ArtefaCToS. Revista de estudios de la ciencia y la tecnología
eISSN: 1989-3612
Vol. 10, No. 2 (2021), 2.ª Época, 21-43
DOI: https://doi.org/10.14201/art20211022143

Bajo observación: inteligencia artificial, reconocimiento facial y sesgos

Under Observation: Artificial Intelligence, Facial Recognition and Biases

Tomás BALMACEDA[*]; Tobías SCHLEIDER[**]; Karina PEDACE[***]

[*] Universidad de Buenos Aires, / IIF/ SADAF – CONICET, Argentina
tomasbalmaceda@gmail.com

[**] UNS-UNMDP-ILSED, Argentina
tSchleider@gmail.com

[***] UBA/UNLaM/ IIF SADAF – CONICET, Argentina
karinapedace@gmail.com

Recibido: 06/11/2020. Revisado: 11/05/2021. Aceptado: 09/07/2021

Resumen

En este trabajo nos concentramos en la tecnología de reconocimiento facial y hacemos hincapié en que en las últimas décadas se instaló en el discurso público la idea de que este tipo de tecnología es "objetiva" y carente de los errores y sesgos humanos, adjudicándole cualquier responsabilidad al "buen o mal uso" que se haga de ella. Frente a esta retórica, argumentamos que a la hora de hablar de artefactos tecnológicos la neutralidad es imposible, aun en la instancia misma del diseño. A tal efecto, a partir de las ideas del pragmatismo norteamericano analizamos un caso real, para mostrar que no es posible sostener una dicotomía entre hechos y valores y que es necesario reconocer que no existe una distinción tajante entre el diseño de la tecnología y el uso que hacemos de ella.

Palabras clave: filosofía de la tecnología; algoritmo; vigilancia; reconocimiento facial; sesgos.

Abstract

In this paper we concentrate on facial recognition technology and emphasize that in recent decades, the idea that this type of technology is "objective" and devoid of human errors and biases has been installed in public discourse, assigning it any responsibility to the "good or bad use" that is made of it. Faced with this rhetoric, we argue that when it comes to talking about technological artifacts, neutrality is impossible, even in the very instance of design. To this end, based on the ideas of North American pragmatism, we analyze a real case to show that it is not possible to sustain a dichotomy between facts and values, and that it is necessary to recognize that there is no sharp distinction between the design of technology and the use that we make of it.

Keywords: philosophy of technology; algorithm; surveillance; facial recognition; biases.

1. Introducción

Entre los problemas más espinosos que deben afrontar las sociedades contemporáneas, el de la seguridad tiene el raro privilegio de ser uno de los abordados con más superficialidad. En palabras de Alberto Binder,

> (…) la sociedad fluctúa entre la sensación de peste y el mesianismo que promete una salvación milagrosa, sin advertir que ambos extremos forman parte de una misma actitud frente al problema: *un conservadurismo carente de ideas, poco dispuesto a profundizar en el análisis y menos aún dispuesto a arriesgar en el diseño de políticas complejas que nos permitan enfrentar un fenómeno social también complejo y multifacético*. (Binder, 2009)

En este sentido, prevalece la idea de que todo conflicto —en especial, todo conflicto vinculado con la seguridad— es una alteración inaceptable de un orden (ideal). La política de seguridad no es más, entonces, que una política de restablecimiento del orden.

Ahora bien, un control democrático de la criminalidad y las violencias fundado en la idea de orden es problemático en varias dimensiones. En la política, porque va en contra de una visión democrática de la seguridad[1]. En la pragmática, porque se cae con facilidad —y, como se verá, con el auspicio esmerado de la tecnología— en la falacia de que una sociedad más vigilada es una sociedad más segura.

Es que la noción de seguridad ha cambiado más en las últimas décadas que en los últimos siglos. De referirse a la defensa de intereses centrales de los Estados,

[1] Así, Binder (2009), quien propone reemplazar el "paradigma del orden" por el de la gestión de la conflictividad, que evite la resolución de los conflictos por la fuerza o favoreciendo a quien ostente, en algún sentido, mayor poder.

como la autonomía y la soberanía, ha pasado a ocuparse del ser humano y la comunidad que lo contiene y, en especial, al desarrollo de ambos. Hoy involucra mucho más que la protección, por parte del Estado, de la integridad física y patrimonial de las personas. Comprende aspectos que hacen a la convivencia pacífica y al goce pleno de derechos diversos. En consecuencia, se presta atención a los fines, pero también a los medios[2].

En la primera sección de este trabajo, rescataremos la caracterización de la "sociedad de la exposición" de Bernard Harcourt para presentar un análisis del escenario actual, en el que la vida cotidiana está cruzada por vigilancia y tecnología y en donde se han resignificado valores como el de la privacidad o la intimidad. Luego, en la segunda sección, intentaremos trazar una genealogía del vínculo establecido entre los rostros humanos y la seguridad, a partir de las posibilidades que ofrecieron distintos artefactos técnicos, desde retratos en el siglo XIX hasta los actuales algoritmos de aprendizaje automatizado que prometen identificar a sospechosos. Nos detendremos en la tecnología de reconocimiento facial señalando cómo, frente a otras técnicas de reconocimiento biométrico disponibles, fue elegida de manera casi unánime no por su eficacia o por el estado de desarrollo en el que se encontraba, sino porque su uso prescinde del consentimiento de los involucrados. Además, haremos hincapié en cómo en las últimas décadas se instaló en el discurso público la idea de que este tipo de tecnología es "objetiva" y carente de los errores y sesgos humanos, adjudicándole cualquier responsabilidad al buen o mal uso que se haga de ella. Finalmente, en la tercera sección, argumentaremos por qué este pretendido halo de ausencia de valores es insostenible. Nos apoyaremos en las ideas de Hilary Putnam para mostrar que no es posible sostener una dicotomía entre hecho y valor, sino que es necesario reconocer que no existe distinción tajante entre el diseño de la tecnología y el uso que hacemos de ella. Nuestra propuesta será que, a la hora de hablar de artefactos tecnológicos, la neutralidad es imposible.

2. Vigilancia y exposición

La vigilancia ocupa cada vez más lugar en nuestras vidas. Y, para ocuparlo, está quitándole su espacio a la privacidad. Este no es un proceso que comenzó hace poco, pero se ha acelerado de manera exponencial en los últimos lustros.

[2] El Programa de las Naciones Unidas para el Desarrollo ha destacado el pasaje de la centralidad de la seguridad pública al de la seguridad humana, que define como la "condición de vivir libre de temor y de necesidad frente a amenazas al desarrollo humano —desde, por ejemplo, desastres ambientales o carencias alimentarias hasta violencias y delitos—" (PNUD, 1995). En este marco, la seguridad ciudadana —noción a la que aludimos centralmente en este trabajo—, que comprende a las violencias (sean o no consideradas delitos por los Estados), los delitos y el temor al crimen (la sensación o sentimiento de inseguridad, más allá del riesgo actual o potencial corrido), ha pasado a ser una condición necesaria, aunque no suficiente, de la seguridad humana.

En parte, por el mero avance de los recursos tecnológicos, aunque mayormente por cómo están transformando esos avances a la sociedad que nos contiene y a nosotros como individuos.

Cada cosa que hacemos en la nueva era digital puede ser registrada, almacenada y monitoreada. Y nosotros estamos poniendo todo de nosotros para *facilitar* ese proceso. El abogado y profesor norteamericano Bernard Harcourt comienza con esas dos ideas su libro *Exposed. Desire and Disobedience in the Digital Age* (2015, 1, 14, *passim*), y las desarrolla a lo largo de sus páginas (y de obras anteriores; esp. Harcourt 2007 y Harcourt 2011). Los planteos de Harcourt son interesantes porque cuestionan, de manera fina e inteligente, las explicaciones tradicionales de la vigilancia en la era digital. En lo que sigue, rescataremos algunos de los tópicos que aborda en la medida en que nos resultan útiles para nuestros propósitos (que, es crucial aclarar, no son los suyos).

Harcourt sostiene que "Para muchos de nosotros, la existencia digital se ha convertido en nuestra vida, el pulso, el flujo sanguíneo, la corriente de nuestras rutinas diarias" (2015, 18). En este sentido, rescata la definición "astuta" de Kevin Haggerty y Richard Ericson (2000, 611): nos hemos convertido en *Homo digitalis*. Y esta nueva criatura se inserta en una sociedad que ha tratado de calificarse de modos diversos, y de explicarse con metáforas atractivas. Las propuestas de Harcourt en los dos sentidos aspiran a ser superadoras.

Así, Guy Debord (1967) habló de la "sociedad del espectáculo". Michel Foucault (2013), a partir de la idea de sociedad panóptica o disciplinaria, configuró la noción de "sociedad punitiva". Gilles Deleuze (1992), por su parte, se refirió a las "sociedades del control". Harcourt propone, en cambio, el título de "sociedad de la exposición" (2015, 19 y *passim*). En ella, "… una nueva forma de poder … inserta [el castigo] en nuestros placeres diarios. Los dos, placer y castigo, ya no pueden ser separados. Se recubren uno al otro, operan juntos. Se han vuelto inextricablemente entrelazados" (2015, 21). Por otra parte, agrega, "El ideal liberal —que puede existir un dominio protegido de autonomía individual— ya no tiene tracción en un mundo en el cual el comercio no puede ser distinguido de la gobernanza, el policiamiento o la vigilancia…" (2015, 26).

Más allá de la terminología y en su empresa de despeje conceptual, Harcourt analiza, antes que nada, la propiedad de las metáforas que se emplean para dar cuenta de esta nueva sociedad de la exposición. Y refiere que ellas son o bien inadecuadas, o bien insuficientes. Que han quedado desfasadas con respecto a la realidad cambiante a la cual quieren referir. Algunos autores comparan el monitoreo constante y extendido de nuestras vidas por parte de agencias nacionales e internacionales y de corporaciones con la amenaza del Gran Hermano y los avatares de la novela de George Orwell *1984* (1949). Otros, en cambio, trazan un parangón con la "caja negra" que Franz Kafka ubicó en

el centro de la burocracia infinita de su novela *El Proceso* (1925). Otros más, abrevan en la idea de panóptico que Michel Foucault tomó de Jeremy Bentham (1791). Finalmente, algunos combinan estos símbolos (Harcourt, 2015, 27).

Orwell fue profético en muchos sentidos: con respecto a la tecnificación y la ubicuidad de la vigilancia; en la idea de la criminalización de los pensamientos; en la manera en que castigamos (al menos, respecto de las modalidades de castigo durante varias décadas posteriores a la publicación de su libro más conocido), esto es, con la pretensión de transformar al sujeto del castigo, de curarlo. Pero, para Harcourt, falló en captar el núcleo que define a la era digital. En *1984*, el objetivo de los poderosos era aplastar y erradicar el deseo. En cambio, hoy es precisamente a través de nuestras pasiones, deseos e intereses la vía por la cual nos hemos convertido en transparentes y, de este modo, nos exponemos a la vigilancia omnipresente. No solo la del Estado y la policía, sino la de corporaciones privadas, medios de comunicación, vendedores, organismos no gubernamentales, países extranjeros y nuestro prójimo. En palabras de Harcourt, "Son nuestras pasiones —en tensión con nuestras dudas, ansiedades e incomodidades, naturalmente— las que alimentan a la sociedad de la exposición" (Harcourt, 2015, 34-37)[3].

Las ideas de Orwell fueron, en efecto, brillantes y visionarias. No obstante, siguiendo a Harcourt, parece que, si alguien diseñó la nueva era de la vigilancia que nos toca vivir, aprendió del error de *1984*. La supresión del placer y, en última instancia, del deseo, fue reemplazada por su exacerbación. Los estímulos para su persecución son los que provocan el exhibicionismo, la vida *hacia afuera*, las prácticas que ponen nuestra información al alcance de quien quiera tomarla. Esta sería la manera en la cual la vigilancia funciona hoy, en las sociedades democráticas: a través de los deseos más elementales, que son curados para nosotros, nos son recomendados. Por ejemplo, los jóvenes activistas son, en muchos sentidos, incentivados en su expresividad, no reprimidos (Harcourt, 2015, 39-48), con el objeto de que muestren sus ideas, que las saquen a la superficie. Dice Harcourt:

No, no vivimos en un mundo orwelliano, soso. Vivimos en un mundo digital hermoso, colorido, estimulante, que está conectado, enchufado, enlazado, en línea y conectado al wifi. Un mundo rico, brillante, vibrante, lleno de pasión y

[3] En el siglo XIX, los datos eran generados por el Estado a través de estadísticas, registros vitales, informes de censos, gráficos militares. Producir información era costoso. Hoy todos somos nuestros propios publicistas. La producción de datos se ha democratizado. El costo lo pagamos nosotros, de manera directa y no monetaria (Harcourt, 2015, 140). Al respecto, deviene interesante la noción de "sociedad de la transparencia" acuñada por Byung-Chul Han (2015). El filósofo coreano-alemán postula que la sociedad contemporánea es transparente, y la califica de positiva (despejada de negatividad), expuesta (todo es medido por su valor de exhibición), evidente (nada queda reservado a lo estimulante del secreto), pornográfica (mostrada obscenamente en su desnudez), acelerada (calculadora antes que pensante), íntima (en cuanto reveladora de la interioridad), informada (antes que iluminada; inmanente y falta de poética), reveladora (todo queda expuesto a la vista de los demás) y controlada (en un panóptico global e interconectado).

de disfrute —medios a través de los cuales nos revelamos y nos convertimos en virtualmente transparentes a la vigilancia. … Vivimos en un mundo que ha rectificado el error del Gran Hermano (Harcourt, 2015, 56).

La segunda metáfora que Harcourt analiza es la del "Estado de vigilancia". Un Estado que se presenta como mentiroso, artero, embaucador, que tergiversa y exacerba, que decora la realidad y amenaza nuestras libertades individuales[4]. Para una posición conservadora, por el contrario, el Estado de vigilancia no tergiversa, no miente ni engaña: es nuestro protector desalmado[5]. Entre las que proponen liberales y conservadores,[6] se presenta una tercera imagen del Estado de vigilancia que lo considera un regulador o administrador[7]. Esta visión se enfoca en una "gestión de riesgos" de varios tipos: "riesgos que corre la seguridad nacional", pero también el comercio y el libre cambio, las relaciones internacionales, la privacidad y la libertad, y los derechos humanos. De la mano de estas ideas viene la de que esos riesgos han de gestionarse no sobre la base de intuiciones y anécdotas, sino luego de considerar datos y evidencia. Se promueve, así, la transparencia, la rendición de cuentas, la aspiración de ganar la confianza pública; el gobierno *de los civiles*, no de los militares; un Estado *abierto*, no secretista. Un Estado de vigilancia, pero *protector*. La metáfora es obviamente correcta en ciertos sentidos, pero también inadecuada y poco pertinente en otros. Es insuficiente hablar de un Estado vigilante, porque no se trata solo del Estado (sin que esto signifique restarle importancia al poder estatal, sino, tal vez, al contrario). Como señala Harcourt, quizás el mote de "complejo militar-industrial" que acuñó el presidente Dwight Eisenhower en su discurso de clausura de mandato fuese más ajustado, pero aún no capta todo lo que sucede en torno de la vigilancia en la era digital (Harcourt, 2015, 55 y ss.).

El recurso al panóptico de Bentham, sobre todo en la versión de Foucault, es también común para ilustrar la realidad de la vigilancia contemporánea. Pero se queda corta: hoy somos menos vigilados que lo que nos exponemos, por propia voluntad y con toda dedicación. Sabemos que nos estamos desnudando ante una "oligarquía que practica el voyerismo digital" y nos ponemos a su merced. Nos enfrentamos con menos frecuencia a la vigilancia que al aprovechamiento de nuestro exhibicionismo. Por eso, ya no somos una sociedad del espectáculo, ni una sociedad punitiva. Antes, somos una sociedad de la exposición. Esta afirmación no debe interpretarse como una minimización del elemento punitivo, ni las formas omnipresentes de represión (policiamiento, encarcelamiento) de las

[4] Harcourt se refiere aquí a la posición del juez Richard Leon (liberal, aunque nombrado a propuesta del presidente republicano George W. Bush).

[5] Harcourt evoca la posición del juez William H. Pauley III (conservador, aunque nombrado a propuesta del presidente demócrata Bill Clinton).

[6] En el texto hacemos referencia a "liberales" y "conservadores" norteamericanos. Al respecto puede verse, por todos, Caplow *et al.*, 1994.

[7] Para ilustrar esta posición, Harcourt refiere a "los asesores" del presidente Barack Obama.

democracias, y las más incontrolables de las dictaduras. En cambio, Harcourt resalta con estas ideas que la metáfora del panóptico falla, sobre todo, porque no percibimos de manera constante que estamos siendo observados (Harcourt, 2015, 80 y ss.)[8]. Antes bien, muchos de nosotros ni siquiera sabemos que estamos siendo acechados. Harcourt rescata el término "criptóptico", de Siva Vaidhyanathan (2011), para bautizar la situación nueva. O, apelando a una metáfora arquitectónica actualizada, sugiere imaginar una suerte de pabellón vidriado que nos deja ver a los demás, nos expone y, a la vez, nos refleja. Estamos siendo monitoreados con más intensidad que en un encierro, y aun en una situación de vigilancia digital tradicional (mediante tobilleras electrónicas, seguimiento satelital o recursos similares). Las tecnologías digitales replican la intrusividad y el modo de ejercer el poder del monitoreo carcelario: un aparato omnipresente que todo lo ve (Harcourt, 2015, 107 y ss.; 239 y ss.), pero sin su materialidad (que, en última instancia, permite una detección más sencilla y, en potencia, la posibilidad de evitación y de evasión).

El nacimiento de la sociedad de la exposición ha venido de la mano de una erosión gradual de los valores análogos que alguna vez atesoramos: la privacidad, la autonomía, cierto anonimato, el secretismo, la dignidad, la habitación propia, el derecho a la soledad. Nos tratamos de convencer de que no tenemos de qué preocuparnos si no tenemos nada que ocultar. Que la transparencia nos beneficiará y que, en todo caso, nuestra información se perderá en la masa de datos, como una aguja en un pajar. Pero del otro lado, al estar expuestos, ser observados, grabados y tratados como el objeto de predicciones, nuestra subjetividad ha comenzado a ser moldeada por las nuevas tecnologías digitales. La incapacidad de controlar nuestra información íntima y la percepción creciente de ser perseguidos (o acechados) refuerzan nuestra sensación de vulnerabilidad. El *ser digital* está tomando el lugar de la existencia física analógica: es más permanente, durable, y hasta tangible y demostrable. En el extremo opuesto, la prisión analógica va

[8] Se han expuesto otras maneras de destacar un pretendido desfase entre la idea del panóptico, tal como la rescata Foucault, y la situación actual del control social, que podría estar dejando atrás no solo a la omnipresencia de la vigilancia, sino a la propia vigilancia. Pat O'Malley, por ejemplo, sostiene —aunque con el eje puesto en la crítica a ciertas tecnologías de prevención delictual que exceden el núcleo de este trabajo— que, en la actualidad, la atención se desplazó del individuo que delinque, sus motivaciones y sus condicionantes, hacia las situaciones que propician que el individuo delinca. En otras palabras, se deja de atender a las causas sociales y se enfoca a las circunstancias materiales: la ocasión, la oportunidad. Los enfoques biográfico-causales dejan paso a los blancos objetivos del delito. El ofensor se construye como un ente abstracto, universal y racional, libre para cometer el delito o no cometerlo, abiográfico. Este pensamiento, que O'Malley califica de "neoconservador", desligaría al individuo de la sociedad y, por lo tanto, eximiría a la sociedad de la responsabilidad por el delito. El ofensor sería, así, merecedor del castigo: se lo habría buscado. Por lo tanto, el carácter punitivo del castigo ganaría legitimidad frente a la alternativa de la corrección (e. g., O'Malley, 1992). De esto parece seguirse que, para una posición como la comentada, la aplicación de la idea de panóptico a la noción contemporánea de vigilancia carecería de utilidad, ya que no sería crucial atender a las personas sino a las situaciones: al ambiente de las víctimas potenciales, para modificarlo y obstaculizar los planes espurios de los ofensores.

deslizándose, lentamente, hacia formas de supervisión digital (Harcourt, 2015, 166, 181, 217, 237 y ss.). Y ya no se necesita que el individuo sea consciente de que es vigilado, porque porta un adminículo adosado a su pierna o aun un teléfono móvil: el seguimiento recurre a insumos que difícilmente podamos dejar atrás, como los latidos de nuestro corazón (e. g., Israel e Irvine, 2012; Ramli et ál., 2016; Nait-ali, 2020, esp. 17-44 y 45-69) o nuestros rostros.[9] Quizás la consecuencia peor de este proceso complejo sea que la *ilusión* de libertad es cada vez mayor.

3. Videovigilancia y reconocimiento facial

Si bien el vínculo entre los rostros y la búsqueda de la seguridad se remonta a muchos años atrás, una confluencia de factores precipitó su ligazón en la forma actual que conocemos bajo el nombre de tecnología de reconocimiento facial (TRF). La identificación de "delincuentes" o "sospechosos" con las fotografías de sus caras se volvió popular a mediados del siglo XIX, cuando Allan Pinkerton, un fabricante de barriles escocés que vivía en Chicago, llegó a la portada de los periódicos al descubrir un elaborado complot de falsificación de alcohol. Con este triunfo, su fama de investigador creció. Así, decidió fundar la Agencia Nacional de Detectives de Pinkerton, en la que impartía clases acerca de métodos que él mismo había creado y otros sobre los que había leído o escuchado. Pronto tuvo un equipo de trabajo que gozó de reconocimiento popular gracias a noticias coloridas en las que se contaban sus victorias y hazañas. Aunque la sistematización de diferentes formas de trabajo que habían conseguido era, en efecto, original, lo cierto es que su eficacia también se explicaba porque, al ser un organismo paraestatal, podía perseguir criminales más allá de las fronteras estaduales y sin preocuparse por cumplir con la ley al pie de la letra. Convertido en una suerte de estrella mediática, Pinkerton llamó la atención de las autoridades. Durante la Guerra Civil de los Estados Unidos fue convocado para organizar un servicio secreto dentro del ejército, y reunió información sobre defensas, suministros y rutas de transporte del enemigo.

Una vez finalizado el conflicto, Pinkerton volvió a su trabajo como detective privado y perfeccionó una de sus técnicas más efectivas para dar con el paradero de una persona *sospechosa*: recolectar datos de testimonios e imágenes de recortes de los diarios o archivos familiares y armar archivos con información personal. En el caso de que lo creyera necesario, llenaba un pueblo de volantes con el rostro de la persona buscada. Su agencia adoptó como signo de identificación un ojo abierto con el lema "*We Never Sleep*", que rápidamente se popularizó y derivó en el uso de *private eye* como sinónimo de detective privado en el lenguaje coloquial de los Estados Unidos (Petersen, 2001, cap. 1). Para finales de 1870, este escocés tenía la colección más grande de imágenes de supuestos delincuentes,

[9] También, a través de una combinación de latidos y rostros (e. g., Medina, 2019).

que comenzaron a ser conocidas como "*mugshots*"[10] o "fotos de prontuario", una práctica que había comenzado en Bélgica e Inglaterra algunos años antes, y que cobró fuerza a finales de ese siglo, con la inversión, por parte de departamentos de policía de las principales ciudades de los Estados Unidos y Europa, en la tecnología necesaria para tomar daguerrotipos de los criminales. La estandarización de su formato, con un plano de frente y dos planos de costado, llegó en 1888 de la mano del francés Alphonse Bertillion, un policía pionero en las investigaciones biométricas (Rhodes, 1956, 27).

Si bien los *mugshots* siguieron vigentes, en el siglo XX creció el interés por las huellas dactilares como herramienta de identificación de las personas, una técnica desarrollada por el antropólogo argentino de origen croata Juan Vucetich y que fue utilizada con éxito por primera vez en 1892 (Vucetich, 1904). En 1905, el Departamento de Justicia de los Estados Unidos creó un registro unificado federal de huellas digitales de todas las personas que eran apresadas, con o sin condena firme, una práctica que se fue extendiendo por el globo. Mientras que las dos guerras mundiales fomentaron otras formas de vigilancia —vinculadas con las ondas de radio, el radar, los rayos ultravioletas y los rayos X— el uso de huellas digitales se mantuvo como un método económico y confiable para identificar, e incluso se vio potenciado por la aparición de las primeras computadoras en la década del 70. La aparición de las tecnologías antecesoras de lo que hoy conocemos como Internet, como las redes ARPANET y MILNET, también abrieron nuevas posibilidades de comunicación para compartir información y registros.

A finales del año 2001, luego de los cuatro atentados terroristas suicidas cometidos el 11 de septiembre en los Estados Unidos mediante el secuestro de aviones comerciales, creció la preocupación por la seguridad a nivel mundial, junto con voces que señalaron que las técnicas usuales de identificación de las personas eran insuficientes. En una presentación realizada en el Congreso de los Estados Unidos el 14 de noviembre de ese año, y transmitida en vivo por las principales cadenas de noticias de ese país, la senadora Dianne Feinstein pronunció un fuerte discurso en el que aseguró que la tragedia se hubiese evitado con la implementación de tecnología que había sido desarrollada en el país y ya estaba disponible, la TRF. Sus palabras se dieron en el marco de la audiencia pública "Identificación biométrica y la nueva cara del terror", de la que participaron los miembros del subcomité de tecnología, terrorismo e información de gobierno. Esto tuvo

[10] En inglés, mug significa recipiente o vaso cilíndrico, usado generalmente para beber cerveza. A fines del siglo XIX, era común que las tabernas sirvieran sus bebidas en mugs tallados con rostros humanos grotescos. Desde allí, el nombre se convirtió en una forma de llamar a los rostros, en particular a los de personas consideradas marginales. Combinada con shot, en su acepción de "toma fotográfica", la palabra se popularizó, a mediados del siglo XX, para denominar a las fotos policiales de los delincuentes, que engrosaban los mugbooks o libros de prontuarios (véase, e. g., Online Etimology Dictionary, 2020).

un alto impacto en el discurso público y popularizó la idea que la TRF era una manera confiable y mucho más eficiente de identificar a cualquier persona que intentara ingresar a los Estados Unidos (Feinstein, 2001).

A efectos de clarificar la inserción de la TRF en el vasto ámbito de la así llamada "inteligencia artificial", resulta pertinente trazar la siguiente precisión conceptual. De manera más o menos estándar se suele hablar, por un lado, de inteligencia artificial "general" en relación con el ambicioso proyecto consistente en crear un artefacto que, en todos los aspectos, actúe y piense de una manera humana (al modo de los replicantes de la película *Blade Runner)*. Por otro lado, se denomina inteligencia artificial "estrecha" a aquella que trata de buscar sistemas que realicen una tarea particular de forma inteligente. Puesto que parece claro que el primer proyecto sigue estando confinado al ámbito de la ciencia ficción, en lo que sigue nos centraremos en la inscripción de la TRF al interior del proyecto más acotado, que estriba en replicar artificialmente ciertas capacidades específicas humanas.[11]

Los primeros intentos para enseñarle a dispositivos a "ver rostros humanos" se dieron en 1960, pero recién comenzaron a dar resultados interesantes hacia el final del siglo XX, cuando la tecnología permitió que distintos sistemas reconocieran determinados objetos en fotografías y videos. El desarrollo de la TRF no se trató de un programa unificado, sino más bien de un campo interdisciplinario de investigación y de un conjunto de experimentos tecnológicos que compartía el interés por conseguir herramientas automatizadas de percepción facial. Hasta la audiencia en el Congreso de los Estados Unidos en noviembre de 2001, el foco no estaba puesto en la identificación de las personas a partir de su rostro, sino que científicos y académicos parecían más interesados en explorar las expresiones de sentimientos y emociones para lograr formas más sofisticadas de interacción entre personas y computadoras. La irrupción de actores inesperados en el área, como las agencias de seguridad y las fuerzas militares, cambió el foco hacia la posibilidad de crear desarrollos de la vigilancia que pudiesen generar sistemas de monitoreo con menor aporte humano, bajo la promesa de obtener "sistemas de vigilancia que funcionen de manera más efectiva, extendiendo su alcance en el tiempo y el espacio" (Gates, 2011, 1).

En paralelo con estos desarrollos vinculados con la seguridad y los requerimientos de las autoridades, algunos años más tarde la TRF irrumpió en plataformas y dispositivos muy populares, familiarizando a muchas personas con sus posibilidades. En 2010, Facebook desplegó en todos los perfiles de sus usuarios

[11] En este sentido, nos restringiremos a la consideración de sistemas de aprendizaje automatizado y a la noción subyacente de algoritmo. Desde su caracterización formal, un algoritmo es considerado un constructo matemático a efectos de lograr un propósito dado bajo ciertos suministros. Desde su empleo concreto, alude a qué tiene que ser implementado y ejecutado para llevar a cabo una acción y obtener ciertos efectos. En lo que sigue, cuando hablemos de "algoritmos" haremos referencia a esta segunda acepción, esto es, a un algoritmo realizado en un artefacto concreto que forma parte de nuestro mundo material.

un algoritmo que reconocía el rostro de las personas en las fotografías que se subían a su red y sugería su etiquetación. Si bien no faltaron voces que llamaron la atención sobre la falta de información acerca de cómo se utilizaba esa base de datos, su innegable atractivo lo volvió todo un suceso. Una década más tarde, en mayo de 2020, se calcula que más de 350 millones de fotos se cargan en Facebook y se etiquetan con este algoritmo que mejoró muchísimo desde su debut.

En 2017, mientras tanto, salió a la venta el modelo X del teléfono celular iPhone, el primer dispositivo masivo en utilizar TRF para desbloquear su pantalla. La tecnología, popularizada como FaceID —un juego de palabras entre identidad y rostro, un resumen perfecto que solo el marketing de la compañía fundada por Steve Jobs puede lograr— pronto fue imitada por todos los dispositivos de alta gama[12].

Como señala la historiadora y socióloga estadounidense Kelly A. Gates (2011), la TRF no era la única técnica biométrica disponible a comienzos del siglo XXI y aún hoy no es la que demostró ser más robusta en sus resultados, pero ha sido la que ha recibido mayor cantidad de apoyo e inversión sostenida por parte de gobiernos y agencias de seguridad. ¿Cuál habrá sido el interés en financiar el esfuerzo extremadamente complejo de crear algoritmos que puedan reconocer rostros y distinguirlos unos de otros cuando, por ejemplo, las huellas digitales, el iris y la voz son alternativas válidas y menos costosas?

Al parecer, dos factores explican esta decisión. El primero es que el reconocimiento de rostros es la única manera, de todas las opciones posibles dos décadas atrás, de identificar personas sin necesidad del consentimiento que se asume que está presente cuando se deja registro de las huellas, el iris o la voz. Sin que nos demos cuenta, nuestros rostros son captados hoy por miles de cámaras que nos rodean en edificios públicos y privados, en calles, aeropuertos, estaciones de tren e incluso en zonas de nuestra casa que compartimos con vecinos. No existe un consentimiento explícito de nuestra conformidad para que nos tomen imágenes cuando retiramos dinero de un cajero automático, por ejemplo, o entramos a un local a comprar una prenda. Esa increíble facilidad para obtener rostros incontables volvió a la TRF mucho más atractiva que sus competidores.

[12] Deliberadamente dejaremos de lado aquí la distinción, que suele pasar inadvertida, entre las tecnologías de reconocimiento facial (TRF) y las de autenticación facial. En rigor, la TRF es la que capta, analiza y almacena escaneos faciales para identificar la identidad de una persona, a través de la comparación de la información obtenida con la almacenada previamente en una base de datos. La autenticación facial, por su parte, se usa para franquear el acceso a un espacio (e. g., una habitación), a un dispositivo (e. g., un teléfono móvil) o a un objeto (e. g., una caja fuerte) mediante el rostro, que previamente fue almacenado, o bien en un servidor, o bien en el propio artefacto. En concreto, actualmente el FaceID de Apple (como las tecnologías semejantes de sus competidores) encripta los datos obtenidos desde el rostro del usuario en un microchip del propio teléfono. Con esto, pretende evadirse de las críticas por la vulnerabilidad de la TRF que, por caso, afectaron —de un modo por demás relativo— a Facebook (ver, por todos, Li y Jain, 2014).

El segundo factor relevante para que esta tecnología se haya impuesto es que su uso actual es compatible con grabaciones y archivos que existen desde finales del siglo XIX, como las fotografías de nuestros pasaportes, licencias de conducir y documentos de identidad, por ejemplo. Si bien es cierto que el registro de huellas digitales se profesionalizó en la década de 1980 con la adopción de normas internacionales y el abandono de la tinta tradicional por un escaneo óptico que permitió su digitalización, el archivo fotográfico de rostros de las fuerzas policiales de todo el mundo era amplio a comienzos del milenio. Para la burocracia estatal global, nuestras caras son nuestra identidad desde hace muchas décadas. De acuerdo con Gates, existe una conexión establecida desde hace más de un siglo entre las formas documentales de identificación y las imágenes faciales estandarizadas (Gates, 2011, 44).

En pocos años, y gracias a la financiación de grandes Estados, la TRF se volvió una de las formas preferidas por organismos públicos y privados para confirmar la identidad de las personas en espacios comunes, eventos públicos de gran escala e incluso en los espacios de estudio y trabajo. Mientras que el uso de documentos en papel comenzó a ser visto como ineficiente y manipulable, los algoritmos de reconocimiento de rostros crecieron en popularidad y demanda.

En el año 2014, por ejemplo, se llevó adelante el primer *International Workshop on Face and Facial Expression Recognition* (FFER) en Estocolmo, que se establecería con frecuencia bianual hasta el día de hoy. Se trata de una reunión que tiene como objetivo "convocar a investigadores que están trabajando en el desarrollo de sistemas de reconocimiento de rostros y expresiones faciales en condiciones no ideales, como las que podrían estar presentes en un video".[13] Sus principales ponencias son publicadas en un tomo por la editorial Springer al año siguiente, y se volvieron material de consulta en el área. Entre esas contribuciones, por ejemplo, de la edición 2014 se destaca "Probabilistic Elastic Part Model for Real-World Face Recognition", del científico informático chino-estadounidense Gang Hua, quien comienza su investigación con la siguiente afirmación:

La popularidad cada vez más creciente de las redes sociales y la ubicuidad de las cámaras de video proporcionan una fuente única de información que de otro modo no estaría disponible para usos militares, de seguridad y forenses. Esto implica que el reconocimiento robusto de los rostros presentados en estas imágenes y videos se ha convertido en una necesidad emergente. Por ejemplo, tanto los disturbios de Vancouver, Canadá, en 2011, como la tragedia del atentado con bombas en la Maratón de Boston, en 2013, exigen tecnologías robustas de reconocimiento facial para identificar a los sospechosos a partir de imágenes y videos de baja calidad de fuentes no restringidas (Hua, 2015, 4).

El párrafo es revelador porque condensa en pocas líneas tanto por qué la TRF tiene ventaja sobre otras tecnologías biométricas (las imágenes que inundan las

[13] Pueden consultarse detalles en https://vap.aau.dk/ffer14/

redes sociales y la multiplicidad de cámaras de vigilancia) como cuál sería el problema que viene a resolver este tipo de dispositivos: la amenaza del terrorismo y los levantamientos sociales. En la retórica de este artículo científico solo falta el otro componente que está presente en la manera en la que este tipo de tecnología es presentada en el discurso público: su supuesta neutralidad.

En la página web de Belatrix —una compañía dedicada al software fundada en Mendoza, Argentina, y que, luego de ser adquirida por la multinacional Globant, tiene presencia hoy en países como Perú, Colombia y España— se publicó en agosto de 2019 un extenso artículo titulado "Facial recognition: A two-sided story", en el que se intenta dar cuenta de las ventajas de la TRF admitiendo también sus retos. Firmado por Alejandra Rodríguez, miembro del departamento de marketing de la compañía, el texto aparece como uno de los primeros resultados que arroja una búsqueda del término "tecnología de reconocimiento facial" en Google.

Si bien el reconocimiento facial es efectivo en muchos casos, la tecnología aún está madurando. Lo que hace que el debate a su alrededor sea tan interesante es que, cuando es preciso, crea controversia porque a las personas les preocupa quién podría tener acceso a sus datos. Y en los casos en que no funciona, las personas están preocupadas por lo que significa esta falta de precisión y si puede conducir a la discriminación o la identificación errónea (Rodríguez, 2019).

En la visión que desea compartir Belatrix, los temores y las dudas que genera la TRF nacen de las personas que podrían estar detrás de sus aplicaciones. "No se trata de la tecnología, se trata de su uso. Si bien el reconocimiento facial hoy puede generar controversia, lo mismo sucedió con cada tecnología disruptiva a lo largo de la historia, siendo fuente de conmoción, asombro y debate", concluye.

La empresa de desarrollo de software Cognitec, por su parte, lleva en su logo el eslogan "The face recognition company" y en su página web asegura que su objetivo es "luchar contra el crimen y reducir los sesgos humanos con tecnologías de reconocimiento facial". Entre las razones para que las autoridades adquieran sus servicios destaca que "los investigadores humanos a menudo se ven abrumados por situaciones de emergencia tensas o la acumulación de casos sin resolver, lo que resulta en cansancio mental y estrés, que tiene un impacto negativo en las investigaciones. Los algoritmos no se cansan: siempre ofrecerán resultados consistentes para que los equipos de policía confíen". Tal como sucede con Belatrix, Cognitec no puede ignorar que existe resistencia a sus productos, pero no parece tener problemas con eso: "Los peligros del mal uso y las limitaciones técnicas de las tecnologías de reconocimiento facial son bien conocidos, están bien documentados y bien discutidos. ¡Pero también lo están sus muchos beneficios!" (Cognitec, 2020).

4. Valores y hechos: el colapso de una dicotomía

Según acabamos de ver, dada la ausencia de nuestro consentimiento explícito, la TRF obtuvo con notable facilidad innumerables imágenes de rostros y se tornó más atractiva que sus competidores tradicionales. Ahora bien, frente a los razonables reparos que nos suscita este estado de la cuestión —en virtud de la incapacidad de controlar nuestra información íntima y la percepción de ser acechados, que exacerban nuestra sensación de vulnerabilidad—, la visión que desean compartir quienes comercializan esta tecnología pretende restringir las sombras de nuestros temores y dudas solo sobre las personas que están detrás de sus aplicaciones: su punto es que no se trataría de la tecnología misma, sino de su (mal) uso. La presunción que subyace a esta afirmación es que es posible trazar una distinción tajante entre el diseño de la tecnología y el uso que hacemos de ella.

Según esta visión estratificada, el diseño estaría desprovisto de valores y, por lo tanto, de sesgos, que solo tendrían lugar en la estricta instancia del uso por parte de quienes están detrás de la TRF. De este modo, la socavación de nuestra privacidad, nuestra autonomía y cierto anonimato, así como la gravitación de prejuicios, por ejemplo, en materia de selectividad penal (e. g., Baratta, 2004), no serían cuestiones latentes en el diseño mismo, sino consecuencias del (mal) uso hecho por quienes operan la TRF.

La idea que está detrás de esta afirmación es —en suma— que el diseño está ayuno de valores y, por tanto, es objetivo y neutral, mientras que aquello que resulta subjetivo, cargado de valores y sesgado es el uso humano de la tecnología. El diseño reposa en hechos, el uso en valores, y hay una dicotomía clara entre unos y otros.

En lo que sigue, vamos a poner en cuestión la viabilidad de esta dicotomía, tanto desde una dimensión filosófica conceptual como desde una instancia empírica. Nuestro objetivo es alegar la imposible neutralidad, no solo del uso, sino del diseño mismo de la TRF.

Veamos, en primera instancia, algunas consideraciones conceptuales. El asunto de los hechos y los valores, formulado de un modo suficientemente amplio, nos concierne a todos. En esto se diferencia de manera diáfana de muchos problemas filosóficos. En efecto: el problema de los hechos y los valores parece de elección forzosa, en el sentido de que cualquier persona reflexiva ha de tener una opinión frente a él. Ocurre que una particular solución a este problema se ha arrogado el estatus de institución cultural, y es la siguiente: hechos y valores pertenecen a esferas totalmente distintas y, por lo tanto, no hay una base objetiva

para decidir si las cosas son buenas o malas, mejores o peores. En *El desplome de la dicotomía hecho-valor*, Hilary Putnam (2002) arremete contra esta dicotomía en un espíritu conceptual que reconstruimos, muy brevemente, a continuación[14].

En primer lugar, es importante tener presente la siguiente aclaración terminológica: una *dicotomía* es caracterizada como una distinción tajante y omnipresente que pretende poder aplicarse a todo enunciado significativo, en absolutamente toda área. En cambio, una *distinción* ordinaria tiene rangos de aplicación y no nos sorprende si no se aplica siempre.

La estrategia general de Putnam para difuminar la dicotomía hecho-valor y rehabilitarla, entonces, como una mera distinción puede resumirse del modo siguiente. Centra su atención en los juicios de hecho (allí donde quienes intentaron desmantelar la dicotomía en cuestión habían apuntado a los juicios de valor) y ataca la tradición empirista al calor de la cual se gestó esa noción en su nudo gordiano.

En efecto, el diagnóstico de Putnam es que el gran error se cometió respecto de cómo se conceptualizó a los juicios de hecho desde aquella tradición filosófica. Se asumió que son *objetivos,* en el sentido de que serían garantizables más allá de toda discusión ulterior y, correlativamente, se incurrió en un error que es parasitario de aquél: se consideró que los juicios de *valor* son ineludiblemente *subjetivos*.

Frente a ello, la estrategia argumentativa de Putnam puede articularse en dos instancias: una parte crítica o negativa destinada a mostrar que la dicotomía está sustentada en argumentos indefendibles y una parte constructiva o positiva a los efectos de revelar la imbricación [*entanglement*] que habría entre lo fáctico y lo valorativo.

En la parte positiva de su argumento, Putnam ancla su análisis en nuestro lenguaje ordinario, a los efectos de mostrar el carácter difuso de la distinción hecho-valor. Su idea es que si reparamos en el vocabulario de nuestro lenguaje como un *todo* —y no meramente en la parte que fue supuesta por el empirismo lógico como suficiente para la descripción de los "hechos"—, vamos a encontrar que hay un entretejido profundo entre hechos y valores, incluso en el nivel de los predicados individuales. Dicha "imbricación" se patentiza, a su juicio, al considerar términos éticos "densos" [*thick*] como "cruel".

---

[14] Cabe consignar que el de Putnam es un planteo ya clásico sobre este tópico. Muchos otros autores y autoras han criticado algunas de las dicotomías modernas centrales y sus derivas. En tal sentido, pueden consultarse -*inter alia*- las objeciones de Donna Haraway (1991) a los dualismos que hemos heredado y el abordaje de Andrew Feenberg (2002) a partir de su noción de "código técnico", que conjuga conocimiento, técnica y poder. Asimismo, puede indagarse el aporte de Bruno Latour y Steve Woolgar (1979) en lo que se refiere, en particular, a la distinción hecho/valor en las controversias tecnocientíficas (así como la exploración de sus conexiones con la distinción entre *matters of fact* vs. *matters of concern*, a partir del legado de John Dewey).

Veamos, concisamente, de qué se trata. El núcleo de su propuesta consiste en señalar el "enredo" de descripción y evaluación que estaría a la base del uso competente de tales términos. Dicho de otro modo: sostiene que lo característico tanto de descripciones "negativas" (como "cruel") como de descripciones "positivas" (como "valiente") es que para poder usarlas competentemente debemos ser capaces de identificarnos imaginariamente con un punto de vista *evaluativo*. Esta es la razón por la cual, por ejemplo, quien piensa que "valiente" significa "no temeroso de arriesgar su vida" no podría comprender la distinción socrática entre la simple temeridad y la genuina valentía.

Dado este "entretejido" en nuestro lenguaje ordinario, no es posible —por principio— escindir una instancia meramente objetiva, neutral y técnica (tal como sería la de un supuesto *diseño* "aséptico") de una instancia subjetiva y valorativamente cargada en la que quedaría confinado el *uso* que se hace de ella. En términos de una de las enseñanzas cruciales del pragmatismo norteamericano: los valores lo permean todo[15].

Si extrapolamos estas consideraciones conceptuales a la TRF, resulta que, lejos de las pretensiones de la retórica de neutralidad, en el corazón mismo de su diseño hay gravitación de valores. En otras palabras, desde su propia génesis se pone en juego el tipo de sociedad que queremos habitar.

Cabe señalar, entonces, que —por principio— los valores gravitan en todas las instancias de injerencia humana acerca del software biométrico (y no quedan restringidos al mero uso de la TRF). Esto es, se hallan presentes en (a) la delimitación del problema; (b) la elección del tipo de algoritmo; (c) la selección de los datos de entrenamiento; (d) la supervisión humana del resultado (*output*).

Veámoslo más detenidamente ahora desde su dimensión empírica, a partir de examinar -muy sumariamente- esta tecnología. En primer lugar, circunscribir la seguridad —en clave de restablecimiento del orden— como uno de los problemas cruciales a los cuales se quiere responder a partir de la videovigilancia mediante la incorporación de la TRF resulta una cuestión valorativamente cargada y muy controvertida. En efecto, a la vera de la idea de un control democrático de la criminalidad y las violencias, la delimitación misma del problema parece propiciar la falacia, ya comentada, de que una sociedad más vigilada es una sociedad más segura.

---

[15] Cabe aclarar que de aquí no se sigue: (i) que no haya dimensiones fácticas atendibles, (ii) ni que haya necesariamente una imbricación, a su vez, entre distintos tipos de valores (por ejemplo, epistémicos y ético-políticos), (iii) ni un compromiso, del tipo que abraza Putnam, con alguna variedad del objetivismo moral (cf. Putnam, 2004). A los efectos de nuestra argumentación, simplemente resulta suficiente que se conceda la tesis más débil según la cual hay una mera distinción de grado (y no de clase), esto es, un entretejido entre hechos y valores, que —para el caso que nos ocupa— torna inaplicable una demarcación tajante entre la presunta objetividad del diseño de la TRF y su uso subjetivo (concebido, este último, como la única instancia en la que tendrían injerencia nuestros valores).

Una vez que desde su origen se entrelazan conceptualmente seguridad y vigilancia, la elección del tipo de algoritmo subyacente a la TRF tampoco estará desprovista de valores. El reconocimiento facial funciona mediante un software alimentado por datos, y su procesamiento tiene lugar a través de un algoritmo que está entrenado para reconocer rostros e individualizar sus rasgos. Una vez que se realiza el mapeo de los rasgos faciales, el software genera una plantilla con la representación matemática para ese rostro único. Esa plantilla es el dato biométrico dentro de la tecnología de reconocimiento facial. Con la plantilla biométrica el rostro ya puede ser leído por una computadora y contrastado con una base de datos que previamente almacenó todo un conjunto de rostros.

Ahora bien, la selección de datos de entrenamiento tampoco resulta valorativamente neutra. El software puede llevar a cabo una comparación en tiempo real con todos los rostros almacenados en esa base de datos para determinar si una persona se encuentra registrada allí. La biometría es un proceso de probabilidades, por lo que una vez que el software encuentra una potencial coincidencia, arroja un porcentaje que define qué tan probable es que corresponda a la misma persona (e. g., ADC, 2020). Ahora bien, como en todas las aplicaciones de aprendizaje automático, un prejuicio inicial en los datos de entrenamiento genera predicciones inexorablemente sesgadas, que pueden exacerbar la discriminación de los sujetos y grupos más vulnerables.

Finalmente, la instancia humana de supervisión, dada su condición inherentemente subjetiva, también ha de estar cargada de valores. Ahora bien, nótese que es en esta dimensión —y solo en ella— donde la prédica presente en las empresas que venden TRF está dispuesta a conceder su estatus problemático: "no se trata de la tecnología, se trata de su uso". Otra falacia a advertir es, pues, aquella que constriñe el alcance de los valores al mero *uso* por parte de operadores humanos, frente a cuya impureza y fragilidad se impondría la "asepsia", la neutralidad y el carácter incansable del algoritmo.

Si nuestra línea de argumentación es apropiada, entre los riesgos asociados a esta tecnología no solo cabe consignar su uso encubierto o sin consentimiento de la población, la inversión de la carga de la prueba (somos culpables hasta que el algoritmo diga que no lo somos), la discriminación involucrada en un alto porcentaje de falsos positivos (personas no blancas, mujeres, otras comunidades vulnerables), su uso sin una base legal regulatoria específica (tal como acontece en el caso argentino), eventuales hackeos a la base de datos (por pobre o nula implementación de medidas de seguridad), su empleo ubicuo (en tiempo real o con imágenes grabadas), la individualización y seguimiento de cada persona y la concomitante facilitación de la vigilancia masiva automatizada (véase, e. g., ADC, 2020).

Una dimensión inadvertida y adicional es aquella sobre la que, centralmente, queremos llamar la atención: estos riesgos ínsitos en el *uso* están inextricablemente unidos al *diseño* mismo de la TRF. Si esta cuestión queda velada, se obtura por

razones *a priori* toda discusión acerca de la sociedad en la que queremos vivir. En otros términos: no hay meras cuestiones técnicas de diseño, relativas a su eficiencia y eficacia, que puedan deslindarse tajantemente de las cuestiones valorativas inherentes a su uso. El diseño mismo está atravesado por valores y, si ello es así, hay —por principio— una irreductibilidad de apreciaciones en torno a la "buena vida", que deseamos explicitar tal que sea posible abrir un debate que, de otra manera, ni siquiera podría plantearse.

No deseamos, por tanto, incurrir en aquello mismo que objetamos, esto es, en afirmar, sin mediación crítica reflexiva, que haya algo intrínsecamente "malo", "espurio" e "indeseable" en la TRF. Frente a los agoreros que acríticamente celebran sus bondades y ven problemas solo extrínsecos, en relación con su uso, nuestra invitación es, por el contrario, a que demos el debate en torno a cuáles han de ser los valores que le imprimiríamos y, más aún, por qué y para qué nuestra sociedad tendría —en su caso— que abrazar esta tecnología.

Cabe advertir ahora, desde su ilustración empírica, cuán implausible resulta la idea de que la información recogida, a su vez, por la TRF, esté exenta de valores. En efecto, a la luz del desmantelamiento de la dicotomía hecho-valor, la metáfora de la "sociedad de la exposición" que consignamos al principio de este trabajo revela la imposible neutralidad de la tecnología.

En efecto, la metáfora de la sociedad de la exposición subraya que nuestras subjetividades resultan moldeadas a la vera del afán de ser "nuestros propios publicistas" y, de este modo, la información que generamos y transparentamos está valorativamente cargada desde su propia génesis. Nos exhibimos exaltando ciertos rasgos que apreciamos o evaluamos como positivos y, con ello, la presunción de que el problema resida en el mero uso de la TRF y no, desde el inicio, en el diseño de la tecnología misma, se revela como una completa quimera que termina automatizando la desigualdad y los sesgos que anidan en nuestras sociedades.

5. Valores y sesgos: ¿la automatización de la desigualdad?

Tal como acabamos de ver, los algoritmos están cargados de valores. En este sentido, el ya popular eslogan informal "*garbage in, garbage out*" permite asir de una manera bastante intuitiva qué es lo que está en juego en este caso: las conclusiones solo pueden ser fiables en función de los datos sobre los cuales están basadas. En consecuencia, si la base de datos está viciada, conducirá a resultados sesgados y, en ocasiones, injustos e inequitativos. Se solapan aquí, pues, tanto cuestiones epistémicas como valorativas.

Si asumimos, entonces, que hay una carga inevitablemente normativa en las tecnologías de la información en general y en el desarrollo de algoritmos en particular (e. g., Newell y Marabelli, 2015), parece que los algoritmos conducen inexorablemente a decisiones sesgadas. El diseño y la funcionalidad de un

algoritmo reflejan los valores de la sociedad en la que están inmersos quienes los diseñan y de sus usos pretendidos. Como adelantamos ya, el desarrollo no es neutral: no hay una elección objetivamente correcta en ninguna instancia del desarrollo, sino muchas posibles elecciones (Johnson, 2006). Por lo tanto, es muy arduo detectar los sesgos latentes en los algoritmos y en los modelos que producen.

Friedman y Nissenbaum (1996) sostienen que los sesgos pueden tener lugar a partir de, al menos, tres instancias: (i) valores sociales preexistentes de las instituciones sociales, prácticas y actitudes de las que surge la tecnología, (ii) restricciones técnicas y (iii) aspectos emergentes de un contexto de uso.

La cuestión que nos interesa enfatizar aquí es, especialmente, la primera instancia concerniente a los valores sociales preexistentes que gravitan en ya en el diseño, y que se extrapolan al desarrollo y los usos de la TRF. En efecto, a partir de tomas de decisión sesgadas pueden tener lugar resultados fuertemente discriminatorios. En especial, los algoritmos que arman perfiles son usualmente citados como fuentes de discriminación. Estos algoritmos identifican correlaciones y hacen predicciones acerca de la conducta de un determinado grupo. El sesgo puede concebirse, así, como una dimensión de la toma de decisión misma, en tanto que la discriminación sería un efecto de la decisión, en términos de un impacto desproporcionadamente adverso resultante de la toma de decisión algorítmica (GIFT, 2020).

En *Automating inequality*, Virginia Eubanks (2018, 11) sostiene que en *1984*, George Orwell se equivocó en una cosa. El Gran Hermano no *te* está mirando, sino que *nos* está mirando. La mayoría de las personas están sujetas a escrutinio digital como miembros de grupos sociales, no como individuos. Gente de color, inmigrantes, grupos religiosos impopulares, minorías sexuales, los pobres y otras poblaciones oprimidas y explotadas soportan una carga mucho mayor de seguimiento que los grupos favorecidos. Los grupos marginados se enfrentan, pues, con niveles más altos de recopilación de datos cuando, por ejemplo, acceden a beneficios públicos, caminan por vecindarios altamente vigilados, ingresan al sistema de atención médica, o cruzan las fronteras nacionales. Los datos actúan, así, para reforzar su marginalidad cuando se utilizan para identificarlos por mera sospecha y con un escrutinio adicional. De este modo, estos grupos considerados indignos son señalados por políticas públicas punitivas y sujetos a una vigilancia más intensa. A la vera de valores sociales preexistentes (sumamente cuestionables), se fragua un ciclo de retroalimentación de la injusticia.

Veámoslo a través de un episodio real. En mayo de 2020, un mes después de las protestas masivas en contra de la policía en Estados Unidos tras la muerte de George Floyd, una persona fue arrestada bajo el cargo de agresión a un oficial. La cadena NBC 6 realizó una investigación y determinó que la identificación de la persona se produjo mediante imágenes obtenidas con la cámara que el policía llevaba en su uniforme y a través del uso del software de reconocimiento facial

de Clearview AI. La investigación concluyó que los departamentos de policía en todo el sur de la Florida están usando la misma tecnología que identifica a las personas a través de fotos disponibles públicamente, incluyendo sus redes sociales. El abogado de la detenida —una mujer de origen latino— advirtió que, de no haber sido por la investigación periodística, su cliente no habría estado al tanto de la situación, porque la policía omitió mencionar en el informe del arresto la tecnología empleada (Fossi y Prazan, 2020).

El caso de TRF que traemos a colación parece ilustrar, de manera diáfana, los diversos modos en que convergen la metáfora de la sociedad de la exposición a la que aludimos (la policía de valió de fotos disponibles públicamente en redes sociales) con el carácter irreductible de valores sociales preexistentes que permean el diseño mismo de la tecnología, dando lugar a sesgos que automatizan la desigualdad.

En este entramado se revela, pues, una suerte de corresponsabilidad (que, no obstante, no parece ser simétrica) entre las personas usuarias, por ejemplo, de redes sociales y quienes diseñan, desarrollan y tienen en sus manos el uso de tecnología tal como la TRF.

La incapacidad de controlar nuestra información íntima en conjunción con la expansión ingente de diversas formas de supervisión digital invita —por lo menos— a las dos cuestiones siguientes. A poner en jaque la falacia según la cual una sociedad más vigilada es una sociedad más segura y a ubicar en el centro de la escena la necesidad acuciante de dar un debate público en torno a los valores que deseamos que nos configuren como sociedad y que trasunten en nuestra tecnología.

Referencias bibliográficas

ADC (2020). Con mi cara no. Asociación por los Derechos Civiles. http://conmicarano.adc.org.ar/

Baratta, Alessandro (2003). Principios de derecho penal mínimo. En Alessandro Baratta, *Criminología y Sistema Penal* (*compilación* in memoriam) (pp. 299-333). Buenos Aires: B de F.

Bentham, Jeremy (1791). *Panopticon, or the Inspection House*. London: T. Payne.

Binder, Alberto (2009). El control de la criminalidad en una sociedad democrática. En Gabriel Kessler (comp.), *Seguridad y Ciudadanía. Nuevos paradigmas, reforma policial y políticas innovadoras* (pp. 25-54). Buenos Aires: Edhasa.

Caplow, Theodore, Bahr, Howard M., Modell, John y Chadwick, Bruce A. (1994). *Recent Social Trends in the United States 1960-1990*. Montreal: McGill-Queen's University Press.

Cognitec (2020). Cognitec supports fighting crime and curtailing human bias with face recognition technologies. https://www.cognitec.com/news-reader/fighting-crime-and-curtailing-human-bias-with-face-recognition.html

Debord, Guy (1967). *La société du spectacle*. Paris: Buchet.

Deleuze, Gilles (1992). Postscript on the Societies of Control. *October*, *59*, 3-7.

Eubanks, Virginia (2018). *Automating inequality: How High-Tech Tools Profile, Police and Punish the Poor*. New York: St. Martin's Press. https://doi.org/10.1080/10999922.2018.1511671

Feenberg, Andrew (2002). *Transforming technology. A critical theory revisited*. USA: Oxford University Press.

Feinstein, Dianne (2001). Biometrics Identifiers and the Modern Face of Terror: New Technologies in the Global War on Terrorism — Open Statement, 14 de noviembre de 2001. Washington, DC: U. S. Government Printing Office. https://www.govinfo.gov/content/pkg/CHRG-107shrg81678/html/CHRG-107shrg81678.htm

Friedman, Batya y Nissenbaum, Helen (1996). Bias in computer systems. *ACM Transactions on Information Systems (TOIS)*, *14*(3), 330-347. https://doi.org/10.1145/230538.230561

Fossi, Connie y Prazan Phil (2020). Miami Police Used Facial Recognition Technology in Protester's Arrest. https://www.nbcmiami.com/investigations/miami-police-used-facial-recognition-technology-in-protesters-arrest/2278848/

Foucault, Michel (2013). *La société punitive: Cours au Collège de France*, *1972-1973*, ed. Bernard E. Harcourt. Paris: Gallimard - Le Seuil.

Gates, Kelly A. (2011). *Our Biometric Future: Facial Recognition Technology and the Culture of Surveillance*. New York: NYU Press. https://doi.org/10.18574/9780814733035

GIFT (2020). *Caja de herramientas humanísticas*. https://grupo.gift

Haggerty, Kevin D. y Ericson, Richard V. (2000). The Surveillant Assemblage. *British Journal of Sociology*, *51*(4), 605-622. https://doi.org/10.1080/00071310020015280

Han, Byung-Chul (2015). *The Transparent Society*. Stanford: Stanford University Press.

Haraway, Donna (1991). A Cyborg Manifesto: Science, Technology, and Socialist-Feminism in the Late Twentieth Century. En *Simians, Cyborgs and Women: The Reinvention of Nature* (pp. 149-181). New York: Routledge.

Harcourt, Bernard E. (2007). *Against Prediction. Profiling, Policing, and Punishing in an Actuarial Age*. Chicago: University of Chicago Press. https://doi.org/10.7208/chicago/9780226315997.001.0001

Harcourt, Bernard E. (2011): *The Illusion of Free Markets. Punishment and the Myth of the Natural Order*. Cambridge: Harvard University Press. https://10.2307/j.ctvjhzpv2

Harcourt, Bernard E. (2015): *Exposed. Desire and Disobedience in the Digital Age*. Cambridge: Harvard University Press. https://doi.org/10.4159/9780674915077

Hua, Gang (2015). *Probabilistic Elastic Part Model for Real-World Face Recognition*. En Ji, Qiang et ál (eds.), *Face and Facial Expression Recognition from Real World Videos. FFER 2014*. *Lecture Notes in Computer Science*, vol. 8912, 3-10. Cham: Springer. https://doi.org/10.1007/978-3-319-13737-7_1

Israel, Steve A. e Irvine, John (2012). Heartbeat biometrics: a sensing system perspective. *Int. J. Cognitive Biometrics*, *1*(1), 39-65. https://doi.org/10.1504/IJCB.2012.046514

Johnson, Jeffrey A. (2006). Technology and pragmatism: From value neutrality to value criticality. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.2154654

Kafka, Franz (1925). *Der Process*. Berlin: Verlag Die Schmiede.

Latour, Bruno y Woolgar, Steve (1979). *Laboratory life. The social construction of scientific facts*. Beverly Hills, California, Sage Publications.

Li, Stan Z. y Jain, Anil K. (eds.) (2004). Introduction. En *Handbook of Face Recognition* (pp. 1-18). London: Springer Verlag. https://doi.org/10.1007/978-0-85729-932-1_1

Medina, R., III, *et al.* (2019). Authentication based on heartbeat detection and facial recognition in video data. United States Patent 10,268,916 B1. https://patentimages.storage.googleapis.com/a8/12/73/ca411b4b82e4e4/US10268910.pdf

Nait-ali, Amine (ed.) (2020). *Hidden Biometrics*. *When Biometrics Security Meets Biomedical Engineering*, Singapur: Springer. https://doi.org/10.1007/978-981-13-0956-4

Newell Sue y Marabelli, Marco (2015). Strategic opportunities (and challenges) of algorithmic decision-making: A call for action on the long-term societal effects of 'datification'. *The Journal of Strategic Information Systems*, *24*, 3-14. https://doi.org/10.1016/j.jsis.2015.02.001

Online Etymology Dictionary (2020). Entrada "mug". https://www.etymonline.com/word/mug

O'Malley, Pat (1992). Risk, Power and Crime Prevention. *Economy and Society*, *21*(3), 252-275. https://doi.org/10.1080/03085149200000013

Orwell, George (1949). *1984*. Londres: Secker.

Petersen, Julie K (2001). *Understanding Survilleance Technologies: Spy Devices, their origins & applications*. Boca Raton: CRC Press.

PNUD (1995). *Informe sobre Desarrollo Humano 1994*. Washington: Programa de las Naciones Unidas para el Desarrollo.

Putnam, Hilary (2002). *The Collapse of the Fact/Value Dichotomy and Other Essays.* New York: Harvard University Press. https://doi.org/10.1201/978142003881

Ramli, Dzati A., Hooi, Man Y. y Chee, Kai J. (2016). Development of Heartbeat Detection Kit for Biometric Authentication System. *Procedia Computer Science*, 96, 305-314. https://doi.org/10.1016/j.procs.2016.08.143

Rhodes, Henry T. F. (1956). *Alphonse Bertillon: Father of Scientific Detection*. New York: Abelard-Schuman.

Rodríguez, Alejandra (2019). Facial recognition: A two-sided story. https://www.belatrixsf.com/blog/facial-recognition

Vaidhyanathan, Siva (2011). *The Googlization of Everything —and Why We Should Worry*. Berkeley: University of California Press. https://doi.org/10.1525/9780520948693

Vucetich, Juan (1904). *Dactiloscopía comparada, el nuevo sistema argentino: trabajo hecho expresamente para el 2do. Congreso Médico Latino-americano, Buenos Aires, 3-10 de abril de 1904*. Buenos Aires: Peuser.