CIBERRASTREO ANALÍTICO: *WEB SCRAPING* Y *BIG DATA* COMO TÉCNICAS DE INVESTIGACIÓN EN EL DERECHO PROCESAL PENAL CHILENO

ANALITICAL CYBER HARVESTING: WEB SCRAPING AND BIG DATA AS CRIMINAL INVESTIGATIVE TECHNIQUES IN CHILEAN CRIMINAL PROCEDURAL LAW

Roberto Navarro-Dolmestch[*]

Resumen: Este artículo analiza la aplicación de las técnicas de *web scraping* a la investigación de delitos y a su juzgamiento. Se sostiene que esta es una técnica intrusiva porque su uso es una injerencia sobre el derecho a la intimidad y tensiona el debido proceso. Considerando este carácter, y que carece de regulación legal expresa en el derecho procesal penal chileno, la admisibilidad de su uso se encuentra sometido a un control jurisdiccional sobre la calidad de la información recogida y tratada, la explicabilidad de sus resultados y el cumplimiento de otros deberes jurídicos.

Palabras clave: *web scraping*, *web crawling*, derecho procesal penal chileno, inteligencia artificial, investigación criminal.

Abstract: An analysis of the application of web scraping techniques to crime investigation and adjudication is presented in this paper. It asserts that web

[*] Doctor en Derecho. Profesor adjunto de Derecho Penal, Universidad Católica del Maule. Correo electrónico: ronavaroro@ucm.cl https://orcid.org/0000-0003-0907-5714

Este artículo se ha elaborado en el marco del proyecto de investigación "La responsabilidad de la inteligencia artificial: un desafío para las ciencias penales" (PID2020-112637RB-I00), financiado por el Programa Estatal de Fomento de la Investigación Científica y Técnica de Excelencia, Subprograma Estatal de Generación de Conocimiento, del Ministerio de Economía y Competitividad, España.

scraping is an intrusive technique because it interferes with privacy rights and undermines due process of law. Considering its intrusive nature and the absence of explicit legal regulation in Chilean criminal procedural law, its admissibility has been argued to be subject to judicial control regarding the quality and explainability of the collected and processed information, as well as compliance with other legal obligations.

Keywords: Web scraping, web crawling, Chilean criminal procedural law, artificial intelligence, criminal investigation.

Introducción

Este artículo parte de la premisa, cada vez más aceptada[1], que la

> "inteligencia artificial es una tecnología disruptiva que está [...] haciendo surgir problemas fundamentales y desafiantes a considerar por el derecho"[2].

El derecho procesal y los sistemas de justicia no escapan ni a esta nueva realidad tecnológica[3] ni a los desafíos que la inteligencia artificial (IA) les plantea.

En ese sentido, y como lo ha sintetizado John Danaher[4], el desarrollo de la IA ha hecho surgir, al menos, tres debates:

a) el de los sus efectos en la intimidad y el vigilantismo,
b) el de los sesgos y la discriminación y
c) el de la transparencia y el procedimiento.

En la aplicación de la IA a la investigación penal y al proceso penal, esos tres debates se muestran especialmente incidentes. El primero de esos debates conduce a la necesidad de evaluar de qué forma las regulaciones y concepciones sobre la intimidad protegen de modo efectivo al imputado en un contexto en el que el Estado-investigador dispone, ahora, de herramientas de búsqueda de información basadas en IA más sofisticadas y poderosas, y de ingentes volúmenes de datos en los sistemas informáticos y en la web. El segundo de los debates, sobre la forma en la que los sesgos pueden llevar al desarrollo y aplicación de herramientas inteligentes que produzcan efectos arbitrariamente discriminatorios, concediendo tratamientos perjudiciales o beneficiosos, cons-

[1] A modo de ejemplo: Yeung (2019) p. 28; Schirmer (2020) p. 128; Mcginnis & Pearce (2014).

[2] Barfield (2018) p. 22.

[3] Luis (2023) p. 2.

[4] Danaher (2022) pp. 250-251.

titucionalmente inadmisibles, a determinadas categorías de imputados. El tercero, por su parte, al problema del impacto de la inexplicabilidad sobre la construcción de las conclusiones a que puede arribar una herramienta de IA sobre el ejercicio efectivo del derecho a la defensa en lo que se refiere a la contradicción del material incriminatorio generado por ese tipo de herramientas y, en consecuencia, su aptitud para contribuir a formar la convicción en el tribunal.

La relación de la IA con el sistema de persecución penal es también compleja. La complejidad consiste en que, como lo ha explicado Valentina Faggiani, la IA tiene una naturaleza contradictoria: por un lado, puede ofrecer nuevas herramientas más eficaces para la investigación y el juzgamiento de delitos; por otro:

> "puede afectar a la comunidad en su conjunto, vulnerando no solo los derechos fundamentales y las garantías procesales de un número indeterminado e indeterminable de personas, que pueden verse mermados por la opacidad ('efecto caja negra'), la complejidad, la imprevisibilidad"[5].

Este artículo se construye sobre la base de esos debates y el problema que aborda puede ser expresado como sigue. El mundo tecnológico en el que nos desenvolvemos ha permitido la generación de abrumadores volúmenes de información digital apta para ser tratada o procesada de manera automatizada a través de técnicas y procedimientos informáticos. Adicionalmente, la IA ha permitido desarrollar herramientas informáticas cada vez más poderosas para el procesamiento automatizado de esa información, con el potencial de transformar la actividad policial tradicional, tanto en la investigación de delitos como en la inteligencia policial[6]. Esto produce una doble consecuencia: hay más información disponible a la que se puede acceder que nunca antes en nuestra historia, por un lado, y, por otro, contamos con herramientas y técnicas que permiten hacer cada vez más eficiente el procesamiento de esa información de forma automatizada, es decir, sin intervención humana directa. El procesamiento de esa información inicial permite generar nueva información que surge, por ejemplo, de los patrones que las máquinas inteligentes pueden establecer con un rendimiento muy superior al que tendría un operador humano haciendo esa tarea (la distinción entre dato e información que se analizará en el §III). Tanto los datos como la información pueden ser especialmente relevantes en la investigación de delitos y en su juzgamiento, ya que pueden orientar las indagaciones de hechos que revisten los caracteres de delito, contribuyendo a la cons-

[5] Faggiani (2022) p. 520.

[6] Rowe & Muir (2021) p. 256. Por ejemplo, puede consultarse el estudio de Nussbaum & Udoh (2020) sobre la vigilancia digital en la investigación de ciberdelitos y delitos cometidos en contextos informáticos y digitales.

trucción de una teoría del caso; pueden justificar el ejercicio de la acción penal actuando como elementos de respaldo para deducir acusación y pueden, por último, fundar una sentencia condenatoria porque esas informaciones pueden actuar como elementos incriminatorios capaces de generar convicción en el tribunal sobre la existencia del delito y la participación del acusado.

Ese estado de cosas es de reciente surgimiento, y las tecnologías basadas en IA lentamente están comenzado a ser usadas con esas finalidades. El debido proceso (o, en términos de la Constitución chilena en vigor, un procedimiento y una investigación racionales y justos) y las normas que lo desarrollan, actúan como un conjunto de criterios de legitimidad de la actividad del Estado en sus facetas de investigador y juzgador. En síntesis, la legitimidad se fundamenta en la expectativa de "proteger a los individuos, miembros de una comunidad determinada, contra la utilización arbitraria del poder penal del Estado"[7] o, en otras palabras:

> "un conjunto de parámetros o estándares básicos que deben ser cumplidos por todo proceso para asegurar que la discusión y la determinación de derechos que están en cuestión se haya realizado en un entorno de razonabilidad y justicia para las personas que intervienen en su desarrollo"[8].

Tanto los derechos fundamentales generales (como la intimidad y la igualdad) como las garantías procesales (como el debido proceso o el derecho de defensa) cobran relevancia para este estudio porque su objetivo es determinar en qué medida las regulaciones procesales que pretenden garantizarlos son efectivas, considerando las tecnologías informáticas. En concreto, la pregunta es si a la aplicación de herramientas de búsqueda y procesamiento de información digital en la investigación y al proceso penal debería aplicársele el estatuto previsto para las medidas intrusivas y, en caso afirmativo, si ese estatuto es o no adecuado para la protección de derechos.

El problema surge porque en los escenarios muy conectados en línea en los que nos desenvolvemos, se genera mucha información digital o informática; una parte de ella es la que producimos con cada una de nuestras acciones en el ciberespacio y en el mundo físico. Cada vez que hacemos una compra electrónica, enviamos un correo electrónico, nos conectamos a una red social o, simplemente, llevamos nuestro teléfono inteligente, estamos dejando una traza de información que se conoce como huella digital o digital *footprint*[9]; el "anonimato" digital requiere que la persona interesada utilice procedimientos espe-

[7] Maier (2016) pp. 442-443.
[8] Duce, Marín y Riego (2008) p. 17.
[9] Rowe & Muir (2021) p. 254.

ciales para evitar dejar esa huella[10]. Las agencias gubernamentales disponen de datos e información de personas y actividades producidos por el ejercicio de sus propias funciones, mientras que el sector privado expande el ámbito de la huella digital en ámbitos como las transacciones financieras y la comunicación electrónica[11].

Este artículo comienza:

- §I con una descripción sobre lo que denominaremos el ciberrastreo analítico, esto es, una herramienta de búsqueda y procesamiento de información digital;
- §II un análisis de ella como técnica de investigación utilizable por el sistema de persecución penal;
- §III se adopta una posición sobre el carácter intrusivo del ciberrastreo, fundada en dos elementos:
  - §III.1 la relación de conflicto que ella posee con la vigencia de derechos fundamentales
  - §III.2 su naturaleza disruptiva en el concierto de la actividad investigativa y jurisdiccional;
- §IV destina a analizar la ausencia de una autorización legal expresa del ciberrastreo y, por consiguiente, la inexistencia de regulación legal de sus aspectos operativos;
- §V analiza la admisibilidad del ciberrastreo en el sistema de persecución penal chilena.

El artículo finaliza con la propuesta, a modo de conclusión, de una fórmula que toma partido por la admisibilidad del ciberrastreo como técnica de investigación, pero sometido a un test jurisdiccional.

I. *Web scraping*, *big data* e IA

El desarrollo de las tecnologías de la información y la comunicación, su relativo bajo costo de adquisición y acceso, y la gran penetración de internet han permitido la generación de ingentes volúmenes de información digital almacenada en el ciberespacio o datosfera. Se estima que en 2025 se habrán acumulado 163 zettabytes de información digital[12]. Tomando como referencia ordenadores personales estándar en el mercado, dotados de discos duros con una capacidad de un terabyte, se necesitarían alrededor de ciento sesenta y tres mil millones de ordenadores para almacenar toda la información digital que se espera esté disponible en la datosfera en 2025 (163 settabytes = 1,63 x 10^{11} terabytes).

[10] Guinchard (2021) p. 48.
[11] Sachoulidou (2023).
[12] Reinsel, Gantz & Rydning (2017) p. 3.

La información digital ofrece enormes potencialidades en variadas áreas: en actividades comerciales o crediticias[13], en mercadotecnia[14], en la política, o en la investigación científica[15] o criminológica[16], por nombrar algunas[17]. Y también tanto para la prevención delictiva (dando origen a la denominada *predic tive policing*[18]) o la investigación de delitos.

La variedad y el alto volumen de información que se ha generado condujo al desarrollo de técnicas y tecnologías de búsqueda, ordenación y procesamiento de esa información. Es en ese contexto en el que el *web scraping* adquiere relevancia. Es una

> "técnica para extraer información desde la World Wide Web (www) y guardarla en un sistema de archivos o base de datos para su posterior consulta o análisis"[19].

permite la transformación de la información desestructurada –tal como se presenta en la red– en información estructurada[20], y que tiene múltiples usos. Por ejemplo, hace posible el funcionamiento de los motores de búsqueda, como Google o MS Bing, sin los cuales no sería posible encontrar información en la web[21]. De hecho,se estima que entre un 25 %[22] al 40 %[23] del tráfico total en internet corresponde a

[13] Como lo destacan, por ejemplo: Cong, Li & Zhang (2021) pp. 225-226; Regmi, Rai & Khanal (2021) pp. 77-78; Matousek & Xiang (2021) pp. 89-90.

[14] Fleisher (2008).

[15] Luscombe, Duncan & Walby (2022); Ackerman & Pinson (2016).

[16] Brewer, Westlake, Hart & Arauza (2021) pp. 437-438; Gundur, Berry & Taodang (2021) p. 150; Leukfeldt & Kleemans (2021) p. 128; Miró (2019) pp. 99-100; Hayward & Maas (2020). A modo de ejemplo, consúltese el estudio realizado, usando triangulación de información proveniente de redes sociales, foros y otros recursos en línea, sobre delitos informáticos y sus actores: Holt, Smirnova, Strumsky and Kilger. (2014).

[17] A modo de ejemplo, consúltese el estudio realizado usando triangulación de información proveniente de redes sociales, foros y otros recursos en línea sobre delitos informáticos y sus actores: Holt, Smirnova, Strumsky and Kilger. (2014).

[18] Al respecto, véase: Asaro (2019); Erdoğan (2022); Slobogin (2022); Dülger (2024) p. 106 ss; Utset (2021).

[19] Brewer, Westlake, Hart & Arauza (2021); Gold & Latonero (2018); Gorro, Sabellano, Maderazo, Ceniza & Gorro (2017); Khder (2021); Zhao (2022) p. 951. Se suele encontrar también la denominación de *web crawling*, aunque su distinción del *web scraping* no es determinante. La diferencia consistiría en que el primero es la misma acción de rastreo, pero ejecutada en forma masiva en muchos sitios web, mientras que el último, se concentraría en la búsqueda dentro de un sitio especifico.

[20] Sirisuriya (2015).

[21] Rahman & Tomar (2020) p. 1; Thelwall & Stuart (2006) p. 1771.

[22] Sellars (2018) p. 375.

[23] Rahman & Tomar (2020).

bots ejecutando acciones de *scraping* o *crawling* en la red[24] con fines lícitos o ilícitos[25].

En el ciberespacio puede hallarse información conducente o relevante para la averiguación de la existencia de un delito o para la imputación personal de un hecho que reviste tales caracteres. La búsqueda de esa información –el *web scraping*– y su procesamiento para hallar en ella patrones o relaciones –el *big data*–, potenciadas por tecnologías de inteligencia artificial, pueden configurarse como una poderosa técnica de investigación de hechos que revisten los caracteres de delito[26], parte de la "radical metamorfosis que ha sufrido la investigación criminal debido al continuo progreso de la tecnología"[27].

Se propone denominar a esta técnica "ciberrastreo analítico" y, en su forma abreviada, "ciberrastreo". Esta expresión no solo tiene la ventaja de estar en castellano, sino que, también, recoger los dos elementos que son relevantes en su conceptualización: "ciberrastreo" porque se refiere a la búsqueda ("rastreo") de información digital almacenada en el ciberespacio (de ahí, "ciber") y "analítico" que se refiere al procesamiento de los datos.

Los datos que puedan recogerse mediante el ciberrastreo solo tienen, en principio, un valor investigativo de carácter potencial, por la dispersión y desintegración iniciales de ellos. El rastreo de datos en la red o en bases de datos no presta, necesariamente, utilidad porque ellos no son significativos *per se*: solo representan información potencial[28]. Esos datos deben ser analizados: según

[24] Como sinónimos, en Khder (2021).

[25] Un *bot* "es una aplicación, software o proceso que ha sido creado expresamente con el propósito de automatizar tareas repetitivas": Lukings & Lashkari (2022) p. 74). Gráficamente, a juicio de Dunham & Melnick (2009) p. 1: los *bots* "son abejas obreras altamente adaptables que cumplen las órdenes de su amo en una amplia 'red' ". Aunque inicialmente estuvieron destinados a la ejecución de tareas lícitas –o, al menos, inocuas en términos de dañosidad–, han ido adquiriendo relevancia como una de las principales formas que pueden adoptar las amenazas informáticas, sobre todo cuando se usan en red, llamados *botnets*: Alexandrou (2022) p. 66; Bandler & Merzon (2020) p. 11; Clough (2015) p. 41; Lukings & Lashkari (2022) p. 74.

[26] Como la aplicación Voyager, diseñada y comercializada por la empresa Web IQ HQ (https://web-iq.com) y que es utilizada por SafeLine, canal digital de denuncias de usuarios que funciona en Grecia (www.safeline.gr/en/ make-a-report/), cuyo objetivo es la detección de material abusivo y pornográfico infantil en la red internet: Kokolaki, Daskalaki, Psaroudaki, Christodoulaki & Fragopoulou (2020). CSAM (Child Sexual Abuse Materials) es la denominación genérica que se ha difundido para referirse a "cualquier representación visual que muestre la participación de un menor en conductas sexuales explícitas, incluidos, sin limitaciones, fotos, videos e imágenes generadas por computadora" (https://support.google.com/transparencyreport/answer/10330933?hl=es-419#zippy=%2Cqué-es-csam).

Otros ejemplos de programas y aplicaciones que se utilizan en el ciberrastreo pueden encontrarse en Gundur, Berry & Taodang (2021) pp. 151-152.

[27] Centorame (2021) p. 124.

[28] Albers (2014) p. 222.

la analogía propuesta por Aurelien Portuese, los datos recogidos son petróleo crudo; después de su análisis, llegan a ser combustible[29]; y de ahí lo 'analítico' el concepto propuesto. Es el tratamiento de los datos recopilados lo que genera indicios e informaciones útiles y conducentes para la investigación de un hecho que reviste los caracteres de delito. Esa es la razón por la que el ciberrastreo necesita, adicionalmente, el recurso a técnicas de procesamiento que son las que brinda el *big data*[30], un conjunto de tecnologías de entorno digital basadas en algoritmos por lo general de aprendizaje automático (*machine learning*), que hacen posible analizar grandes volúmenes de datos[31]

> "destinado a descubrir patrones (regularidades estadísticas entre variables) en conjuntos masivos de datos, reconciliar la variedad en diversas fuentes de datos y administrar datos generados a alta velocidad"[32].

El *big data* y las capacidades de procesamiento que ofrece la inteligencia artificial, ha generado una intensa discusión dogmática[33].

En el ámbito penal, el ciberrastreo puede ser usado en términos preventivos (por ejemplo, PredPol[34]) anticipando la probabilidad de ocurrencia de hechos delictivos[35] (en el modelo alemán, con la *Schleppnetzfahndung*[36]), sobre todo en materia de terrorismo o tráfico de drogas (el llamado *Smart law enforcement*[37]); como una técnica de investigación de delitos ya cometidos[38] o en el proceso de adopción de decisiones judiciales[39].

El ciberrastreo puede ser una técnica especialmente útil en delitos que, por su propia naturaleza o su dinámica delictual, generan abundante información, como ocurre de manera paradigmática con los fenómenos que tradicionalmente se denominan "crimen organizado", pero que, cuando se ejecutan a través de medios informáticos, adoptan formas descentralizadas y distribuidas, lo que ha llevado a la doctrina a hablar de criminalidad "desorganizada"[40, 41].

---

[29] Portuese (2022) p. 2.

[30] Završnik (2018b); Kokolaki, Daskalaki, Psaroudaki, Christodoulaki & Fragopoulou (2020).

[31] Hylton (2019) p. 273.

[32] Maass, Parsons, Purao, Rosales, Storey & Woo (2022) p. 75.

[33] Boyd & Crawford (2012); Brewer, Westlake, Hart & Arauza (2021); Carrero (2020); Gold & Latonero (2018); Khder (2021); Krotov, Johnson & Silva (2020); Macapinlac (2019); Parks (2022).

[34] Završnik (2018a).

[35] Andrejevic (2018); Wilson (2018).

[36] Frisch (2014).

[37] Rademacher (2020).

[38] Cano (2003); Frisch (2014) p. 29.

[39] Završnik (2019).

[40] Wall (2014), (2015).

[41] Por ejemplo, la IA puede ser en especial útil en el reconocimiento de intentos de ataques informáticos. Como lo explica Ishaq Azhar Mohammed, las huellas digitales que dejan los ata-

La inteligencia artificial no es de la esencia del *web scraping*, ya que esta técnica puede ser ejecutada, también, manualmente o con herramientas informáticas basadas en programación tradicional. En este último caso, los *scripts* o las API usadas para "rastrear" la web ejecutan las funciones de acuerdo con los parámetros predefinidos en la programación de la herramienta. El uso de la inteligencia artificial –que puede ser una abundante fuente de nuevas técnicas de investigación delictual[42]– en algunos o en todos los pasos del ciberrastreo produce como efecto la potenciación de su alcance y de sus resultados. El uso de aprendizaje automático de máquinas en arquitecturas coneccionistas (*machine learning*, *deep learning*), de algoritmia genética o, en general de tecnologías de inteligencia artificial, permite:

a) construir herramientas informáticas más robustas y rápidas, limitadas solo por los anchos de banda de conexión a la red de que se trate y las capacidades de procedimiento informático de que se disponga, aumentando, en consecuencia, la capacidad de rastreo[43] y
b) diseñar aplicaciones que, sin necesidad de una programación previa específica, sean capaces de aprender a seleccionar la información a rastrear, de superar las defensas que los administradores de sitios web habitualmente implementan para impedir el *scraping*[44] y con la posibilidad de adaptarse a los entornos en los que operan.

II. Ciberrastreo como técnica de investigación

El sistema de persecución penal chileno considera una primera fase de carácter no-jurisdiccional distinta del proceso penal propiamente tal[45], en la que los fiscales del Ministerio Público, por sí mismos o a través de las policías, llevan a cabo "diligencias de investigación que consideraren conducentes al esclarecimiento de los hechos" (art. 180 del *Código Procesal Penal* (en adelante *CPP*)).

cantes pueden ser recopiladas y ordenadas en base de datos. Estos datos pueden permitir el entrenamiento de aplicaciones de IA para reconocer y detectar intrusiones en tiempo real. Asimismo, el *scraping* asistido por IA es capaz de escanear toda la red en busca de vulnerabilidades, evitando así la mayoría de los tipos típicos de ataques: Mohammed (2020) p. 174.

[42] Middleton (2021) p. 213.

[43] Diouf, Sarr, Sall, Birregah, Bousso, & Mbaye (2019); Farley & Pierotte (2017); Luscombe, Dick & Walby (2022).

[44] Una descripción de tales medidas de defensa en Thelwall & Stuart (2006).

[45] En este sentido, Navarro-Dolmestch (2018) pp. 63-64. En contra, Castro (2023) p. 101, para quien la extensión por la Constitución chilena de las garantías a la investigación junto con el proceso "da la impresión que la investigación no forma parte del procedimiento penal".

El conjunto de tales diligencias o actuaciones conforma la "investigación penal". De acuerdo con el diseño legal, la investigación penal tiene, entre otras, la función de "consignar y asegurar todo cuanto condujere a la comprobación del hecho y a la identificación de los partícipes en el mismo" (art. 181 del C*PP*).

De esta forma, se puede comprender la investigación penal como el proceso de búsqueda y recogida de información relevante sobre la existencia o inexistencia de un delito y sobre la identidad de la persona a la que puede imputársele ese delito[46]. La información recogida debe tener respaldo en sus respectivas fuentes (*v.gr.*: testigos, informes policiales, registros audiovisuales, etc.), ya que son ellas las que se desempañarán como medios de prueba, aportando información en la audiencia de juicio oral si esta llega a producirse[47]. Si la investigación no es capaz de reunir suficiente información que respalde la imputación, la información reunida no está respaldada de manera adecuada en fuentes que puedan actuar como medios de prueba (por ejemplo, los testigos no están disponibles) o las fuentes de respaldo no son aptas para generar convicción en un tribunal (como cuando los testigos no son fiables porque tienen estímulos para mentir), el Ministerio Público tiene la facultad de decidir no ejercer la acción penal y, en consecuencia, el caso no llegará a juicio.

Para comprender adecuadamente la lógica del ciberrastreo como técnica investigativa, es necesario distinguir entre "dato" e "información", como ya se adelantó. Esta distinción también será relevante para comprender la posición que se adoptará sobre el carácter intrusivo del ciberrastreo.

Un "dato" es un elemento de hecho del que se tiene noticia, una unidad básica de conocimiento, como cuando el ciberrastreo permite saber que una persona estuvo un día determinado en un lugar específico. En la lógica del ciberrastreo, los datos se recogen del ciberespacio o de bases de datos digitales. En el ejemplo, se supo la ubicación de esa persona rastreando los lugares que visitó y de los que dejó registro, por ejemplo, en sus redes sociales. La "información", por su parte, es más que la suma agregada de varios datos, es un conocimiento nuevo que se puede adquirir sobre la base del análisis de uno más datos. Siuiendo con el ejemplo, supóngase que el ciberrastreo permitió determinar que una determinada persona ha estado tres veces en los últimos nueve meses en un determinado centro médico según los registros de sus redes sociales; que según la web de ese centro médico, este se especializa en el diagnóstico y tratamiento de enfermedades oncológicas; que la persona investigada es un activista por la defensa de los derechos de los enfermos de cáncer, ya que así lo muestran sus publicaciones en redes sociales; y que esa persona recibió su vacuna contra la COVID-19 al comienzo del proceso de inmunización, en el periodo reservado

[46] Núñez y Correa (2017) p. 199.

[47] Duce y Riego (2007) pp. 120-121.

para personas inmunodeprimidas, lo que se sabe, nuevamente, porque la persona lo divulgó en una red social. Cada una de esas unidades representan datos. A partir de ellos, se puede concluir que esa persona padece cáncer. Tal conclusión sobre el diagnóstico de la persona será, en el concepto sostenido aquí, una información. Como lo expone Marion Albers, explicando la diferencia entre dato e información:

> "las unidades de información pueden basarse en datos (o en observaciones o comunicaciones), pero los datos solo adquieren significado si son explicados e interpretados por el destinatario o usuario de los datos que los utiliza para obtener información"[48].

Los datos, considerados como unidades básicas, no necesariamente son útiles para la investigación penal. La interrelación de varios datos entre sí y la capacidad de ellos para generar información es lo que en realidad presta utilidad a la investigación. El ciberrastreo analítico, como se dijo, cubre todas las fases: desde la recogida de los datos, el proceso de estructuración de esos datos y su análisis, esto es, el proceso que permite inferir de ellos información.

El paso desde los datos a la información puede ser explicado con ayuda de la argumentación. Unos determinados datos son las premisas iniciales; la información obtenida, la conclusión y el proceso para generar esa nueva información, una inferencia[49]. Si la veracidad de la conclusión (esto es, de la información producida) no está completamente garantizada, aun cuando las premisas sean todas verdaderas, debe dejarse establecido que la relación de inferencia tiene un carácter inductivo, esto es:

> "es que es probable (en un grado mayor o menor) que si los enunciados fácticos son verdaderos (o las normas, válidas o correctas), entonces lo sea también la conclusión"[50].

Este esquema es aplicable a los seres humanos que argumentan en sentido propio, pero no necesariamente a las máquinas. Los algoritmos que permiten que estas ejecuten las funciones les permitirán llegar a esa misma conclusión o a otra diversa. El paso de un dato a una información expresada en una conclusión, un pronóstico o un contenido queda entregada al funcionamiento del propio algoritmo, de acuerdo con su programación inicial o con la forma en la que la máquina "aprendió" a razonar. Qué hay en ese proceso que permita confiar en esa conclusión no es del todo claro, aspecto sobre el que se volverá más adelante.

---

[48] Albers (2014) p. 222. En el mismo sentido, Betkier (2019) p. 9.
[49] Atienza (2013) p. 173.
[50] *Op. cit*. p. 177.

Como la información generada se refiere a hechos en el marco de una investigación, el carácter inductivo de ella está normativamente determinado: la veracidad de una información solo podrá predicarse cuando ella sea declarada en la sentencia firme dictada por un tribunal competente. Una cuestión distinta es la del grado de confiabilidad de una inferencia, conclusión o información así generada, que apunta a cómo esa información va a influir en el proceso de adopción de una decisión por un tribunal. En otras palabras, en qué medida y bajo qué condiciones una información va a ser suficiente para formar "convicción más allá de toda duda razonable" en el tribunal llamado a enjuiciar el caso.

La recogida de los datos es un proceso que si se realiza de forma manual puede ser fatigoso, en extremo caro y poco efectivo. Si se realiza en forma automatizada, todas esas variables comienzan a mejorar: salvo por un inicial costo de instalación (adquisición de *software*, entrenamiento y, eventualmente, diseño de aplicaciones), el uso de mecanismos automatizados puede representar grandes ahorros de recursos para la investigación; la recogida automatizada de datos solo puede producir fatiga de *software* o *hardware*, pero no necesariamente de personas que estén frente a la pantalla de un computador visitando extensas listas de sitios web y registrando de ellas los datos que pueden ser relevantes; y la automatización amplía las posibilidades de efectividad, al permitir una recogida mucho mayor de datos en el mismo tiempo que tomaría hacerlo en forma manual. El proceso de estructuración, por su parte, consiste en uniformar los datos de manera que puedan ser entendidos por máquinas y comparados entre ellos. Se requiere uniformar las distintas formas y soportes en las que pueden presentarse los datos: un registro fotográfico podrá requerir reconocimiento visual por máquinas que interprete la imagen, un registro de audio va a requerir reconocimiento de lenguaje natural para obtener una transcripción y así obtener el dato desde ese registro, etc. Por último, el proceso de análisis es el que permite estructurar conclusiones sobre la base de los datos, esto es, obtener información que, hasta ese momento, era desconocida. El proceso de inferencia que concluye con la obtención de información puede hacerse por medio de aplicaciones de inteligencia artificial, caso en el que estaremos en presencia de un análisis automatizado de la información. Aquí, los procesos lógicos de vincular un dato con otro y extraer de ellos conclusiones quedan entregados a una aplicación sin una intervención humana directa. Pero también puede hacerse de modo manual o con ayuda de aplicaciones, pero con una intervención directa de operadores humanos. Cada una de estas formas de análisis será especialmente incidente en la oportunidad procesal en la que la información generada por el ciberrastreo deba ser objeto de control jurisdiccional para determinar su pertinencia y admisibilidad como prueba a ser rendida en el juicio.

La información, esto es, las conclusiones obtenidas de los datos, puede cumplir varias funciones en una investigación penal. En primer lugar, puede

orientar (o contribuir) la propia investigación, mostrando otras líneas alternativas de indagación o sugiriendo abandonar una ya iniciada, o mostrar la necesidad o la conveniencia de realizar otras diligencias de investigación. En segundo lugar, la información generada por el ciberrastreo puede ser aplicada de manera directa en la construcción del hecho delictual, que es la base de la imputación acusatoria o puede ser usada para simular cursos causales hipotéticos para la construcción de escenarios fácticos alternativos para establecer (y probar, eventualmente) la relación de causalidad como base de la imputación.

En síntesis, el ciberrastreo analítico permite la búsqueda y recolección masiva de datos provenientes de las interacciones de un investigado en el ciberespacio y su análisis por *big data* permite inferir informaciones, en principio, desconocidas. Sin el uso de tales tecnologías, la construcción inferencial de información sería imposible o extremadamente costosa[51].

A mi juicio, hay dos elementos que, con referencia al ordenamiento jurídico chileno, describen al ciberrastreo como técnica de investigación: tanto su carácter intrusivo como desregulado. Estas características serán analizadas en este apartado.

III. Carácter intrusivo

La primera característica del ciberrastreo analítico es su naturaleza intrusiva, esto es, que es una técnica de investigación "limitativa de derechos"[52]. En consecuencia, el ciberrastreo expresa el conflicto propio de todas las técnicas intrusivas, esto es, la colisión entre el interés de dar cumplimiento a las reglas de convivencia expresadas democráticamente en la ley, por un lado, y la vigencia de los derechos fundamentales de las personas, por otro[53].

La relación de conflicto que le otorga a la técnica de ciberrastreo un carácter intrusivo puede construirse analíticamente por dos vías: la primera, con relación a la restricción que su ejecución implica directamente sobre el derecho a la intimidad; la segunda, a la naturaleza disruptiva que tiene dicha técnica que altera la necesaria simetría de las relaciones entre investigación/acusación y defensa, lo que representa un riesgo potencial inminente a la vigencia efectiva del debido proceso.

[51] Custers (2021) p. 65.
[52] Horvitz y López (2002) p. 507 ss.
[53] Duce y Riego (2007) p. 218.

1. Relación de conflicto del ciberrastreo con la intimidad

Este primer aspecto del fundamento de la tesis del carácter intrusivo del ciberrastreo surge de la interrelación de tres dimensiones:

a) la naturaleza del dato,
b) su accesibilidad y disponibilidad y
c) el escenario digital en el que esos datos son almacenados, distribuidos o accedidos.

Estos tres elementos conducen a la necesidad de reconstruir el contenido esencial del derecho a la intimidad para adaptarlo al mundo digitalizado en el que nos desenvolvemos, si lo que se quiere es proteger efectivamente tal derecho.

1.1. Naturaleza del dato

La dimensión de la naturaleza del dato tiene un carácter esencialmente normativo, cuya determinación se hace con referencia a las regulaciones contenidas en la Ley n.° 19628, sobre protección de la vida privada (1999)[54]. Este cuerpo normativo prevé un concepto implícito de "dato" (dato genérico o "dato personal en general"[55]) que correspondería al género y que consiste en cualquier registro sobre un hecho de la realidad. A partir de este género, la ley regula dos especies: "dato personal" y "dato sensible". El primero, es el relativo "a cualquier información concerniente a personas naturales, identificadas o identificables", mientras que el segundo, es aquel que se refiere a las

[54] Jervis (2005) propone una taxonomía de los datos basada en la Ley n.° 19628, distinguiendo entre datos personales provenientes de fuentes accesibles al público, datos personales tratados por personas jurídicas privadas, datos personales relativos a obligaciones de carácter económico, financiero, bancario o comercial, datos personales sensibles y datos personales en general. Su autora construye esta taxonomía usando como "criterio basal distinguidor la mayor o menor exigencia de la autorización del titular de los datos, como elemento legitimante del tratamiento de datos personales", p. 118. Aunque tal propuesta tiene, desde luego, un completo respaldo normativo, a mi juicio tiene una capacidad disminuida para describir la fenomenología en torno al tratamiento de los datos y a su difusión a través de las tecnologías de información y comunicación y, en particular, de las redes informáticas como internet. En efecto, el problema es que la taxonomía propuesta se construye desde la expectativa de pleno cumplimiento de la legislación o, en otras palabras, de un mundo real plenamente simétrico con las estructuras jurídicas. Sin embargo, tal simetría no existe en realidad (si existiera, no habría conflictos jurídicos) y es por ello que es posible encontrarse en la red con datos sensibles, aunque su tratamiento (y más aún su publicación) no está permitida por la ley por regla general. Esa es la razón por la que es preferible distinguir entre la naturaleza (jurídica) del dato y su disponibilidad, como se propone en este artículo.

[55] Jervis (2005) p. 144.

> "características físicas o morales de las personas o a hechos o circunstancias de su vida privada o intimidad, tales como los hábitos personales, el origen racial, las ideologías y opiniones políticas, las creencias o convicciones religiosas, los estados de salud físicos o psíquicos y la vida sexual" (art. 2.º, literales f y g de la Ley n.º 19628, respectivamente).

No cabe duda que la protección jurídica que dispensa el derecho a la intimidad está especialmente reforzada con relación a los datos sensibles, pues dicha ley prohíbe, por regla general, el tratamiento de tales datos[56]. Los datos personales tienen un menor nivel de protección por parte del derecho a la intimidad, en la medida que, por ejemplo, a diferencia de los datos sensibles, el tratamiento de los datos personales sí está permitido, cumpliéndose determinados requisitos y supuestos. Debe concluirse, en consecuencia, que el dato genérico no tendría, *prima facie*, protección por el derecho a la intimidad.

Esta primera dimensión es insuficiente para construir un argumento en apoyo a la tesis del carácter intrusivo del ciberrastreo, porque ella tiene un carácter estático y desconoce la realidad fenomenológica del ciberespacio y el tratamiento digitalizado de los datos.

1.2. Accesibilidad y disponibilidad

Por eso es necesaria una segunda dimensión, que consiste en determinar la accesibilidad digital de los datos. Esta segunda dimensión no tiene carácter jurídico, sino que está determinada por aspectos técnicos de las aplicaciones que funcionan en entornos de redes de datos –en particular, de internet– y consiste en la variable de accesibilidad o disponibilidad de los datos. Independiente de su naturaleza jurídica, de acuerdo con su disponibilidad, un dato puede esta almacenado en una fuente abierta y ser de acceso público o en una fuente cerrada, caso en el que el dato será de acceso restringido. Aunque no existe unanimidad sobre el contenido exacto del concepto de dato de fuente abierta[57], existe cierto acuerdo en que él se compone de dos elementos centrales: un dato es de fuente abierta cuando está abierto al público y es una información secundaria, en el sentido que ella fue recopilada o registrada por otros[58]. En consecuencia, un dato de acceso público –por estar registrado en una fuente abierta– puede ser un dato genérico, un dato personal o un dato sensible.

---

[56] Jervis (2005) p. 137.

[57] Bandler & Merzon (2020) p. 179 ss consideran como "fuentes abiertas" a motores de búsqueda, redes sociales, medios de comunicación, registros judiciales, registros públicos de propiedad, registros civiles (nacimiento, matrimonios, defunciones, etc.), mapas en línea, bases de datos para profesionales, the wayback machine (archive.org), etcétera.

[58] Chermak, Freilich, Parkin & Lynch (2012) p. 194; Greene-Colozzi, Freilich & Chermak (2021) p. 170.

Esta dimensión significa que en el mundo en línea el criterio tradicional, de que el acceso a información no disponible públicamente implica una restricción a la intimidad no sirve. En la datosfera es posible encontrar datos no protegidos por la intimidad, como aquellos que el titular ha decidido a prpósito hacer públicos, pero también se puede hallar datos jurídicamente protegidos en atención a su contenido. Asimismo, la reunión de datos de acceso público puede producir información que, por su propia naturaleza, estaría protegida. En el ejemplo de la persona enferma de cáncer (§III), su diagnóstico es un dato protegido porque es sensible según la definición del literal g) del art. 2.° de la Ley n.° 19628.

1.3. Escenario digital

Por último, la tercera dimensión a considerar –conceptualmente vinculada con la anterior– es la del escenario digital en el que los datos son almacenados, se distribuyen o son accedidos. Esta también tiene un carácter técnico y consiste en la arquitectura que tiene internet. Según el nivel de accesibilidad a la información, más o menos restringido, se han descrito tres categorías en capa: *Surface Web*, *Deep Web* y *Dark Web*[59], expresiones que, a menudo, "se usan indistintamente para denotar un amplio espectro de conceptos un tanto exclusivos, con definiciones cambiantes"[60].

Un internauta común que visita distintos sitios web, sin necesidad de superar barreras tecnológicas de acceso ni de poseer conocimientos especiales, estará moviéndose en la primera de esas capas: la *Surface Web*. Esta es la que todas y todos conocemos como internet y en la que navegamos habitualmente.

Cuando ese internauta decide entrar a su cuenta en el banco, ver una película en plataformas de video *streaming* (como Netflix o Amazon Prime) o publicar información en redes sociales (como Facebook, Instagram o TikTok), está navegando en la *Deep Web*, porque el común denominador de los servicios antes mencionados es la exigencia de una clave de acceso, ya sea que ella esté o no asociada al pago del servicio. La *Deep Web*, término acuñado por Michael Bergman[61], se refiere a aquel sector o sección de internet en el que reside la mayor parte de su información, que requiere que el usuario proporcione determinada información (como una clave de acceso) para obtener acceso a recursos específicos[62]. Las páginas de la *Deep Web* no pueden ser encontradas usando los motores de búsqueda debido a su legibilidad, su naturaleza dinámica o su contenido propietario:

[59] Lukings & Lashkari (2022) p. 153.
[60] Dalins, Wilson & Carman (2018) p. 62.
[61] Bergman (2001).
[62] Cole (2017b) p. 80.

> "el problema de la legibilidad tiene que ver con los tipos de archivos; el tema de la naturaleza dinámica tiene que ver con actualizaciones constantes de contenido; y el tema de los contenidos propietarios tiene que ver con la idea de sitios freemium o de pago por uso que requieren registro con usuario y contraseña"[63].

La capa oscura de internet es la *Dark Web*. Es una pequeña porción de la *Deep Web* en la que, gracias a distintas tecnologías[64], sus sitios web no pueden ser indexadas por los motores de búsqueda y, por tanto, no pueden ser encontrados por estos[65], y que permiten a sus usuarios interactuar en ellos de forma anónima[66] gracias a técnicas de encriptación y de ciberseguridad[67]. No todo el material existente en la *Dark Web* es ilícito, aunque sus características técnicas y el anonimato que ella proporciona actúan como importantes elementos criminógenos[68].

En todas las capas de internet pueden encontrarse todos los tipos de datos, tanto de acuerdo con su naturaleza jurídica como de su accesibilidad; la única diferencia es que para consultar los datos de la *Dark Web* se requieren elementos tecnológicos específicos (un navegador especial como: Tor, Subgraph, I2P, Whonix, Tails o Waterfox) y conocimientos especiales, pero fáciles de obtener.

1.4. Reconstrucción del contenido de la intimidad

Las tres variables descritas en los subapartados previos llevan a la necesidad de reconstruir el contenido del derecho a la intimidad para dotarlo de efectiva fun-

---

[63] Mead & Agarwal (2022) p. 901.

[64] Como lo explican Mead & Agarwal (2022) p. 902, tales tecnologías son: "(1) El host de la red no utiliza la configuración de enrutador estándar (protocolo de puerta de enlace fronteriza); (2) la dirección IP del servidor host no tiene un punto de entrada DNS estándar porque no está asignado; (3) El servidor host se ha configurado para no responder a los ping del Administrador de contactos inteligente; y (4) se emplean técnicas de Fastfluxing DNS que permiten que la dirección IP del servidor host cambie continua y rápidamente".

[65] Edwards (2020) p. 237.

[66] Cole (2017a) p. 79.

[67] Lukings & Lashkari (2022) p. 150.

[68] Una descripción fenomenológica de los delitos cuya comisión se producen más comúnmente en la *Dark Web* puede encontrarse en Liggett, Lee, Roddy & Wallin (2020), atrayendo el desarrollo de actividades criminales, Lukings & Lashkari (2022) p. 150. Tal como lo pone de manifiesto Miró (2019) p. 108, el desarrollo de actividades criminales o ilícitas no es patrimonio exclusivo de la *Dark Web*: "podemos encontrar patrones a nivel meso en el ciberespacio como tráfico de armas o drogas en la Dark web; radicalización en Telegram; odio en Twitter; pornografía infantil en foros; sexting en Snapchat etc.".

cionalidad protectora en un mundo altamente digitalizado. En otras palabras, aunque muy lejos ha quedado ya la original concepción de la intimidad como secreto o confidencialidad[69], el surgimiento de las tecnologías de la información y el potenciamiento de las capacidades de tratamiento que permite la IA, requieren ahora analizar de qué forma debería protegerse la intimidad, es decir, cuál es su contenido esencial. La vigencia de ese derecho se enfrenta a nuevas realidades tecnológicas como la adopción de decisiones de forma automatizada, el reconocimiento facial, el procesamiento de imágenes de vídeo o las tecnologías de *blockchain*[70] y el *scraping* asistido por IA.

Tanto en la dogmática como en los desarrollos jurisprudenciales ha existido la tendencia a considerar una radical diferenciación entre las conductas ejecutadas en público y las ejecutadas en privado; y que el derecho a la intimidad solo protege a estas últimas, toda vez que solo de ellas puede desprenderse una expectativa razonable de privacidad. En el sistema estadounidense, esta distinción se conoce bajo la fórmula de la regla *no privacy in public*[71]. Una clara aplicación de estas ideas es la posición de John Bandler y Antonia Merzon, para quienes:

> "cuando personas o entidades colocan información en el dominio público, consienten, si no buscan, el acceso público. Como resultado, no hay expectativas de privacidad que considerar ni barreras legales para re copilar esta evidencia"[72].

De acuerdo con esta doctrina, deberían distinguirse tres formas en las que se encuentra la información en internet.

Una primera, es la información divulgada y que es accesible libremente por cualquier internauta. El ejemplo paradigmático serían las publicaciones en un blog, las fotografías difundidas en Instagram o las historias puestas en Facebook a las que se tiene acceso público. O aquella información que no es relativa o perteneciente a personas determinadas[73].

Una segunda forma sería aquella información publicada, pero protegida por algún mecanismo de acceso a ella, de pública obtención. Sería el caso de la información contenida en sitios, aplicaciones o redes sociales, pero a la que se puede acceder solo mediante una suscripción, gratuita o pagada, por lo general pensadas para convertir a un internauta en usuario de esa aplicación o plataforma.

Finalmente, una tercera forma sería el de la información contenida en bases de datos accesibles a través de la red, pero que está reservada y disponible

[69] Abdel-Basset, Moustafa, Hawash & Ding (2022) p. 28.
[70] Kiesow (2021) p. 271 ss.
[71] Xiao (2021) p. 703; Parks (2022).
[72] Bandler & Merzon (2020) p. 114.
[73] Parks (2022) p. 915.

solo para su titular o para quienes él decida. Historiales médicos, cuentas bancarias o archivos almacenados en nubes de datos serían ejemplos de esta tercera forma.

La distinción público/privado en la que se basa la regla *no privacy in public* no se ajusta propiamente a la realidad de la información digital. Tal distinción se creó, en realidad, para el mundo físico, donde es posible aplicarla, de todos modos, no exenta de complicaciones. Pero en el mundo digital en red, las cosas funcionan distinto. En el contexto del ciberespacio, se ha tendido a ampliar la esfera de lo público, de modo que se ha concluido que la información publicada en la red y accesible, con o sin membresía, no está protegida por la intimidad.

Aplicando esta premisa a la investigación penal efectuada por policías o fiscales, se puede concluir, *prima facie*, que la obtención de información desde ciberfuentes accesibles no restringe el derecho a la intimidad.

Sin embargo, creo que el asunto no puede ser tan fácilmente resuelto. Tal como lo ha propuesto Daniel Solove, la vigencia efectiva del derecho a la intimidad en un mundo digitalizado requiere que él sea escindido del "paradigma del secreto" en que se fundamenta la regla de *no privacy in public*. En este paradigma, solo:

> "se produce una violación de la privacidad cuando se revelan datos ocultos a otros. Si la información no se oculta previamente, la recopilación o difusión de la información no implica ningún interés de privacidad"[74].

No obstante, esta visión restringida de la intimidad, afirma el autor, "ha limitado el reconocimiento de las violaciones de la privacidad". La vigilancia, aun en ámbitos públicos, produce un desbalance del poder que hace aumentar el riesgo de un abuso de poder[75]. David Wall sostiene que el desarrollo de internet y de las tecnologías conectadas en red han producido una reedición del modelo de vigilantismo en un doble sentido: por un lado, un panopticismo donde la masa (los "muchos") no sabe que unos "pocos" los están vigilando y, con ello, moldeando sus comportamientos y, por el otro, un sinopticismo, donde "muchos" también pueden observar y vigilar a los "pocos" con el mismo efecto sobre sus comportamientos[76]. Pamela Ugwudike advierte que el uso de tecnologías como la inteligencia artificial es capaz de fomentar una "dominación epistémica digitalizada", esto es:

> "el poder y la capacidad de actores estatales y no estatales influyentes para crear algoritmos basados en datos, cuyas inferencias a partir de

[74] Solove (2006) p. 497.
[75] *Op. cit*. p. 487.
[76] Wall (2014) pp. 228-229.

> patrones sobre conjuntos de datos, construyen discursos clave que evolucionan hacia el conocimiento sobre el riesgo y la gestión eficiente de grupos etiquetados como riesgosos"[77].

Siguiendo la tesis de Daniel Solove, estimo que los datos que pueden rastrearse en la red deben considerarse *prima facie* protegidos por la intimidad, en la medida que el ciberrastreo permite construir información sobre una persona que, de otra forma, sería desconocida. Se trata de una especie de puzle que se va armando a partir de distintas unidades individuales de información. Por eso, una de las piedras angulares sobre las que se construyen los sistemas jurídicos de protección a la intimidad es el principio de limitación del propósito, según el cual los datos solo deben ser recolectados para propósitos especificados, explícitos y legítimos[78], al menos en el modelo europeo de la General Data Protection Regulation. Esto ocurre porque las personas y las organizaciones están permanentemente en contacto con el mundo digital, y esa relación se traduce en un aporte permanente de información que hacen de manera consciente o inconsciente, voluntaria e involuntaria. Las acciones deliberadas de interacción con el mundo digital, como hacer una compra en línea, enviar un correo electrónico, navegar por internet o publicar información en una red social, dejan huellas de esas actividades (huella digital activa)[79]. Pero también dejan esas huellas las acciones no necesariamente conscientes, como la información sobre desplazamientos que recogen automatizadamente los *smartphones* o los datos que generan los dispositivos inmersos en el ecosistema de la Internet of Things (IoT) (huella digital pasiva)[80], aspectos propios de lo que se denomina Web 3.0[81].

El conjunto de esas huellas es lo que se denomina huella digital o digital footprint. No por casualidad, el modelo de negocios de Gmail fue disruptivo respecto de su competencia al momento del lanzamiento de ese servicio de correo electrónico en 2004. Gmail ofreció gratuitamente a sus usuarios grandes cantidades de almacenamiento de datos en sus cuentas de correo, en un momento en el que los otros servicios estaban buscando cobrar a sus usuarios por ese almacenamiento[82].

Por ejemplo, si para acreditar la existencia de un delito tributario debe probarse que el contribuyente pasó más de seis meses en el estado que reclama

---

[77] Ugwudike (2021) pp. 81-82.

[78] Colonna (2014) p. 299; Busser (2014) p. 103 ss; Terwangue (2022) p. 20 ss; Paal (2022) pp. 294-295.

[79] Rathi, Lata, Soni, Jain & Telang. (2023) p. 277.

[80] *Op. cit*. p. 278.

[81] Genlin & Baker (2021) p. 127.

[82] Andrejevic (2018) p. 93.

los impuestos evadidos, el ciberrastreo es una buena alternativa. A partir del rastreo de la información en la web, la permanencia del contribuyente en el territorio físico de un estado se puede inducir, esto es, afirmar, con una alta probabilidad de veracidad, a partir de elementos como las fotos que haya publicado en sus redes sociales en un periodo determinado, los puntos IP desde los que se haya conectado a la red (el *web scraping* puede acceder a metadatos normalmente ocultos al ojo de un internauta), las historias que haya publicado, etcétera.

Esta tesis es la que el Tribunal Constitucional federal alemán desarrolló a propósito de la sentencia del censo de 1983[83], dando origen a la autodeterminación informativa como concepción de la intimidad. Esta misma doctrina fue aplicada en la sentencia de 2006 sobre perfilamiento de potenciales terroristas[84].

En suma, un dato aislado obtenido desde la red no está, por sí mismo, protegido por la intimidad; pero sí lo está la información que se pueda conocer inductivamente a través del análisis de un conjunto de información. A través de la unión de varios datos recogidos de distintas fuentes, se puede conocer, por ejemplo, el estado de salud de una persona, su tendencia política, sus relaciones interpersonales, su orientación sexual, etc. De acuerdo con la taxonomía propuesta por Daniel Solove, el *web scraping* puede ser asimilado a la vigilancia. Ciertas formas de vigilancia están específicamente reguladas por los ordenamientos jurídicos. En el caso chileno, por ejemplo, la interceptación de comunicaciones postales, electrónicas o digitales, requiere de autorización judicial previa en el marco de una investigación penal[85]. Con la misma exigencia está regulado el acceso a los registros que un proveedor de servicios de internet debe, por ley, mantener de las conexiones de sus clientes. Pero el ciberrastreo como tal, no está previsto por la legislación procesal penal.

2. Naturaleza disruptiva del ciberrastreo

Los beneficios *pro societatis*, generados por la utilización del ciberrastreo en contextos de inteligencia artificial en la investigación penal, producen externalidades que afectan el enfoque *pro libertatis* que se encuentra en la base de un proceso penal democrático.

Ciertas características de las tecnologías de IA que se comunican al ciberrastreo como técnica de investigación, constituyen el problema: complejidad por volumen, sesgos y opacidad algorítmica. Estos tres elementos, en conjun-

[83] *BVerfGE* 65, 1-71. BVerfG, Order of the First Senate of 15 December 1983 - 1 BvR 209/83 -, paras. 1-214 [http://www.bverfg.de/e/rs19831215_1bvr020983en.html].

[84] *BVerfGE* 115, 320-381. BVerfG, Order of the First Senate of 4 April 2006 - 1 BvR 518/02 -, paras. 1-182 [http://www.bverfg.de/e/rs20060404_1bvr051802en.html].

[85] Alvarado (2014).

to, producen un sustancial aumento de los costos y cargas que debe soportar la defensa. Al respecto, la dogmática ya ha llamado la atención. Steven Wright estima que el avance continuo de la tecnología de IA y la recopilación de datos determina una transición tecnológica potencialmente disruptiva para la que espera se desarrollen

> "colectivamente la ética de la información que necesitaremos para evaluar y guiar esta transformación tecnológica en beneficio de todos"[86].

La información proveniente del ciberrastreo puede llegar a ser extremadamente voluminosa, dificultando la tarea de la defensa de procesarla y confrontarla dentro del proceso de preparación de la estrategia de defensa y la construcción de la teoría del caso. De esta forma, el uso de tales tecnologías produce una complejización de la defensa, que, si no cuenta con los mismos recursos tecnológicos con los que contó la fiscalía para generar esa información, se verá impedida de controlarla.

La inexistencia de metodologías que permitan identificar y anular los sesgos de que puede adolecer la IA, principalmente aquella basada en aprendizaje de máquinas (*machine learning*), hace que tales sesgos puedan proyectarse hasta la sentencia definitiva y que puedan afectar la imparcialidad judicial. No debe olvidarse, como lo ha puesto de relieve la doctrina, la existencia de:

> "suposiciones epistemológicas defectuosas que se basan en varios mitos, como la supuesta representatividad de los grandes datos, la idea de que tales conjuntos de datos carecen de sesgo humano y carecen de especificidad de contexto»[87].

Ya sea por protección de propiedad intelectual, por el propio funcionamiento del algoritmo que determina la impredictibilidad e inexplicabilidad de sus resultados o la necesidad de contar con recursos humanos muy capacitados para comprenderlos, la opacidad algorítmica dificulta el ejercicio de la defensa. Aunque se han hecho esfuerzos por avanzar hacia una IA explicable (XAI, eXplainable Artificial Intelligence)[88], una tecnología capaz de dar explicaciones de sus conclusiones o resultados[89], lo cierto es que la IA no ha logrado, al menos por el momento, desprenderse de su efecto de caja negra.

En consecuencia, el uso en la investigación penal del ciberrastreo y, en general de tecnologías de IA, puede producir dos efectos profundamente disruptivos.

---

[86] Wright (2020) p. 2163.
[87] Ugwudike (2021) p. 82.
[88] Barredo, Díaz-Rodríguez, Del Ser, Bennetot, Tabik, Barbado, García (2020).
[89] Baum, Mantel, Schmidt & Speith (2022).

El primero, la instalación de una discriminación estructural en el proceso penal por la generación de un escenario en el que solo imputados adinerados puedan financiar una defensa efectiva frente a una persecución penal dotada de instrumentos de IA y, por lo que el derecho a la defensa efectiva pase a ser un privilegio de unos pocos.

El segundo efecto disruptivo, el uso masivo de IA opera como un incentivo para malas prácticas en el proceso penal, pues imputados que no pueden solventar una defensa efectiva buscarán medios alternativos, algunos de ellos, incluso, ilícitos, como la generación de pruebas falsas, para hacer frente a la imputación penal. Aunque este problema tiene un carácter metajurídico puede, sin embargo, tener una solución normativa que pasa por extender las reglas de exclusión de prueba por infracción de derechos fundamentales, con que cuentan los ordenamientos jurídicos, a una investigación inabordable para una defensa.

Un sistema de persecución penal democrático no debería estar dispuesto a tolerar una investigación que disponga de información que es imposible de confrontar por la defensa, ya sea por su volumen, su complejidad o su inexplicabilidad. Asimismo, esa información tampoco podría servir, razonablemente, para la formación de convicción por un tribunal. En la base del problema se encuentra el valor de igualdad de armas que encierra el debido proceso. Siguiendo la analogía de Macarena Vargas y Claudio Fuentes, una fiscalía dotada de herramientas de IA sería como un partido de fútbol en el que fiscalía cuenta con once jugadores y se enfrenta a una defensa que cuenta con solo seis de ellos[90].

IV. Carácter desregulado del ciberrastreo

La segunda característica del ciberrastro analítico como técnica de investigación penal consiste en que, como anticipamos, carece en el ordenamiento procesal penal chileno de una regulación legal específicamente diseñada de acuerdo con su propia naturaleza.

A pesar de las variadas modificaciones que la Ley n.º 21577, que fortalece la persecución de los delitos de delincuencia organizada, establece técnicas especiales para su investigación y robustece comiso de ganancias (2023), introdujo a las técnicas de investigación vinculadas con el secreto de las comunicaciones y los sistemas informáticos, no previó una disposición que regulara el rastreo de información en la web o en otras redes informáticas.

[90] Vargas y Fuentes (2018) pp. 141-42, 147.

En el derecho comparado, una ley de 1992 introdujo los parágrafos 98a y 98b a la ordenanza procesal alemana y que constituyen la regulación legal del *Rasterfahndung*[91] que, en lo esencial, coincide con el ciberrastreo. En Alemania, el rastreo de información digital y su análisis (*Rasterfahndung*) como técnica de investigación se usó sin que ella tuviera reconocimiento legal expreso. La situación cambió con el dictado de la sentencia del Tribunal Constitucional federal de 1983 sobre el censo de población, que puso en evidencia la necesidad de que el *Rasterfahndung* tuviera un reconocimiento legal porque implicaba una restricción de la intimidad.

En todo caso, debe destacarse que la ausencia de una regulación del ciberrastreo como técnica de investigación no entrega información sobre su licitud o ilicitud en general y, en consecuencia, si ella puede o no utilizarse válidamente en Chile. La protección jurídica dispensada a los sistemas informáticos a través de delitos de acceso ilícito, medidas administrativas de control, el derecho de daños o las propias condiciones de uso de los sitios web y plataformas, pueden contener prohibiciones del ciberrastreo, más o menos extensas, tema que, por su amplitud, no es posible abordar en esta oportunidad[92].

V. Admisibilidad del ciberrastreo

Partiendo de su carácter desregulado del ciberrastreo, pueden adoptarse, al menos, dos posiciones jurídicas. La primera, que negaría la posibilidad de utilizar tal técnica por los efectos contrarios que ella produce para la vigencia de derechos fundamentales en el marco de la actividad del Estado-investigador y Estado-juzgador. La segunda, que su utilización estaría permitida, aplicándole analógicamente las regulaciones existentes para otras técnicas de investigación de carácter intrusivo. Dentro de ellas, la candidata más próxima es la regulación de las interceptaciones de comunicaciones.

1. Prohibición de uso

El fundamento para esta primera posición, esto es, la prohibición de usar el ciberrastreo como técnica de investigación, no estaría solo en su carácter intrusivo, sino, además, en sus características y en las dificultades que conlleva para la lógica adversarial del proceso.

[91] Cano (2003) pp. 2-3.

[92] Para el entorno angloamericano, véase, por ejemplo, Macapinlac (2019); Krotov, Johnson & Silva (2020).

Sostener esta posición de prohibición de uso del ciberrastreo deberá lidiar con los discursos *pro societatis* basados en la maximización de la seguridad, poco afines a la vigencia de los derechos fundamentales de los delincuentes.

2. Aplicación analógica de otras regulaciones

Podríamos sostener que la ausencia de una autorización legal expresa de una técnica de investigación intrusiva no es obstáculo para su utilización, pues su admisibilidad puede fundarse en la regla genérica de autorización judicial previa prevista en el art. 9.º del *CPP*; y los aspectos operativos pueden ser suplidos recurriéndose analógicamente a la regulación de otras medidas intrusivas.

Con relación a los aspectos operativos, se puede contraargumentar que se espera una reglamentación sobre aspectos operativos particulares de una determinada técnica de investigación. Esa podría ser la razón por la que el legislador, a pesar de la regla genérica del art. 9.º del *CPP*, también ha regulado, en particular, actuaciones intrusivas, como los exámenes corporales, la entrada y registro en lugares cerrados, la incautación de documentos protegidos por la intimidad, la retención o incautación de correspondencia, la interceptación de comunicaciones, la interceptación y grabación de comunicaciones, de conversaciones o imágenes obtenidos en lugares cerrados o el registro remoto de equipos informáticos. En este plano operativo quedan sin solución legal cuestiones como, por ejemplo, el plazo dentro del cual se podría ejecutar el ciberrastreo, el periodo en el que puede recogerse información (¿desde hace seis meses o desde hace seis años, por ejemplo?), la forma de registro de la información rastreada o las aplicaciones que se usarán para su procesamiento.

De todos modos, y ahora fuera de los aspectos operativos, admitir sin más el ciberrastreo, aun aplicando analógicamente las regulaciones previstas para otras técnicas intrusivas, producirá como consecuencia una recarga de las funciones de control sobre la investigación que la legislación ha atribuido a los tribunales y a la defensa, con posibilidades inciertas de que tales controles lleguen a operar de manera eficaz.

A modo de conclusión: una tercera alternativa

El carácter disyuntivo y extremo de las dos posiciones jurídicas antes expuestas no es compatible, en todo caso, con la realidad. Una propuesta con posibilidades de poder ser aplicada debe considerar los elementos culturales de los actores del sistema de persecución penal y su efectividad, y las expectativas sociales,

por un lado; por el otro, con el imperativo democrático de vigencia de los derechos fundamentales.

Sobre la base de esas premisas, creo que la solución adecuada dentro del ordenamiento jurídico chileno es aceptar el uso del ciberrastreo, pero sometido a la condición de un riguroso control jurisdiccional sobre la admisibilidad de la prueba obtenida mediante dicha técnica en la audiencia de preparación de juicio oral, que es la etapa previa al juzgamiento propiamente tal.

En otras palabras, que la solución consiste en adoptar como criterios operativos que los antecedentes provenientes del ciberrastreo pueden ser admitidos para rendirse como prueba en la audiencia de juicio, solo si dichos antecedentes son capaces de superar un test de control que aborde, al menos, los siguientes aspectos:

a) garantías suficientes sobre la completitud de la información registrada y utilizada en los análisis;
b) que el órgano de persecución penal que ofrece esos antecedentes sea capaz de explicar, con fundamento técnico, la forma en la que los algoritmos usados buscaron la información, la recopilaron y la procesaron, de modo que sean capaz de despejar la opacidad que pueda existir sobre dicha información;
c) que el mismo órgano de persecución penal sea capaz de demostrar que la búsqueda, recolección y análisis de la información se hizo con sujeción al deber legal de objetividad, investigando con igual celo los antecedentes que fundamentan la responsabilidad penal, como la que la eximen o atenúan; esto es, ausencia de sesgos.
d) y, por último, demostrar que la búsqueda y recolección se hizo existiendo previamente una autorización judicial que, en su momento, hizo el test de ponderación y determinó que la intromisión en la intimidad del imputado estaba justificada.

Estos criterios no son más que expresión del test de proporcionalidad que el juez de garantía está obligado a hacer, no solo al momento de autorizar motivada y justificadamente una medida intrusiva[93], sino, también, al decidir la inclusión de una determinada información que restrinja derechos fundamentales en el auto de apertura de juicio oral[94].

Si la información proveniente del uso del ciberrastreo no es capaz de sortear de manera exitosa los aspectos previamente señalados, el tribunal de control debería declarar la imposibilidad de valerse de ellos como medios de prueba

[93] Núñez, Beltrán y Santander (2019) p. 154.

[94] En el mismo sentido de la exigencia de un test de proporcionalidad, Núñez, Beltrán y Santander (2019) p. 160.

en el juicio, recurriendo a las reglas de exclusión de prueba que, en el derecho chileno, se encuentra en el art. 276 del *CPP*[95].

Bibliografía

Abdel-Basset, Mohamed; Moustafa, Nour; Hawash, Hossam & Ding, Weiping (2022): *Deep Learning Techniques for IoT Security and Privacy (Cham: Springer)*. Available in https://doi.org/https://doi.org/10.1007/978-3-030-89025-4 [fecha de consulta: 10 de abril de 2024].

Ackerman, Gary A, & Pinson, Lauren E. (2016): "Speaking Truth to Sources: Introducing a Method for the Quantitative Evaluation of Open Sources in Event Data". *Studies in Conflict & Terrorism*, vol. 39, No. 7-8: pp. 617–640. Available in https://doi.org/10.1080/ 1057610X.2016.1141000 [fecha de consulta: 10 de abril de 2024].

Albers, Marion (2014). "Realizing the Complexity of Data Protection", in Gutwirth, Serge; Leenes Ronald & De Hert, Paul (eds.), *Reloading Data Protection. Multidisciplinary Insights and Contemporary Challenges* (Dordrecht: Springer). Available in https://doi.org/https://doi.org/10.1007/978-94-007-7540-4_11 [fecha de consulta: 13 de abril de 2024].

Alexandrou, Alex (2022): *Cybercrime and information technology. Theory and practice: The computer network infrastructure and computer security, cybersecurity laws, Internet of Things (IoT), and mobile devices* (Boca Raton: CRC Press).

Alvarado Urízar, Agustina (2014): "El control de la resolución motivada que autoriza una interceptación telefónica en Chile y duración de la medida". *Revista de Derecho de la Pontificia Universidad Católica de Valparaíso* vol. xliii: pp. 421-464.

Andrejevic, Mark (2018): "Data collection without limits. Automated policing and the politics of framelessness", in Završnik, Aleš (ed.), *Big Data, crime and social control* (Oxon: Routledge) pp. 93-107.

Aristegui Spikin, Juan (2020): "La prueba ilícita ante la bifurcación del tribunal penal". Quaestio facti. *Revista Internacional sobre Razonamiento Probatorio* vol. 1: pp. 177-198. Disponible en https://doi.org/10.33115/udg_bib/qf.i0.22369 [fecha de consulta: 10 de abril de 2024].

Asaro, Peter M. (2019): "AI Ethics in predictive policing. From models of threat to an ethics of care". I*EEE Technology and Society Magazine* vol. 38 Issue 2 June: pp. 40–53. Available in https://doi.org/https://doi.org/10.1109/mts.2019.2915154 [fecha de consulta: 10 de abril de 2024].

Atienza, Manuel (2013): *Curso de argumentación jurídica* (Madrid: Trotta).

Bandler, John & Merzon, Antonia (2020): *Cybercrime investigations. A comprehensive resource for everyone* (Boca Raton: CRC Press).

[95] Aristegui (2020); Echeverría (2011); Zapata (2004).

Barfield, Woodrow (2018): "Towards a law of artificial intelligence", in Barfield, Woodrow (ed.), *Research Handbook on the Law of Artificial Intelligence* (Cheltenham/Northampton: Edward Elgar Publishing). Available in https://doi.org/https://doi.org/10.4337/9781786439055.00011 [fecha de consulta: 10 de abril de 2024].

Barredo Arrieta, Alejandro; Díaz-Rodríguez, Natalia; Del Ser, Javier; Bennetot, Adrien; Tabik, Siham; Barbado, Alberto; García, Salvador (2020): "Explainable Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI". *Information Fusion* vol. 58: pp. 82-115.

Baum, Kevin; Mantel, Susanne; Schmidt, Eva & Speith, Timo (2022): "From Responsibility to Reason-Giving Explainable Artificial Intelligence", *Philosophy & Technology* vol. 35. Available in https://doi.org/10.1007/s13347-022-00510-w [fecha de consulta: 10 de abril de 2024].

Bergman, Michael K. (2001): "White paper: The Deep Web: Surfacing hidden value". *The Journal of Electronic Publishing* vol. 7, No. 1. Availble in https://doi.org/https://doi.org/10.3998/3336451.0007.104 [fecha de consulta: 10 de abril de 2024].

Betkier, Marcin (2019): *Privacy online, law and the effective regulation of online services* (Cambridge: Intersentia).

Blount, Kelly (2024): "Using artificial intelligence to prevent crime: implications for due process and criminal justice", *AI & Society* vol. 39: pp. 359-368. Available in https://doi.org/10.1007/s00146-022-01513-z [fecha de consulta: 10 de abril de 2024].

Boyd, Danah & Crawford, Kate (2012): "Critical questions for big data. Provocations for a cultural, technological, and scholarly phenomenon". *Information, Communication & Society* vol. 15 No. 5: pp. 662-679. Available in https://doi.org/10.1080/1369118X.2012. 678878 [fecha de consulta: 10 de abril de 2024].

Brewer, Russell; Westlake, Bryce; Hart, Tahlia & Arauza, Omar (2021): "The ethics of web crawling and web scraping in cybercrime research: Navigating issues of consent, privacy, and other potential harms associated with automated data collection", in Lavorgna, Anita & Holt, Thomas J. (eds.), *Researching cybercrimes. Methodologies, ethics, and critical approaches* (Cham: Springer). Available in https://doi.org/10.1007/978-3-030-74837-1_22 [fecha de consulta: 10 de abril de 2024].

Busser, Els de (2014): "Open Source Data and Criminal Investigations: Anything You Publish Can and Will Be Used Against You". *Groningen Journal of International Law* vol. 2 No. 2.

Cano Paños, Miguel Ángel (2003): "El Rasterfahndung en el Derecho Procesal Penal alemán y su aplicación práctica en la lucha antiterrorista". *Revista Electrónica de Ciencia Penal y Criminología*, vol. 5, n.° 6: pp. 1-14. Available in http://criminet.ugr.es/recpc/05/recpc05-06.pdf [fecha de consulta: 10 de abril de 2024].

Carrero, Jacquellena (2020): "Access granted: A First Amendment theory of reform of the CFAA Access Provision". *Columbia Law Review* No. 120: pp. 131-172.

Castro Jofré, Javier (2023): *Manual de derecho procesal penal* (Valencia: Tirant lo Blanch).

Centorame, Federica (2021): "Investigaciones criminales intrusivas y búsqueda de pruebas a través de 'software espías' en la experiencia procesal italiana", en Pereira Puigvert, Silvia y Ordóñez Ponz, Francesc, *Investigación y proceso penal en el siglo xxi: nuevas tecnologías y protección de datos* (Pamplona: Aranzadi).

Chermak, Steven M; Freilich, Joshua D.; Parkin, William S. & Lynch, James P. (2012): "American Terrorism and Extremist Crime Data Sources and Selectivity Bias: An Investigation Focusing on Homicide Events Committed by Far-Right Extremists". *Journal of Quantitative Criminology* vol. 28 No. 1. Available in https://doi.org/10.1007/s10940-011-9156-4 [fecha de consulta: 11 de abril de 2024].

Clough, Jonathan (2015): *Principles of cybercrime* (Cambridge: Cambridge University Press, 2nd edition).

Cole, Jeremy (2017a): "Dark web", in Springer, Paul J. (ed.), *Encyclopedia of cyber warfare* (Santa Barbara: ABC-CLIO).

Cole, Jeremy (2017b): "Deep web", in Springer, Paul J. (ed.), *Encyclopedia of cyber warfare* (Santa Barbara: ABC-CLIO).

Colonna, Liana (2014): "Data Mining and Its Paradoxical Relationship to the Purpose Limitation Principle", in Gutwirth, Serge; Leenes, Ronald & De Hert Paul, *Reloading Data Protection. Multidisciplinary Insights and Contemporary Challenges* (Dordrecht: Springer). Available in https://doi.org/https://doi.org/10.1007/978-94-007-7540-4_14 [fecha de consulta: 11 de abril de 2024].

Cong, Lin William; Li Beibei & Zhang, Qingquan Tony (2021): "Alternative Data in FinTech and Business Intelligence", in Pompella, Maurizio & Matousek, Roman (eds,), *The Palgrave Handbook of FinTech and Blockchain* (Cham: Springer). Available in https:// doi.org/10.1007/978-3-030-66433-6_9 [fecha de consulta: 11 de abril de 2024].

Custers, Bart (2021): "Profiling and predictions: Challenges in cybercrime research datafication", in Lavorgna, Anita & Holt, Thomas J. (eds.), *Researching cybercrimes. Methodologies, ethics, and critical approaches* (Cham: Springer) pp. 63-79. Available in https:// doi.org/10.1007/978-3-030-74837-1_4 [fecha de consulta: 11 de abril de 2024].

Dalins, Janis; Wilson, Campbell & Carman, Mark (2018): "Criminal motivation on the dark web: A categorisation model for law enforcement". *Digital Investigation*, No. 24. Available in https://doi.org/https://doi.org/10.1016/j.diin.2017.12.003 [fecha de consulta: 11 de abril de 2024].

Danaher, John (2022): "Freedom in an Age of Algocracy", in Vallor, Shannon (ed.), *The Oxford Handbook of Philosophy of Technology* (New York: Oxford University Press). Available in https://doi.org/https://doi.org/10.1093/oxfordhb/9780190851187.013.16 [fecha de consulta: 11 de abril de 2024].

Diouf, Rabiyatou; Sarr, Edouard Ngor; Sall, Ousmane; Birregah, Babiga; Bousso, Mamadou & Mbaye, Sény Ndiaye (2019): "Web scraping: State-of-the-art and areas of application". 2019 *IEEE International Conference on Big Data (Big Data)* pp. 6040-

6042. Available in https://doi.org/10.1109/BigData47090.2019.9005594 [fecha de consulta: 11 de abril de 2024].

Duce, Mauricio; Marín, Felipe y Riego, Cristián (2008): "Reforma a los procesos civiles: consideraciones desde el debido proceso y calidad de la información", en CEJA (ed.), *Justicia civil: perspectivas para una reforma en América Latina* (Santiago: CEJA): pp. 13-94. https://cejamericas.org/wp-content/uploads/2020/09/41-Justiciacivil 2008_ceja.pdf [fecha de consulta: 11 de abril de 2024].

Duce, Mauricio y Riego, Cristián (2007): *Proceso penal* (Santiago: Editorial Jurídica de Chile).

Dülger, Murat Volkan (2024): "Prevention of Discrimination in the Practices of Predictive Policing", in Kılıç, Muharrem & Kahyaoğlu, Sezer B. (eds.), *Algorithmic Discrimination and Ethical Perspective of Artificial Intelligence* (Gateway East: Springer). Available in https://doi.org/10.1007/978-981-99-6327-0_7 [fecha de consulta: 11 de abril de 2024].

Dunham, Ken & Melnick, Jim (2009): *Malicious bots. An inside look into the cyber-criminal underground of the Internet* (Boca Raton: CRC Press).

Echeverría, Isabel (2011): *Los derechos fundamentales y la prueba ilícita. Con especial referencia a la prueba ilícita aportada por el querellante particular y por la defensa* (Santiago: Ediciones Jurídicas de Santiago).

Edwards, Graeme (2020): *Cybercrime investigators handbook* (Hoboken: Wiley).

Erdoğan, Irmak (2022): "Algorithmic Suspicion in the Era of Predictive Policing", in Borges, Georges & Sorge, Christoph (eds.), *Law and Technology in a Global Digital Society* (Cham: Springer) pp. 89-102. Available in https://doi.org/10.1007/978-3-030-90513-2_5 [fecha de consulta: 11 de abril de 2024].

Faggiani, Valentina (2022): "El derecho a un proceso con todas las garantías ante los cambios de paradigma de la inteligencia artificial", *Teoría y Realidad Constitucional* n.° 50.

Farley, Erin J. & Pierotte, Lisa (2017): "Web scraping. An emerging data collection method for criminal justice researchers". *Justice Research and Statics Association* pp. 1-5. Available in http://hdl.handle.net/20.500.11990/4026 [fecha de consulta: 11 de abril de 2024].

Fleisher, Craig S. (2008): "Using open source data in developing competitive and marketing intelligence". *European Journal of Marketing*, vol. 42, No. 7-8: pp. 852-866.

Frisch, Wolfgang (2014): "Transformaciones del Derecho penal como consecuencia del cambio social". *Revista de Estudios de la Justicia*, n.° 21: pp. 15-40.

Genlin, Liang & Baker, Dennis J. (2021): "Criminalising cybercrime facilitation by omission and its remote harm form in China", in Baker, Dennis J. & Robinson, Paul H. (eds.), *Artificial intelligence and the Law. Cybercrime and criminal liability* (Oxon: Routledge): pp. 126-155. Available in https://doi.org/10.4324/9780429344015-6 [fecha de consulta: 11 de abril de 2024].

Gold, Zachary & Latonero, Mark (2018): "Robots welcome? Ethical and legal considerations for web crawling and scraping". *Washington Journal of Law, Technology & Arts* vol. 13 No. 3: pp. 275-312. Available in https://digitalcommons.law.uw.edu/wjlta/vol13/iss3/4/ [fecha de consulta: 12 de abril de 2024].

Gorro, Ken D.; Sabellano, Mary Jane G.; Maderazo, Christian V.; Ceniza, Angie M. & Gorro, Kim (2017): "Exploring facebook for sharing crime experiences using selenium and support vector machine". *ACM International Conference Proceeding Series*: pp. 218-222. Available in https://doi.org/10.1145/3176653.3176692 [fecha de consulta: 12 de abril de 2024].

Greene-Colozzi, Emily Ann; Freilich, Joshua D. & Chermak, Steven M. (2021): "Developing open-source databases from online sources to study online and offline phenomena", in Lavorgna. Anita & Holt, Thomas J., *Researching cybercrimes. Methodologies, ethics, and critical approaches* (Cham: Palgrave Macmillan). Available in https://doi.org/10. 1007/978-3-030-74837-1_9 [fecha de consulta: 12 de abril de 2024].

Guinchard, Audrey (2021): "The criminalisation of tools under the Computer Misuse Act 1990. The need to rethink cybercrime offences to effectively protect legitimate activities and deter cybercriminals", in Owen, Tim & Marshall Jessica (eds.), *Rethinking cybercrime. Critical debates* (Cham: Palgrave Macmillan). Available in https://doi.org/10.1007/978-3-030-55841-3_3 [fecha de consulta: 12 de abril de 2024].

Gundur, Rajeev V.; Berry, Mark & Taodang, Dean (2021): "Using digital open source and crowdsourced data in studies of deviance and crime", in Lavorgna, Anita & Holt, Thomas J. (eds.), *Researching cybercrimes. Methodologies, ethics, and critical approa ches* (Cham: Palgrave Macmillan). Available in https://doi.org/10.1007/978-3-030-74837-1_8 [fecha de consulta: 12 de abril de 2024].

Hayward, Keith J. & Maas, Matthijs M. (2020): "Artificial intelligence and crime: A primer for criminologists". *Crime, Media, Culture* vol. 17 No 2: pp. 209-233. Available in https://doi.org/10.1177/1741659020917434 [fecha de consulta: 12 de abril de 2024].

Holt, Thomas J.; Smirnova, Olga; Strumsky, Deborah and Kilger, Max (2014): "Case study. Advancing research on hackers through social network data", in Marcum, Catherine D. y Higgins, George E. (eds.), *Social networking as a criminal enterprise* (Boca Raton: CRC Press): pp. 145-163. Available in https://doi.org/10.1201/b16912-14 [fecha de consulta: 12 de abril de 2024].

Horvitz Lennon, Ma. Inés y López Masle, Julián (2002): *Derecho procesal penal chileno. Principios, sujetos procesales, medidas cautelares, etapa de investigación*, tomo I (Santiago: Editorial Jurídica de Chile).

Hylton, Keith N. (2019): "Digital platforms and antitrust law". *Nebraska Law Review* vol. 98 No. 2.

Jervis Ortiz, Paula (2005): "Categorías de datos reconocidas en la Ley 19.628". *Revista Chilena de Derecho Informático*, n.° 6. Disponible en https://doi.org/https://doi.org/10.5354/rchdi.v0i6.10727 [fecha de consulta: 12 de abril de 2024].

Khder, Moaiad Ahmad (2021): "Web scraping or web crawling: State of art, techniques, approaches and application". *International Journal of Advances in Soft Computing and its Applications*, vol. 13 No. 3: pp. 144-168. Available in https://doi.org/10.15849/IJASCA.211128.11 [fecha de consulta: 12 de abril de 2024].

Kiesow Cortez, Elif (2021): "Data Protection Around the World: Future Challenges", en Kiesow Cortez, Elif (ed.), *Data Protection Around the World. Privacy Laws in Action* (The Hague: Asser Press). Available in https://doi.org/https://doi.org/10.1007/978-94-6265-407-5_12 [fecha de consulta: 12 de abril de 2024].

Kokolaki, Emmanouela; Daskalaki, Evangelia; Psaroudaki, Katerina; Christodoulaki Meltini & Fragopoulou, Paraskevi (2020): "Investigating the dynamics of illegal online activity: The power of reporting, dark web, and related legislation". *Computer Law & Security Review*, No. 38: pp. 105440. Available in https://doi.org/https://doi.org/10.1016/j.clsr.2020.105440 [fecha de consulta: 10 de abril de 2024].

Krotov, Vlad; Johnson, Leigh & Silva, Leiser (2020): "Tutorial: Legality and ethics of web scraping". *Communications of the Association for Information Systems* No. 47: pp. 539-563. Available in https://doi.org/https://doi.org/10.17705/1CAIS.04724 [fecha de consulta: 10 de abril de 2024].

Leukfeldt, E. Rutger & Kleemans, Edward R. (2021): "Breaking the walls of silence: Analyzing criminal investigations to improve our understanding of cybercrime", in Lavorgna, Anita & Holt, Thomas J. (eds.), *Researching cybercrimes. Methodologies, ethics, and critical approaches* (Cham: Palgrave Macmillan). Available in https://doi.org/10.1007/978-3-030-74837-1_7 [fecha de consulta: 12 de abril de 2024].

Liggett, Roberta; Lee, Jin R.; Roddy, Ariel L. & Wallin, Mikaela A. (2020): "The Dark Web as a platform for crime: An exploration of illicit drug, firearm, CSAM, and cybercrime markets", in Holt, Thomas J. & Bossler, Adam M. (eds.), *The Palgrave Handbook of international cybercrime and cyberdeviance* (Cham: Palgrave Macmillan). Available in https://doi.org/10.1007/978-3-319-78440-3_17 [fecha de consulta: 10 de abril de 2024].

Luis García, Elena de (2023): "Justicia, inteligencia artificial y derecho de defensa". *Revista de Internet, Derecho y Política*, n.º 39. Disponible en https://doi.org/http://dx.doi.org/10.7238/idp.v0i39.417164 [fecha de consulta: 11 de abril de 2024].

Lukings, Melissa & Lashkari, Arash Habibi (2022): *Understanding cybersecurity law and digital privacy. A Common Law Perspective* (Cham: Springer).

Luscombe, Alex; Dick, Kevin & Walby, Kevin (2022): "Algorithmic thinking in the public interest: Navigating technical, legal, and ethical hurdles to web scraping in the social sciences". *Quality & Quantity* No. 56: pp. 1023-1044. Available in https://doi.org/https://doi.org/10.1007/s11135-021-01164-0 [fecha de consulta: 11 de abril de 2024].

Luscombe, Alex; Duncan, Jamie & Walby, Kevin (2022): "Jumpstarting the justice disciplines: A computational-qualitative approach to collecting and analyzing text and image data in criminology and criminal justice studies". *Journal of Criminal*

Justice Education, vol. 33, No. 2: pp. 151-71. Available in https://doi.org/https://doi.org/10.1080/10511253.2022.2027477 [fecha de consulta: 12 de abril de 2024].

Maass, Wolfgang; Parsons, Jeffrey; Purao, Sandeep; Rosales, Alirio; Storey, Veda C. & Woo, Carson C. (2022): "Big data and theory", in Schintler, Laurie A. & McNeely, Connie L. (eds.), *Encyclopedia of Big Data* (Cham: Springer). Available in https://doi.org/10.1007/978-3-319-32010-6_508 [fecha de consulta: 12 de abril de 2024].

Macapinlac, Tess (2019): "The legality of web scraping: A proposal". *Federal Communications Law Journal* vol. 71 No. 3: pp. 399-422.

Maier, Julio (2016): *Derecho procesal penal. I Fundamentos* (Buenos Aires: Ad-Hoc).

Matousek, Roman & Xiang, Dong (2021): "The Challenges and Competitiveness of Fintech Companies in Europe, UK and USA: An Overview", in Pompella, Maurizio y Matorusek, Roman (eds.), *The Palgrave Handbook of FinTech and Blockchain* (Cham: Springer). Available in https://doi.org/10.1007/978-3-030-66433-6_5 [fecha de consulta: 12 de abril de 2024].

McGinnis, John O. & Pearce Russell G. (2014): "The Great Disruption: How Machine Intelligence Will Transform The Great Disruption: How Machine Intelligence Will Transform the Role of Lawyers in the Delivery of Legal Services the Role of Lawyers in the Delivery of Legal Services", *Fordham Law Review* vol. 82, No 6: pp. 3041-3066. Available in https://ir.lawnet.fordham.edu/cgi/viewcontent.cgi?article=5007&context=flr [fecha de consulta: 12 de abril de 2024].

Mead, Esther & Agarwal, Nitin (2022): "Surface Web vs Deep Web vs Dark Web", in Schintler, Laurie A. y McNeely, Connie L. (eds.), *Encyclopedia of Big Data* (Cham: Springer). Available in https://doi.org/10.1007/978-3-319-32001-4_461-1 [fecha de consulta: 11 de abril de 2024].

Middleton, Stuart E. (2021): "Use of Artificial Intelligence to support cybercrime research", in Lavorgna, Anita & Holt, Thomas J., *Researching cybercrimes. Methodologies, ethics, and critical approaches* (Cham: Springer). Available in https://doi.org/10.1007/978-3-030-74837-1_11 [fecha de consulta: 11 de abril de 2024].

Miró Llinares, Fernando (2019): "El modelo policial que viene: mitos y realidades del impacto de la inteligencia artificial y la ciencia de datos en la prevención policial del crimen", en Martínez Espasa, José (coord.), *Libro blanco de la prevención y seguridad local valenciana. Conclusiones y propuestas del Congreso Valenciano de Seguridad Local: la prevención del siglo XXI* (Valencia: Generalitat Valenciana).

Mohammed, Ishaq Azhar (2020). "Artificial Intelligence for Cybersecurity: A Systematic Mapping of Literature", *International Journal of Innovations in Engineering Research and Technology* vol. 7 No. 9.

Navarro-Dolmestch, Roberto (2018): *Derecho procesal penal chileno I* (Santiago: Ediciones Jurídicas de Santiago).

Núñez Ojeda, Raúl; Beltrán Calfurrapa, Ramón y Santander Akkrass, Nicolás (2019): "Los hallazgos casuales en las diligencias de incautación e intervención de las comunicaciones digitales en Chile. Algunos problemas". *Política Criminal* vol. 14 n.º 28.

Disponible en https://doi.org/10.4067/s0718-33992019000200152 [fecha de consulta: 12 de abril de 2024].

Núñez Ojeda, Raúl y Correa Zacarías, Claudio (2017): "La prueba ilícita en las diligencias limitativas de derechos fundamentales en el proceso penal chileno. Algunos problemas". *Ius et Praxis* vol. 23 n.° 1. Available in https://doi.org/10.4067/s0718-00122017000100007 [fecha de consulta: 12 de abril de 2024].

Nussbaum, Brian & Udoh, Emmanuel Sebastian (2020): "Surveillance, Surveillance Studies, and Cyber Criminality", in Holt, Thomas J. & Bossler, Adam M. (eds.), *The Palgrave Handbook of International Cybercrime and Cyberdeviance* (Cham: Springer) pp. 155-182. Available in https://doi.org/https://doi.org/10.1007/978-3-319-78440-3_16 [fecha de consulta: 10 de abril de 2024].

Paal, Boris P. (2022): "Artificial Intelligence as a Challenge for Data Protection Law", in Voeneky, Silja; Kellmeyer, Philipp; Mueller, Oliver and Burgard, Wolfram (eds.), *The Cambridge Handbook of Responsible Artificial Intelligence* (Cambridge: Cambridge University Press). Available in https://doi.org/https://doi.org/10.1017/9781009207898.023 [fecha de consulta: 10 de abril de 2024].

Parks, Andrew M. (2022): "Unfair collection: Reclaiming control of publicly available personal information from data scrapers". *Michigan Law Review* vol. 120 No. 5: pp. 913-945. Available in https://doi.org/https://doi.org/10.36644/mlr.120.5.unfair [fecha de consulta: 10 de abril de 2024].

Portuese, Aurelien (2022): "Prologue: Algorithmic Antitrust - A Primer", en Portuese, Aureluien (ed.), *Algorithmic Antitrust* (Cham: Springer). Available in https://doi.org/10.1007/978-3-030-85859-9_1 [fecha de consulta: 11 de abril de 2024].

Rademacher, Timo (2020): "Artificial intelligence and law enforcement", in Wischmeyer, Thomas y Rademacher, Timo (eds.), *Regulating artificial intelligence* (Cham: Springer): pp. 225-254 Available in https://doi.org/10.1007/978-3-030-32361-5_10 [fecha de consulta: 11 de abril de 2024].

Rahman, Rizwan Ur & Tomar, Deepak Singh (2020): "A new web forensic framework for bot crime investigation". *Forensic Science International: Digital Investigation* vol. 33. Available in https://doi.org/10.1016/j.fsidi.2020.300943 [fecha de consulta: 11 de abril de 2024].

Rathi, Sudhir Kumar; Lata Pritam Prasad; Soni, Nitin; Jain, Sanat & Telang, Shrikant (2023): "Digital footprints: Opportunities and challenges for online robotics technologies", in Rawat, Romil; Chakrawarti, Rajesh K.; Sarangi, Sanjaya K.; Choudhary, Rahul; Gadwal, Anand S. & Bhardwaj, Vivek (eds.), *Robotic Process Automation* (Hoboken: John Wiley & Sons). Available in https://doi.org/10.1002/9781394166954.ch18 [fecha de consulta: 11 de abril de 2024].

Regmi, Rupesh; Rai, Denesh & Khanal, Shradha (2021): "Fintech and Blockchain: Contemporary Issues, New Paradigms, and Disruption", in Pompella, Maurizio & Matousek, Roman (eds.), *The Palgrave Handbook of FinTech and Blockchain* (Cham: Springer). Available in https://doi.org/10.1007/978-3-030-66433-6_4 [fecha de consulta: 13 de abril de 2024].

Reinsel, David, Gantz, John & Rydning, John (2017): "Data Age 2025: The Evolution of Data to Life-Critical". Available in www.seagate.com/files/www-content/our-story/trends/files/Seagate-WP-DataAge2025-March-2017.pdf [fecha de consulta: 13 de abril de 2024].

Rowe, Michael & Muir, Rick (2021): "Big data policing. Governing the machines?", in McDaniel, John & Pease, Ken (eds.), *Predictive Policing and Artificial Intelligence* (Oxon: Routlegde). Available in https://doi.org/10.4324/9780429265365-13 [fecha de consulta: 13 de abril de 2024].

Sachoulidou, Athina (2023): "Going beyond the 'common suspects': to be presumed innocent in the era of algorithms, big data and artificial intelligence". *Artificial Intelligence and Law.* Available in https://doi.org/10.1007/s10506-023-09347-w [fecha de consulta: 13 de abril de 2024].

Schirmer, Jan-Erik (2020): "Artificial Intelligence and Legal Personality: Introducing 'Teilrechtsfähigkeit': A Partial Legal Status Made in Germany", in Wischmeyer, Thomas & Rademacher, Timo (eds.), *Regulating Artificial Intelligence* (Cham: Springer). Available in https://doi.org/https://doi.org/10.1007/978-3-030-32361-5_6 [fecha de consulta: 13 de abril de 2024].

Sellars, Andrew (2018): "Twenty years of web scraping and the Computer Fraud and Abuse Act". *Boston University Journal of Science & Technology Law* No. 24.

Sirisuriya, S. (2015): "A Comparative Study on Web Scraping". *8th International Research Conference, KDU*, November: pp. 135-140. Available in http://ir.kdu.ac.lk/bitstream/handle/345/1051/com-059.pdf [fecha de consulta: 13 de abril de 2024].

Slobogin, Christopher (2022): "Predictive policing", en Pérez Juan, J. y Sanjuán Andrés, F. (eds.), *La transformación algorítmica del sistema de justicia penal* (Cizur Menor: Aranzadi) pp. 89-99.

Solove, Daniel J. (2006): "A taxonomy of privacy". *University of Pennsylvania Law Review* vol. 154 No. 3.

Terwangne, Cécile de (2022): "Privacy and data protection in Europe: Council of Europe's Convention 108+ and the European Union's GDPR", in González Fuster, Gloria; Van Brakel, Rosamunde & De Hert, Paul (eds.), *Research Handbook on Privacy and Data Protection Law. Values, Norms and Global Politics* (Cheltenham: Edward Elgar Publishing). Available in https://doi.org/https://doi.org/10.4337/9781786438515.00007 [fecha de consulta: 13 de abril de 2024].

Thelwall, Mike & Stuart, David (2006): "Web crawling ethics revisited: Cost, privacy, and denial of service". *Journal of the American Society for Information and Technology* vol. 57 No. 13. Available in https://doi.org/https://doi.org/10.1002/asi.20388 [fecha de consulta: 13 de abril de 2024].

Ugwudike, Pamela (2021): "Data-Driven Technologies in Justice Systems: Intersections of power, data configurations, and knowledge production", in Lavorgna, Anita & Holt, Thomas J. (eds.), Researching cybercrimes. *Methodologies, ethics, and critical approaches* (Cham: Springer). Available in https://doi.org/10.1007/978-3-030-74837-1_5 [fecha de consulta: 13 de abril de 2024].

Utset, Manuel A. (2021): "Predictive policing and criminal law", en McDaniel, John & Pease, Ken G., *Predictive Policing and Artificial Intelligence* (London/New York: Routledge) pp. 163-182. Available in https://doi.org/10.4324/9780429265365-8 [fecha de consulta: 13 de abril de 2024].

Vanden Broucke, Seppe & Baesens, Bart (2018): *Practical Web Scraping for Data Science* (Berkeley: Apress). Available in https://doi.org/https://doi.org/10.1007/978-1-4842-3582-9 [fecha de consulta: 13 de abril de 2024].

Vargas Pávez, Macarena y Fuentes Maureira, Claudio (2018): *Introducción al derecho procesal. Nuevas aproximaciones* (Santiago: DER Ediciones).

Wall, David S. (2014): "Internet mafias? The dis-organisation of crime on the Internet", in Caneppele. Stefano & Calderoni, Francesco (eds.), *Organized crime, corruption and crime prevention. Essays in honor of Ernesto U. Savona* (Cham: Springer) pp. 227-238. Available in https://doi.org/10.1007/978-3-319-01839-3_26 [fecha de consulta: 13 de abril de 2024].

Wall, David S. (2015): "Dis-organised crime: Towards a distributed model of the organization of cybercrime". *The European Review of Organised Crime* vol. 2 No. 2: pp. 71-90.

Wilson, Dean (2018): "Algorithmic patrol. The futures of predictive policing", in Završnik, Aleš (ed.), *Big Data, Crime and Social Control* (London/New York: Routledge): pp. 108-127. Available in https://doi.org/https://doi.org/10.4324/9781315395784-6 [fecha de consulta: 13 de abril de 2024].

Wright, Steven A. (2020): "AI in the Law: Towards Assessing Ethical Risks", in IEEE, *2020 IEEE International Conference on Big Data (Big Data)*. Available in https://doi.org/10.1109/BigData50022.2020.9377950 [fecha de consulta: 13 de abril de 2024].

Xiao, Geoffrey (2021). "Bad Bots: Regulating the Scraping of Public Personal Information". *Harvard Journal of Law & Technology* vol. 34 No. 2. Available in https://jolt.law.harvard.edu/assets/articlePDFs/v34/6.-Xiao-Bad-Bots-Regulating-the-Scraping-of-Public-Personal-Information-edit.pdf [fecha de consulta: 13 de abril de 2024].

Yeung, Karen (2019): *Responsibility and AI*. Available in https://rm.coe.int/responsability-and-ai-en/168097d9c5 [fecha de consulta: 13 de abril de 2024].

Zapata, Francisca (2004): *La prueba ilícita* (Santiago: Lexisnexis).

Završnik, Aleš (2018a): "Algorithmic crime control", in Završnik, Aleš (ed.), *Big Data, crime and social control* (London/New York: Routledge) pp. 131-153 Available in https://doi.org/https://doi.org/10.4324/9781315395784-7 [fecha de consulta: 13 de abril de 2024].

Završnik, Aleš (2018b): "Big data. What is it and why does it matter for crime and social control?", in Završnik, Aleš (ed.), *Big Data, crime and social control* (London/New York: Routledge) pp. 3-28.

Završnik, Aleš (2019): "Algorithmic justice: Algorithms and big data in criminal justice settings". *European Journal of Criminology* vol. 18 No. 5: pp. 623-642. Available in https://doi.org/https://doi.org/10.1177/1477370819876762 [fecha de consulta: 13 de abril de 2024].

Zhao, Bo (2022): "Web scraping", in Schintler, Laurie A. & McNeely, Connie L., *Encyclopedia of Big Data* (Cham: Springer) pp. 951-953. Available in https://doi.org/10.1007/978-3-319-32010-6 [fecha de consulta: 13 de abril de 2024].

Normas

Ley nº 19628, sobre protección de la vida privada, 24 de agosto de 1999.

Ley n.º 21577, fortalece la persecución de los delitos de delincuencia organizada, establece técnicas especiales para su investigación y robustece comiso de ganancias, 15 de junio de 2023.

Jurisprudencia

Tribunal Constitucional Federal de Alemania, BVerfGE 65, 1-71 [www.bverfg. de/e /rs19831215_1bvr020983en.html] [fecha de consulta: 13 de abril de 2024].

Tribunal Constitucional Federal de Alemania, BVerfGE 115, 320-381 [www.bverfg. de/ e/rs20060404_1bvr051802en.html] [fecha de consulta: 13 de abril de 2024].