplo, con la **sostenibilidad ambiental, la accesibilidad para las personas con discapacidad, la participación de las partes interesadas en el diseño y desarrollo de sistemas de IA y la diversidad de equipos de desarrollo** (Memorandum 5.2.7).

Multitud de expertos señalan la falta de diversidad como una de las causas detrás de la discriminación en sistemas de IA, y demandan una amplia participación de las partes interesadas en todo el ciclo de vida de sistemas con IA.[153] El texto, sin embargo, considera la cuestión opcional y voluntaria.

Posición conjunta del supervisor europeo de protección de datos

En junio de 2021, las autoridades europeas de protección de datos, European Data Protection Supervisor (EDPS) y European Data Protection Board (EDPB), publicaron una posición conjunta sobre la propuesta europea de regulación de IA.[154] Consideramos que es una contribución muy valiosa y reproducimos a continuación el resumen ejecutivo, tal como aparece en el documento. Las negritas y cursivas son tal como aparecen en el documento original.

Resumen ejecutivo

El 21 de abril de 2021, la Comisión Europea presentó su Propuesta de Reglamento del Parlamento Europeo y del Consejo por el que se establecen normas armonizadas sobre inteligencia artificial (en adelante, "la Propuesta"). El EDPB

153 Cristina Lago, "Voluntary frameworks will not protect against algorithmic bias", *Tech Monitor*, https://techmonitor.ai/ai/voluntary-frameworks-will-not-protect-against-ai-bias

154 "Joint Opinion 5/2021 on the proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act)", *EDPB-EDPS*, https://edps.europa.eu/system/files/2021-06/2021-06-18-edpb-edps_joint_opinion_ai_regulation_en.pdf

y el EDPS acogen con satisfacción la preocupación del legislador por abordar el uso de la inteligencia artificial (IA) en la Unión Europea (UE) y destacan que la propuesta tiene **implicaciones de protección de datos** muy importantes.

El EDPB y el EDPS señalan que la **base jurídica** de la propuesta es, en primer lugar, el artículo 114 del Tratado de Funcionamiento de la Unión Europea (TFUE). Además, la propuesta también se basa en el artículo 16 del TFUE en la medida en que contiene normas específicas sobre la protección de las personas con respecto al tratamiento de datos personales, en particular restricciones del uso de sistemas de inteligencia artificial para sistemas de identificación biométrica remota en 'tiempo real' en espacios de acceso público con el propósito de hacer cumplir la ley. El EDPB y el EDPS recuerdan que, de acuerdo con la jurisprudencia del Tribunal de Justicia de la UE (TJUE), el artículo 16 del TFUE proporciona una base jurídica adecuada en los casos en los que la protección de datos personales es uno de los objetivos o componentes esenciales del normas adoptadas por el legislador de la UE. La aplicación del artículo 16 del TFUE también implica la **necesidad de garantizar una supervisión independiente del cumplimiento** de los requisitos relativos al tratamiento de datos personales, como también lo exige el artículo 8 de la Carta de los Derechos Fundamentales de la UE.

En cuanto al **alcance de la propuesta**, el EDPB y el EDPS acogen con gran satisfacción el hecho de que se extienda al suministro y uso de sistemas de inteligencia artificial por parte de las instituciones, organismos o agencias de la UE. Sin embargo, la **exclusión de la cooperación policial internacional del ámbito** de aplicación de la propuesta plantea serias preocupaciones para el EDPB y el EDPS, ya que dicha exclusión crea un riesgo significativo de elusión (por ejemplo, terceros países u organizaciones internacionales que operan aplicaciones de alto riesgo en las que se apoyan autoridades públicas de la UE).

El EDPB y el EDPS **acogen favorablemente el enfoque basado en el riesgo** que sustenta la propuesta. Sin embargo, conviene aclarar este enfoque y alinear el concepto de "riesgo para los derechos fundamentales" con el RGPD y el Reglamento (UE) 2018/1725 (EUDPR), ya que entran en juego aspectos relacionados con la protección de datos personales.

El EDPB y el EDPS están de acuerdo con la propuesta cuando establece que la clasificación de un **sistema de inteligencia artificial como de alto riesgo no significa necesariamente que sea legal** per se y que el usuario pueda implementarlo como tal. Es posible que el responsable del tratamiento **deba cumplir otros requisitos derivados de la ley de protección de datos de la UE**. Además, el cumplimiento de las obligaciones legales derivadas de la legislación de la Unión (incluida

la protección de datos personales) debería ser una condición previa para poder entrar en el mercado europeo como producto con la marca CE. A tal fin, el EDPB y el EDPS consideran que **el requisito de garantizar el cumplimiento del RGPD y la EUDPR debe incluirse en el capítulo 2 del título III**. Además, el EDPB y el EDPS consideran necesario adaptar el procedimiento de evaluación de la conformidad de la propuesta para que los terceros siempre realicen evaluaciones *ex ante* de la conformidad de los sistemas de IA de alto riesgo.

Dado el gran riesgo de discriminación, la propuesta prohíbe la "puntuación social" cuando se realiza 'durante un determinado período de tiempo' o 'por las autoridades públicas o en su nombre'. Sin embargo, las empresas privadas, como las redes sociales y los proveedores de servicios en la nube, también pueden procesar grandes cantidades de datos personales y realizar evaluaciones sociales. En consecuencia, **el futuro Reglamento de IA debería prohibir cualquier tipo de puntuación social**.

La identificación biométrica remota de personas en espacios de acceso público plantea un alto riesgo de intrusión en la vida privada de las personas, con graves efectos en la expectativa de la población de permanecer en el anonimato en espacios públicos. Por estas razones, el EDPB y el EDPS **piden una prohibición general de cualquier uso de IA para el reconocimiento automático de rasgos humanos en espacios de acceso público**, como rostros, pero también de marcha, huellas dactilares, ADN, voz, pulsaciones de teclas y otros elementos biométricos o señales de comportamiento - en cualquier contexto. Se recomienda igualmente la **prohibición** de los sistemas de inteligencia artificial que **clasifiquen a las personas según la biometría en grupos** según su origen étnico, género, orientación política o sexual u otros motivos de discriminación en virtud del Artículo 21 de la Carta. Además, el EDPB y el EDPS consideran que el uso de IA para **inferir las emociones de una persona física es muy indeseable y debería prohibirse**.

El EDPB y el EDPS acogen con satisfacción **la designación del EDPS como autoridad competente y autoridad de vigilancia del mercado para la supervisión de las instituciones, agencias y organismos de la Unión**. Sin embargo, conviene aclarar más el papel y las tareas del EDPS, específicamente en lo que respecta a su papel como autoridad de vigilancia del mercado. Además, el futuro Reglamento sobre IA debería establecer claramente la **independencia de las autoridades de supervisión** en el desempeño de sus tareas de supervisión y ejecución.

La designación de las autoridades de protección de datos (APD) como autoridades nacionales de supervisión garantizaría un enfoque reglamentario más armonizado, contribuiría a la interpretación coherente de las disposiciones sobre tratamiento de datos y evitaría contradicciones en su aplicación entre los Estados

miembros. En consecuencia, el EDPB y el EDPS consideran que **las autoridades de protección de datos deben ser designadas como autoridades nacionales de supervisión de conformidad con el artículo 59 de la propuesta**.

La propuesta asigna un papel predominante a la Comisión en el "Consejo Europeo de Inteligencia Artificial" (EAIB). Este papel entra en conflicto con la necesidad de que un organismo europeo de IA sea independiente de cualquier influencia política. Para garantizar su independencia, el futuro Reglamento de IA debería otorgar **más autonomía al EAIB** y garantizar que pueda actuar por iniciativa propia.

Teniendo en cuenta la difusión de los sistemas de inteligencia artificial en el mercado único y la probabilidad de casos transfronterizos, existe una necesidad crucial de una aplicación armonizada y una asignación adecuada de competencias entre las autoridades nacionales de supervisión. El EDPB y el EDPS sugieren contemplar un mecanismo que garantice **un punto de contacto único para las personas afectadas por la legislación y para las empresas, para cada sistema de IA**.

Por lo que se refiere a las **zonas de pruebas**, el EDPB y el EDPS **recomiendan aclarar su alcance y objetivos**. La propuesta también debe establecer claramente que la base legal de tales entornos de pruebas deben cumplir con los requisitos establecidos en el marco de protección de datos existente.

El **sistema de certificación** descrito en la propuesta **carece de una relación clara con la ley de protección de datos de la UE**, así como con otras leyes de la UE y de los Estados miembros aplicables a cada 'área' del sistema de IA de alto riesgo y no tiene en cuenta **los principios de minimización de datos y protección de datos por diseño** como uno de los aspectos a tener en cuenta **antes de obtener el marcado CE**. Por tanto, el EDPB y el EDPS recomiendan modificar la propuesta para aclarar la relación entre los certificados emitidos en virtud de dicho Reglamento y las certificaciones, precintos y marcas de protección de datos. Por último, las APD deben participar en la preparación y el establecimiento de normas armonizadas y especificaciones comunes.

En cuanto a los códigos de conducta, el EDPB y el EDPS consideran **necesario aclarar** si la protección de datos personales debe ser considerada entre los "requisitos adicionales" que pueden ser abordados por estos códigos de conducta, y asegurar que las "especificaciones y soluciones técnicas" no entran en conflicto con las normas y principios del actual marco de protección de datos de la UE.

Conclusión

Hemos presentado las tecnologías relacionadas con el reconocimiento facial realizadas con IA, y ha quedado establecido que los modelos creados con IA son buenas aproximaciones para problemas complejos, pero una caja negra en cuanto a la capacidad de explicar los resultados y de describir qué aprenden los modelos. Los modelos creados con IA pueden ocultar sesgos y discriminar. Se han descrito graves problemas al usar el RF como detenciones equivocadas debidas a identificación errónea realizada con reconocimiento facial.

La IA se está utilizando también para crear modelos de detección de emociones, inferir el estado de ánimo a partir de imágenes de rostros. Esta aplicación es altamente controvertida, ya que la ciencia que respalda esa equivalencia es altamente dudosa.

Hay una creciente crítica científica a la naturaleza de la inteligencia artificial, completamente distinta de la inteligencia humana, que en realidad apenas conocemos. El aprendizaje automático es incremental, mientras que el razonamiento por sentido común es de otra naturaleza. Esto implica serias limitaciones en su aplicación como sustituto de la inteligencia humana. Existe una creciente preocupación por el perjuicio laboral y el impacto medioambiental causado por la IA.

La aplicación de la IA a imágenes de rostros implica la recopilación masiva de información personal, lo que supone una amenaza para la privacidad, autonomía y libertad de los ciudadanos y para la democracia liberal.

Los filósofos de las tecnologías alertan del impacto de la IA en la vida humana. Los objetivos perseguidos no se alinean con una vida humana plena y floreciente. Escribe el profesor Tasioulas: "la tecnología no es neutral. Las tecnologías son formas de construir valores humanos en el mundo". Mediante el aprendizaje automático, la vida, enriquecida por la experiencia y la diversidad, pretende ser sustituida por elecciones "optimizadas". Un peligro para la libertad y la autonomía. En el mejor de los casos, un aburrimiento.

Muchas autoridades científicas sostienen que la normalización del RF desencadenará una preocupante cadena de perjuicios personales y sociales. Puesto que el rostro es la parte más peculiar e identificable de una persona, su ocultación supondría un cambio social importante.

El empleo de tecnologías de reconocimiento facial, reconocidamente expuestas a sesgos, pueden discriminar a las personas por rasgos físicos y resul-

tar en violación de los derechos humanos. Quizás ha sido una de las causas de que IBM, Google y Amazon hayan decidido adoptar una moratoria de esta tecnología en 2020.

Finalmente presentamos cómo se está regulando la tecnología en Canadá. Francia, Reino Unido, Estados Unidos, las recomendaciones del Consejo de Europa y la propuesta de regulación de la IA en Europa.

La propuesta europea propone una regulación un poco vaga y permisiva de hecho, basada en riesgos. La regulación no debería centrarse meramente en soluciones técnicas o de manipulación de datos, sino en desarrollar procesos socio-técnicos y la responsabilidad corporativa, para asegurar que cualquier resultado discriminatorio o injusto sea evitado o, al menos, mitigado.[155]

La posición de las autoridades europeas de protección de datos es decididamente contraria al uso del reconocimiento facial y resalta la estrecha relación con los datos personales y la obligación de protección de la privacidad como se recoge en diversas normas comunitarias.

El futuro es impredecible, pero cuanto más tiempo se tarde en percibir lo que supone la IA en la sociedad, un debate más profundo que la tecnología y los reglamentos, más difícil será controlar una tecnología que otorga tanto poder para modelar la sociedad.

155 Virginia Dignum, "What we need to talk about when we talk about AI", *LinkedIn*, https://www.linkedin.com/pulse/what-we-need-talk-when-ai-regulatory-purposes-virginia-dignum/

Índice general